

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-528823

(P2004-528823A)

(43) 公表日 平成16年9月24日(2004.9.24)

(51) Int. Cl. ⁷		F I		テーマコード (参考)	
C 1 2 N	15/09	C 1 2 N	15/00	Z N A A	4 B O 2 4
A 6 1 K	39/00	A 6 1 K	39/00	H	4 B O 6 3
A 6 1 K	39/395	A 6 1 K	39/395	D	4 B O 6 4
A 6 1 K	48/00	A 6 1 K	39/395	N	4 B O 6 5
A 6 1 P	35/00	A 6 1 K	48/00		4 C O 8 4
		審査請求	未請求	予備審査請求	有 (全 367 頁) 最終頁に続く

(21) 出願番号	特願2002-550085 (P2002-550085)	(71) 出願人	502160545
(86) (22) 出願日	平成13年10月30日 (2001. 10. 30)		ディアデクサス インコーポレーテッド
(85) 翻訳文提出日	平成15年4月30日 (2003. 4. 30)		アメリカ合衆国 カリフォルニア州 94
(86) 国際出願番号	PCT/US2001/051341		O 8 O、サウス サン フランシスコ、オ
(87) 国際公開番号	W02002/048370		イスター ポイント ブルバード 343
(87) 国際公開日	平成14年6月20日 (2002. 6. 20)	(74) 代理人	100102842
(31) 優先権主張番号	60/244, 717		弁理士 葛和 清司
(32) 優先日	平成12年10月31日 (2000. 10. 31)	(72) 発明者	スン, ヨンミン
(33) 優先権主張国	米国 (US)		アメリカ合衆国 カリフォルニア州 94
			O 6 5、レッドウッド シティ、ショール
			ドライブ 551
		(72) 発明者	レシボン, ハーヴ
			アメリカ合衆国 カリフォルニア州 94
			115、サン フランシスコ、フォーチュ
			ナ アベニュー 85
			最終頁に続く

(54) 【発明の名称】 大腸特異的遺伝子およびタンパク質に関する組成物および方法

(57) 【要約】

本発明は、正常および癌性大腸細胞に存在する新規に同定された核酸およびポリペプチドに関し、これには、当該核酸およびポリペプチドの断片、変異体および誘導体が含まれる。本発明はまた、本発明のポリペプチドに対する抗体、ならびに本発明のポリペプチドのアゴニストおよびアンタゴニストにも関する。本発明はまた、本発明の核酸、ポリペプチド、抗体、変異体、誘導体、アゴニストおよびアンタゴニスト、ならびにこれらの組成物を使用する方法にも関する。これらの使用は、大腸癌および大腸における非癌性の疾患状態の同定、診断、モニタリング、病期分類、画像化および処置、大腸組織の同定、本発明のポリペプチドのアゴニストおよびアンタゴニストのモニタリングおよび同定および／または設計を含む。当該使用はまた、遺伝子治療、トランスジェニック動物および細胞の作製、および処置および研究用の操作された大腸組織の作製を含む。

【特許請求の範囲】

【請求項 1】

- (a) 配列番号 148～250 のアミノ酸配列をコードする核酸配列を含む核酸分子、
(b) 配列番号 1～147 の核酸配列を含む核酸分子、
(c) (a) または (b) の核酸分子に選択的にハイブリダイズする核酸分子、または
(d) (a) または (b) の核酸分子と少なくとも 60% の配列同一性を有する核酸分子

、を含む、単離核酸分子。

【請求項 2】

核酸分子が cDNA である、請求項 1 に記載の核酸分子。

10

【請求項 3】

核酸分子がゲノム DNA である、請求項 1 に記載の核酸分子。

【請求項 4】

核酸分子が哺乳類の核酸分子である、請求項 1 に記載の核酸分子。

【請求項 5】

核酸分子がヒトの核酸分子である、請求項 4 に記載の核酸分子。

【請求項 6】

試料中の大腸特異的核酸 (CSNA) の存在を決定する方法であって、

(a) 請求項 1 に記載の核酸分子を、該核酸分子が大腸特異的核酸に選択的にハイブリダイズする条件下で、試料と接触させる工程、および

20

(b) 該核酸分子の試料中の CSNA へのハイブリダイゼーションを検出する工程、
を含み、ハイブリダイゼーションの検出が、試料中の CSNA の存在を示す、前記方法。

【請求項 7】

請求項 1 に記載の核酸分子を含む、ベクター。

【請求項 8】

請求項 7 に記載のベクターを含む、宿主細胞。

【請求項 9】

請求項 1 に記載の核酸分子によってコードされるポリペプチドを産生する方法であって、

(a) 1 または 2 以上の発現制御配列に作動可能に連結した核酸分子を含む宿主細胞を提供する工程、および (b) ポリペプチドが産生される条件下で宿主細胞を培養する工程、
を含む、前記方法。

30

【請求項 10】

請求項 1 に記載の核酸分子によってコードされるポリペプチド。

【請求項 11】

(a) 配列番号 148～250 と少なくとも 60% の配列同一性を有するアミノ酸配列を含むポリペプチド、または

(b) 配列番号 1～147 の核酸配列を含む核酸分子によってコードされるアミノ酸配列を含むポリペプチド、

からなる群から選択される単離ポリペプチド。

【請求項 12】

40

請求項 11 に記載のポリペプチドに特異的に結合する抗体またはその断片。

【請求項 13】

試料中の大腸特異的タンパク質の存在を決定する方法であって、

(a) 請求項 12 に記載の抗体を、該抗体が大腸特異的タンパク質に選択的に結合する条件下で、試料と接触させる工程、および

(b) 試料中の大腸特異的タンパク質への抗体の結合を検出する工程、
を含み、結合の検出が試料中の大腸特異的タンパク質の存在を示す、前記方法。

【請求項 14】

患者における大腸癌の存在および転移を診断およびモニタリングする方法であって、

(a) 患者試料中の請求項 1 の核酸分子または請求項 6 のポリペプチドの量を決定する工

50

程、および

(b) 患者試料中の決定された核酸分子またはポリペプチドの量を、正常対象中の大腸特異的マーカーの量と比較する工程を含み、試料中の核酸分子またはポリペプチドの量の、正常対照における核酸分子またはポリペプチドの量に対する差が、大腸癌の存在に関連している、前記方法。

【請求項 15】

患者における癌の危険性または癌の存在を検出するためのキットであって、該キットが、患者試料中の請求項 1 の核酸分子または請求項 6 のポリペプチドの存在を決定するための手段を含む、前記キット。

【請求項 16】

大腸癌を有する患者を処置する方法であって、請求項 12 に記載の組成物をそれが必要な患者に投与する工程を含み、該投与が、核酸分子またはポリペプチドを発現している大腸癌細胞に対する免疫応答を誘発する、前記方法。

【請求項 17】

請求項 11 に記載のポリペプチドまたは該ポリペプチドをコードする核酸を含むワクチン。

【発明の詳細な説明】

【0001】

本出願は、その全体が参照によって本明細書に組込まれる 2000 年 10 月 31 日提出の米国仮出願第 60/244,717 号に基づく優先権の利益を主張する。

【0002】

発明の分野

本発明は、正常のおよび腫瘍性の大腸細胞に存在する、新たに同定された核酸分子およびポリペプチド（該核酸およびポリペプチドの断片、変異体並びに誘導体を含む）に関する。また本発明は、本発明のポリペプチドに対する抗体並びに本発明のポリペプチドのアゴニストおよびアンタゴニストに関する。また本発明は、本発明の核酸、ポリペプチド、抗体、変異体、誘導体、アゴニストおよびアンタゴニストを含む組成物並びにこれら組成物の使用のための方法に関する。これらの使用には、大腸癌および大腸組織における非癌性疾患状態の同定、診断、モニタリング、病期分類、画像化および処置、並びに大腸組織の同定およびモニタリング、並びに本発明のポリペプチドのアゴニストおよびアンタゴニストの同定および／または設計を含む。使用はまた、遺伝子治療、形質転換動物および細胞の作出並びに処置および調査のために操作された大腸組織の作出を含む。

【0003】

発明の背景

結腸直腸癌は、米国において、2 番目に最も多い癌死の原因であり、男性および女性における 3 番目に最も蔓延した癌である。Colon and Rectal Cancer 47 (Peter S. Edelstein ed., 2000) 中、M. L. Davila & A. D. Davila, Screening for Colon and Rectal Cancer。毎年およそ 100,000 人の患者が大腸癌に苦しみ、およそその半数がその疾患により死亡する。Colon and Rectal Cancer 157 (Peter S. Edelstein ed., 2000) 中、Hanna Ngoc Ha & Bard C. Cosman, Treatment of Colon Cancer。結腸直腸癌のほとんど全てのケースが、腺腫性ポリープから起き、そのいくつかは大きなポリープに成熟し、異常に増殖および発達し、そして最終的に癌に進行する。Davila & Davila、前述の 55-56。この進行は、ほとんどの患者において少なくとも 10 年かかるように思われ、癌が局在している場合、早期に診断すれば、癌の容易に処置可能な形態となる。同書 56; Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 125 (1998)。

【0004】

大腸癌の病因についての我々の理解は、継続的に洗練されてきているが、この領域での高度な研究によって、年齢、遺伝的および非遺伝的条件、並びに環境／食事因子を含む因子の組合せが指摘されている。年齢は、結腸直腸癌の発達における鍵となる危険因子であり（Davila & Davila, 前述48）、40歳を超えた男性および女性は、この癌に対する感受性が高まるようになる（Burdette, 前述126）。罹患率は、人生の各10年毎に大幅に増加する。Davila et al., 前述48。多数の遺伝的および非遺伝的条件もまた、家族性大腸腺腫症（FAP）、遺伝性非腺腫性結腸直腸癌（リンチ症候群またはHNPCC）、結腸直腸癌もしくは腺腫性ポリープの個人歴および／または家族歴、炎症性腸疾患、糖尿病および肥満を含み、結腸直腸癌の発生の危険を高めることに関連している。同書47；Colon and Rectal Cancer 67-68（Peter S. Edelstein ed., 2000）中、Henry T. Lynch & Jane F. Lynch, Hereditary Nonpolyposis Colorectal Cancer（Lynch Syndrome）。

【0005】

FAPのケースにおいて、5q21に位置する腫瘍抑制遺伝子APC（大腸腺腫様ポリポース（adenomatous polyposis coli））は、変異により不活性化されるか、また欠失される（Alberts et al., Molecular Biology of the Cell 1288（3rd ed., 1994））。APCタンパク質は、細胞接着、アポトーシス、およびc-myc癌遺伝子の抑制を含む多数の機能の役割を担っている。N. R. Hall & R. D. Madoff, Genetics and the Polyp-Cancer Sequence, Colon and Rectal Cancer 8（Peter S. Edelstein, ed., 2000）。正常APC遺伝子を有する結腸直腸癌をもつ患者において、65%を超える者が、他の組織においては無いのであるが、癌細胞においては、そのような変異を有している。Alberts et al., 前述1288。HNPCCのケースにおいて、患者は、腫瘍抑制遺伝子HNPCCにおいて異常を示すが、腫瘍の約15%だけが変異した遺伝子を含む。同書。他の遺伝子の宿主はまた、結腸直腸癌と関連し、K-ras、N-ras、H-rasおよびc-myc癌遺伝子、並びに腫瘍抑制遺伝子DCC（大腸癌において欠失（deleted in colon carcinoma））およびp53を含む。Hall & Madoff、前述8-9；Alberts et al., 前述1288。

【0006】

結腸直腸癌の危険を増大するのに関連する環境／食事因子は、高脂食、高赤肉食の摂取、および定住型のライフスタイルを含む。Davila & Davila, supra at 47；Reddy, B. S., Prev. Med. 16（4）：460-7（1987）。反対に、結腸直腸癌の危険を低減するのに関連する環境／食事要因は、線維、葉酸、カルシウムの高い食事および閉経後の女性におけるホルモン代替療法を含む。Davila & Davila、前述50-55。大腸癌の危険性を低減することにおける酸化防止剤の効果は明確ではない。同書53。

早期の、限局した病期に検出した場合、大腸癌は極めて処置しやすいので、スクリーニングは、50歳から始まる全ての成人、特に結腸直腸癌の第1度近親者をもつ人に対する、日常的なケアの一部とされるべきである。結腸直腸癌の他のタイプの癌におけるその対応物に対する結腸直腸癌のスクリーニングの1つの大きな利点は、前癌状態の病変を検出し得るだけでなく、それらを十分に除去し得ることである。Davila & Davila、上記56。今日用いられている鍵となる結腸直腸癌スクリーニングテストは、便潜血検査、S字結腸鏡検査、結腸鏡検査、二重造影バリウム注腸および癌胎児性抗原（CEA）テストである。同書、Burdette、上記125。

【0007】

便潜血検査（FOBT）は、大便中の血液量を検出することによって結腸直腸癌をスクリー

ーニングするが、腫瘍性組織、特に悪性の組織が、ポリープの大きさおよび癌の病期とともに出血の量が増大するのに伴ない、典型的な粘膜よりも多く出血することが前提となる。Davila & Davila、上記56-57。早い病期にある腫瘍を検出するのに効果的である一方、FOBTは、腺腫性ポリープ（前癌病変）を検出することはできず、そして、排泄物試料の内容物に依存して、偽陽性を示すことになる。同書56-59。対照的に、S字結腸鏡検査および結腸鏡検査は、腸の直接的な視覚化を可能にし、腺腫性ポリープを検出、バイオプシーおよび除去することを可能にする。同書、59-60、61。これらの方法の利点にもかかわらず、これらは否定的な面を伴う：S字結腸鏡検査は、その定義のとおり、S字結腸以下に限定されるし、結腸鏡検査は、比較的高価な方法であり、両方とも腸の穿孔および大出血の可能性のリスクを有している。同書59-60。二重造影バリウム注腸（DCBE）は、FOBTよりも優れて、結腸鏡検査とほとんど同等に、病変の検出が可能であるが、曲がりくねった直腸S字結腸接合部（winding rectosigmoid region）を評価することに限られてしまう。同書60。癌胎児性抗原に対する血液のスクリーニングを含むCEA血液検査は、早い病期の結腸直腸癌の検出において限定された有用性を有するという、FOBTの否定的な面を有する。Burdette、上記125。

【0008】

一度大腸癌が診断されれば、処置の決定は、典型的には、癌進行の病期を参照して行なわれる。多数の技術が、癌の病期分類に用いられ（そのいくつかは、大腸癌のスクリーニングに用いられる）、切除された大腸の病理学的検査、S字結腸鏡検査、結腸鏡検査および種々の画像化技術を含む。AJCC Cancer Staging Handbook 84（Irvin D. Fleming et al. eds., 5th ed. 1998）；Montgomery, R. C. and Ridge, J. A., Semin. Surg. Oncol. 15（3）：143-150（1998）。さらに、胸部写真（chest films）、肝機能検査および肝スキャン（liver scans）が、転移の範囲を決定するのに用いられる。Fleming et al. eds., 上記84。コンピュータ断層撮影および磁気共鳴映像法は、結腸直腸癌の病期分類に有用である一方、この両方とも疾患の早い病期を同定するのに受け入れ難い程、病期分類の正確性が低い。これは、両方法が、（1）腸壁の腫瘍の浸潤の深さを示すことおよび（2）悪性のアデノパシーを診断することにおける困難さを有するためである。Thoeni, R. F., Radiol. Clin. N. Am. 35（2）：457-85（1997）。むしろ、経直腸的な超音波（TRUS）などの技術は、転移を含んでいるかもしれない小さなリンパ節を検出することに関しては不正確ではあるが、この状況において好まれる。Colon and Rectal Cancer 316（Peter S. Edelstein ed., 2000）中、David Blumberg & Frank G. Opelka, Neoadjuvant and Adjuvant Therapy for Adenocarcinoma of the Rectum。

【0009】

いくつかの分類システムは、デュークシステム（Dukes' system）および当業界において多くの人により有用な病期分類システムと考えられている、より詳細な国際対癌連合—米国合同委員会の癌TNM病期分類システムを含み、結腸直腸癌の程度を病期分類することを講じてきた。Burdette、上記126-27。臨床的なまたは病理学的な病期分類に用いられるTNMシステムは、4つの病期に分かれている。その夫々は、原発腫瘍（T）、領域リンパ節（N）および遠隔転移（M）に関して癌の成長の程度を評価する。Fleming et al. eds., 上記84-85。該システムは、腸壁への腫瘍浸潤の範囲、隣接構造の侵入、冒された領域リンパ節の数、および遠隔転移が起きているか否かに焦点を当てている。同書81。

【0010】

病期0は、上皮内癌（Tis）によって特徴付けられ、ここで癌細胞は、腺基底膜（上皮内）または固有層（粘膜内）の内側に局在している。同書84-85；Burdet

10

20

30

40

50

te、上記127。この病期において、癌は領域リンパ節に広がっておらず(N0)、遠隔転移もない(M0)。Fleming et al. eds.、上記85; Burdette、上記127。病期Iにおいて、未だ領域リンパ節への癌の広がりはなく、遠隔転移もないが、腫瘍は粘膜下組織に侵入しているか(T1)、または進行し、さらに固有筋層に侵入している(T2)。Fleming et al. eds.、上記84-85; Burdette、上記127。病期IIもまた、領域リンパ節への癌の広がりはなく、遠隔転移もないが、腫瘍は、漿膜下組織または非腹膜結腸周囲(nonperitonealized pericolic)または直腸周囲組織に侵入しているか(T3)、または他の器官もしくは構造に侵入する程に進行しているか、および/または臓側腹膜を穿孔している(T4)。同書。病期3は、任意のTのサブステージによって特徴付けられ、遠隔転移がなく、1-3の領域リンパ節に転移があるか(N1)、または4もしくは5以上の領域リンパ節に転移がある(N2)。Fleming et al. eds.、上記85; Burdette、上記127。最後に、病期4は、遠隔転移とともに、任意のTまたはNのサブステージを含む。同書。

【0011】

近年、病理学的病期分類がより正確な予後を提供するので、大腸癌の病理学的病期分類は、臨床的な病期分類に対し好ましい。病理学的病期分類は、典型的には、切除した大腸切片の検査を含み、これは腹腔の外科的検査を伴う。Fleming et al. eds.、上記84。臨床的な病期分類は、対応物の浸襲性手順に依存しないので、少なくとも病理学的な病期分類と同様に正確であれば、その病期分類は好ましい。

結腸直腸癌の処置になれば、外科的切除によって、およそ50%の患者が治癒する。Burdette、上記125。放射線照射は、結腸直腸癌の処置において、術前にも術後にも用いられる。同書125、132-33。化学療法剤、特に5-フルオロウラシルもまた、結腸直腸癌の処置において強力な武器である。同書125、133。他の剤には、イリノテカン(irinotecan)およびフロクスウリジン、シスプラチン、レバミゾール、メトトレキサート、インターフェロン・アルファおよびロイコボリンが含まれる。同書133。それでもなお、30-40%の患者が、外科的切除の後、大腸癌の再発を発症するだろう。Wayne De Vos, Follow-up After Treatment of Colon Cancer, Colon and Rectal Cancer 225 (Peter S. Edelstein ed., 2000)、多くの患者において死の最大要因である。したがって、大腸癌患者を緊密にモニターし、治療への応答を決定し、持続性または再発性の疾患および転移を検出しなくてはならない。

【0012】

前記より、結腸直腸癌の再発を検出、診断、モニタリング、病期分類、予見および予防するために用いる手法が、患者の転帰に対し極めて重要である。さらに、最近の手法は、これらの各分析に有用であるが、それらの特異性、感受性、浸襲性および/またはそのコストによって限定される。そのため、最小の侵襲性で、および合理的なコストで、細胞、組織または体液における新規マーカーの検出によって行う、高度に特異的で高感度の手法は、非常に望ましい。

したがって、結腸直腸癌の発生しやすいヒトかどうかを予見するための、結腸直腸癌を診断するための、疾患の進行をモニタリングするための、結腸直腸癌を病期分類するための、結腸直腸癌が転移したかどうかを決定するための、および結腸直腸癌を画像化するための、より高感度で正確な方法に対する大きな需要がある。また結腸直腸癌のよりよい処置に対しても需要がある。

【0013】

発明の要約

本発明は、大腸癌および大腸における非癌性の疾患状態を同定、診断、モニタリング、病期分類、画像化および処置するため、大腸組織を同定およびモニタリングするため、そして、本発明のポリペプチドのアゴニストおよびアンタゴニストを同定および設計するために用いることができる、核酸分子およびポリペプチドならびにその抗体、アゴニストおよ

びアンタゴニストを提供することによって、当該技術分野におけるこれらおよび他のニーズを解決する。本発明はまた、遺伝子治療、トランスジェニック動物および細胞の作製方法、および処置および研究用の操作された大腸組織の作製方法を提供する。

従って、本発明の目的の1つは、大腸細胞、大腸組織および／または大腸器官に特異的な核酸分子を提供することである。これらの大腸特異的な核酸(CSNA)は、天然に存在するcDNA、ゲノムDNA、RNAもしくはこれらの核酸のうちの1つの断片であってもよく、または天然に存在しない核酸分子であってもよい。CSNAがゲノムDNAである場合、CSNAは大腸特異的な遺伝子(CSG)である。好ましい態様では、核酸分子は、大腸に特異的なポリペプチドをコードする。より好ましい態様では、核酸分子は、配列番号148～250のアミノ酸配列を含むポリペプチドをコードする。別の非常に好ましい態様では、核酸分子は配列番号1～147の核酸配列を含む。核酸分子はまた、CSPをコードする核酸分子に選択的にハイブリダイズするかまたはそれと実質的な配列類似性を示す配列、またはCSNAをコードする核酸分子に選択的にハイブリダイズするかまたは実質的な配列類似性を示す配列、並びにCSPをコードする核酸分子の相同変異体およびCSNAの相同変異体を含むこととする。CSPをコードする核酸配列の部分を含む核酸分子、またはCSNAの核酸配列の部分を含む核酸分子もまた提供される。

本発明に関連する目的は、CSNAの全部または一部の転写および／または翻訳を制御する1または2以上の発現制御配列を含む核酸分子を提供することである。好ましい態様では、核酸分子は、CSPの全部または断片をコードする核酸分子の転写および／または翻訳を制御する1または2以上の発現制御配列を含む。

【0014】

本発明の他の目的は、本発明の核酸分子を含むベクターおよび／または宿主細胞を提供することである。好ましい態様では、核酸分子はCSPの全部または断片をコードする。他の好ましい態様では、核酸分子はCSNAの全部または一部を含む。

本発明の他の目的は、本発明のポリペプチドを組換え的に製造するために、本発明の核酸分子を含むベクターまたは宿主細胞を用いる方法を提供することである。

本発明の他の目的は、本発明の核酸分子によりコードされたポリペプチドを提供することである。好ましい態様では、ポリペプチドはCSPである。ポリペプチドは、断片または全長タンパク質のいずれか、並びに突然変異タンパク質(ミューテイン)、融合タンパク質、相同タンパク質またはCSPの対立遺伝子変異体によりコードされるポリペプチドを含み得る。

本発明の他の目的は、本発明のポリペプチドに特異的に結合する抗体を提供することである。

本発明の他の目的は、本発明の核酸分子およびポリペプチドのアゴニストおよびアンタゴニストを提供することである。

本発明の他の目的は、核酸分子を用い、本明細書中に記載された核酸分子に比較して類似または同一の核酸配列を有する核酸分子を検出または増幅するための方法を提供することである。好ましい態様では、本発明は、大腸癌および大腸における非癌性の疾病状態を同定、診断、モニタリング、病期分類、画像化および処置するために核酸分子を用いる方法を提供する。他の好ましい態様では、本発明は、大腸組織を同定および／またはモニタリングするために本発明の核酸分子を用いる方法を提供する。本発明の核酸分子はまた、遺伝子治療に、トランスジェニック動物および細胞を作製するために、そして、処置および研究用に操作された大腸組織を作製するために用いることができる。

本発明のポリペプチドおよび／または抗体はまた、大腸癌および大腸における非癌性の疾病状態を同定、診断、モニタリング、病期分類、画像化および処置するために用いることができる。本発明は、大腸組織を同定および／またはモニタリングするために、および操作された大腸組織を作製するために、本発明のポリペプチドを用いる方法を提供する。

本発明のアゴニストおよびアンタゴニストは、大腸癌および大腸における非癌性の疾病状態を処置するために、および操作された大腸組織を作製するために用いることができる。

本発明のさらに別の目的は、本発明の核酸およびアミノ酸配列を保存するためのコンピュ

10

20

30

40

50

ータ可読な手段を提供することである。コンピュータ可読な記録は、本発明の配列と他の配列との比較、アラインメントおよびオーダリングの目的で、配列を読み込みそして表示するためにアクセスすることができる。

【0015】

発明の詳細な説明

定義および一般的な技術

本明細書中に別記のない限り、本発明に関して用いられる科学および技術的用語は、当業者に通常理解されている意味を有するものとする。さらに、状況により特に必要とされない限り、単数形の用語は複数形を含み、複数形の用語は単数形を含むものとする。一般的に、本明細書中に記載された細胞および組織培養、分子生物学、免疫学、微生物学、遺伝子およびタンパク質および核酸化学ならびにハイブリダイゼーションに関して用いられる用語、およびその技術は、当該技術分野においてよく知られ、通常用いられているものとする。一般的に、別記のない限り、本発明の方法および技術は、当該技術分野においてよく知られた慣用の方法に従って、本明細書中で引用され、議論されている種々の一般的なおよびより専門的な参考文献に記載されたとおりに行われる。参照により、そのそれぞれが全体として本明細書に組み込まれる、例えば、Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) および Sambrook et al., Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates (1992, および2000の補遺); Ausubel et al., Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology - 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); および Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999)を参照。

酵素反応および精製技術は、製造者の仕様書に従い、当該技術分野において通常なされているとおりに、または本明細書に記載のとおりに行う。本明細書中に記載された分析化学、合成有機化学ならびに医薬品化学および薬化学に関して用いられる用語、ならびにその実験手順および技術は、当該技術分野においてよく知られ、通常用いられているものである。標準的な技術を、化学合成、化学分析、薬剤の製造、製剤および送達、ならびに患者の処置に用いる。

【0016】

以下の用語は、別記のない限り、以下の意味を有すると理解されるものとする。

本発明の「核酸分子」は、ヌクレオチドの重合形態を指し、RNA、cDNA、ゲノムDNAのセンスおよびアンチセンス鎖の両方、ならびに上記の合形成態および混合重合体を含む。ヌクレオチドは、リボヌクレオチド、デオキシヌクレオチド、またはいずれかのタイプのヌクレオチドの修飾形態を指す。本明細書中で用いる「核酸分子」は、「核酸」および「ポリヌクレオチド」の類義語である。「核酸分子」は、別記のない限り、通常、少なくとも10塩基長の分子を指す。この用語は、DNAの一本鎖および二本鎖形態を含む。加えて、ポリヌクレオチドは、天然に存在するおよび／または天然に存在しないヌクレオチド結合によって結合した天然に存在するヌクレオチドおよび修飾ヌクレオチドの一方または両方を含む得る。

核酸分子は、化学的または生化学的に修飾されてもよく、または非天然のまたは誘導体化

されたヌクレオチド塩基を含んでもよく、これらは当業者によって容易に評価されるだろう。かかる修飾は、例えば、標識、メチル化、1または2以上の天然に存在するヌクレオチドの、類縁体による置換、無電荷結合（例えば、メチルホスホネート、ホスホトリエステル、ホスホロアミデート、カルバメートなど）、電荷結合（例えば、ホスホロチオエート、ホスホジチオエートなど）、ペンダント部分（例えば、ポリペプチド）、インターカレーター（例えば、アクリジン、プソラレンなど）、キレータ、アルキレータ、および修飾された結合（例えば、アルファアノマー核酸など）などのヌクレオチド間修飾を含む。「核酸分子」の用語はまた、一本鎖構造、二本鎖構造、部分的に二重化された構造、三重化された構造、ヘアピン状の構造、環状構造およびパドロック構造を含む、任意のトポロジー構造を含む。水素結合および他の化学的相互作用を介して所定の配列と結合するその能力について、ポリヌクレオチドを模倣する合成分子もまた含まれる。かかる分子は当該技術分野で知られており、例えば、分子の主鎖においてリン酸結合がペプチド結合で置換されているものを含む。

10

【0017】

「遺伝子」は、ポリペプチドをコードする核酸配列、およびポリペプチドをコードする核酸配列の周囲の発現制御配列を含む核酸分子として定義される。例えば、遺伝子は、プロモーター、1または2以上のエンハンサー、ポリペプチドをコードする核酸配列、下流の調節配列、および、場合によっては、RNAの発現の調節に關与する他の核酸配列を含み得る。当該技術分野でよく知られているとおり、真核生物の遺伝子は、通常エクソンおよびイントロンの両方を含有する。「エクソン」の用語は、ゲノムDNA中に見出される核酸配列であって、成熟mRNA転写物のコンティグ配列に寄与することが生物情報学的に予測され、および／または実験的に確認されているものを指す。「イントロン」の用語は、ゲノムDNA中に見出される核酸配列であって、成熟mRNA転写物のコンティグ配列に寄与しないが、むしろ転写の過程に「スプライスアウト」されることが予測され、および／または確認されているものを指す。

20

核酸分子またはポリペプチドは、核酸分子またはポリペプチドが特定の種から単離された場合、または、核酸分子またはポリペプチドが特定の種から単離した核酸分子またはポリペプチドと相同である場合は、その特定の種に「由来」する。

「単離された」または「実質的に純粋な」核酸またはポリヌクレオチド（例えば、RNA、DNAまたは混合された重合体）は、本来の宿主細胞において、ネイティブなポリヌクレオチドに本来付随していた他の細胞成分、例えばリボソーム、ポリメラーゼ、またはそれが本来結びついていたゲノム配列から実質的に分離されたものである。この用語は、（1）天然に存在する環境から取り出された、（2）「単離ポリヌクレオチド」が天然に見出されるポリヌクレオチドの全部または部分と結びついていない、（3）天然では連結していないポリヌクレオチドと作動可能に連結している、（4）天然ではより長い配列の部分として存在しない、または（5）天然では見出せないヌクレオチドまたはヌクレオシド間結合を含む、核酸またはポリヌクレオチドを包含する。用語「単離された」または「実質的に純粋な」はまた、組換えまたはクローンDNA単離物、化学的に合成されたポリヌクレオチド類縁体、または異種の系で生物学的に合成されたポリヌクレオチド類縁体に関して用いることができる。用語「単離核酸分子」は、宿主染色体の非相同的な部位に組み込まれたか、ネイティブ断片の非相同配列への組換え的融合体、エピソームまたは宿主細胞染色体に組み込まれたものとして存在する組換えベクターを含む。

30

40

【0018】

核酸分子の「部分」は、参照核酸分子の少なくとも10塩基の部分的なコンティグ配列を含む核酸分子を指す。好ましくは、1つの部分は参照核酸分子の少なくとも15〜20塩基を含む。理論的には、17ヌクレオチドの核酸配列は、3ギガ塩基のヒトゲノムにおいて1回より少ない頻度でランダムに存在し、従って、複雑なゲノムの核酸混合物において参照配列をユニークに同定することができる核酸プローブを提供するのに十分な長さである。好ましい部分は、少なくとも6個の連続したアミノ酸配列（少なくとも18ヌクレオチドの断片）をコードできる核酸配列を含むものである。なぜなら、それらは参照核酸に

50

よってコードされたポリペプチドのエピトープをマッピングするのに有用である、ペプチドの発現または合成を導くのに有用だからである。その開示の全体が参照により本明細書に組み込まれる、例えば Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1984) および米国特許第4,708,871号および第5,595,915号を参照。部分はまた、参照核酸分子の少なくとも25、30、35または40ヌクレオチド、または、参照核酸分子の少なくとも50、60、70、80、90、100、150、200、250、300、350、400または500ヌクレオチドを含み得る。核酸分子の部分は、他の核酸配列を含まなくてもよい。あるいは、核酸の部分は、他の核酸分子からの他の核酸配列を含んでもよい。

用語「オリゴヌクレオチド」は、一般的に200塩基またはそれより短い長さを含む核酸分子を指す。この用語は、しばしば一本鎖デオキシリボヌクレオチドをさすが、同様に、とりわけ、一本鎖または二本鎖リボヌクレオチド、RNA:DNAハイブリッド、および二本鎖DNAを指すことができる。好ましくは、オリゴヌクレオチドは10~60塩基長であり、最も好ましくは12、13、14、15、16、17、18、19または20塩基長である。他の好ましいオリゴヌクレオチドは、25、30、35、40、45、50、55または60塩基長である。オリゴヌクレオチドは一本鎖でもよく（例えばプローブまたはプライマーとして用いるため）、または二本鎖であってもよい（例えば変異遺伝子の構築に用いるため）。本発明のオリゴヌクレオチドは、センスまたはアンチセンスオリゴヌクレオチドのいずれかであることができる。オリゴヌクレオチドは、核酸分子について上記で論じたとおり、誘導体化または修飾されることができる。

10

20

【0019】

一本鎖DNAプローブオリゴヌクレオチドなどのオリゴヌクレオチドは、しばしば、自動オリゴヌクレオチド合成装置において実行されるものなどの化学的方法により合成される。しかしながら、オリゴヌクレオチドは、例えば *in vitro* 組換えDNA媒介技術を含む種々の他の方法、ならびに細胞および生体でのDNA発現により作製することができる。まず化学的に合成したDNAは、典型的に、5'ホスフェートなしで得られる。かかるオリゴヌクレオチドの5'末端は、組換えDNA分子形成に典型的に用いられるDNAリガーゼを使用する連結反応によるホスホジエステル結合形成の基質ではない。かかるオリゴヌクレオチドの連結が望ましい場合には、キナーゼおよびATPを使用するような標準的技術によりホスフェートを付加することができる。化学的に合成したオリゴヌクレオチドの3'末端は、遊離の水酸基を一般的に有しており、T4DNAリガーゼなどのリガーゼの存在下において、別のオリゴヌクレオチドなどの別のポリヌクレオチドの5'ホスフェートと容易にホスホジエステル結合を形成するだろう。周知の様に、この反応は、望ましい場合、連結前に当該他のポリヌクレオチドの5'ホスフェートを除去することで選択的に阻止することができる。

30

本明細書中で用いられている用語「天然に存在するヌクレオチド」は、天然に存在するデオキシリボヌクレオチドおよびリボヌクレオチドを含む。本明細書中で用いる用語「修飾ヌクレオチド」は、修飾または置換糖基などを有するヌクレオチドを含む。本明細書中で用いる用語「核酸結合」は、ホスホロチオエート、ホスホロジチオエート、ホスホロセレノエート、ホスホロジセレノエート、ホスホロアニロチオエート、ホスホラニラデート、ホスホロアミデートなどを含む。その開示が参照により本明細書に組み込まれる、例えば、LaPlanche et al. Nucl. Acids Res. 14:9081-9093 (1986); Stein et al. Nucl. Acids Res. 16:3209-3221 (1988); Zon et al. AntiCancer Drug Design 6:539-568 (1991); Zon et al., in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); 米国特許第5,151,510号; Uhlmann and Peyman Chemical Reviews 90:543 (1990)を参照。

40

50

【0020】

他に特定されていない限り、センス方向におけるポリヌクレオチド配列の左端は5'末端であり、配列の右端は3'末端である。加えて、センス方向におけるポリヌクレオチド配列の左方向は5'方向と呼び、一方、ポリヌクレオチド配列の右方向は、3'方向と呼ぶ。さらに、他に指示のない限り、本明細書中では、各ヌクレオチド配列は、デオキシリボヌクレオチドの配列として記載されている。しかしながら、与えられた配列は、ポリヌクレオチド組成に適切であるように解釈されることが意図される。例えば、単離された核酸がRNAの組成であれば、与えられた配列は、ウリジンがチミジンについて置換したリボヌクレオチドを意味する。

用語「対立遺伝子変異体」は、2または3以上の代替的に天然に存在する遺伝子の形態の1つであって、各遺伝子がユニークなヌクレオチド配列を保有するものを指す。好ましい態様では、所定の遺伝子の異なる対立遺伝子は、類似または同一の生物学的特性を有する。

核酸配列の場合における、用語「配列同一性パーセント」は、最大に一致するように並べた場合に同一である2つの配列中の残基を指す。配列同一性比較の長さは、少なくとも約9ヌクレオチドの長さを超えてもよく、普通は少なくとも約20ヌクレオチド、より普通には少なくとも約24ヌクレオチド、典型的には少なくとも約28ヌクレオチド、より典型的には少なくとも約32ヌクレオチド、そして好ましくは少なくとも36ヌクレオチドまたは38ヌクレオチド以上であり得る。核酸配列の同一性を測定するのに用いることができる、当該技術分野に知られている数多くの種々のアルゴリズムがある。例えば、ポリヌクレオチド配列は、Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin中のプログラムであるFASTA、GapまたはBestfitを用いて比較することができる。例えば、プログラムFASTA2およびFASTA3を含むFASTAは、照会配列と検索配列との間の最良のオーバーラップ領域の配列および配列同一性パーセントを提供する（本明細書中に参照により組み込むPearson, Methods Enzymol. 183: 63-98 (1990); Pearson, Methods Mol. Biol. 132: 185-219 (2000); Pearson, Methods Enzymol. 266: 227-258 (1996); Pearson, J. Mol. Biol. 276: 71-84 (1998)）。他に特定しない限り、特定のプログラムまたはアルゴリズムのデフォルトパラメータを用いた。例えば、核酸配列同士の配列同一性パーセントは、FASTAを用い、そのデフォルトパラメータ（文字サイズ6およびスコアリングマトリックスについてはNOPAMファクター）により、または、Gapを用いて、GCG Version 6.1で提供されているとおりのデフォルトパラメータにより決定することができ、これらは本明細書中に参照により組み込まれる。

【0021】

核酸配列への参照は、他に特定しない限り、その相補体を包含する。従って、特定の配列を有する核酸分子への参照は、その相補的な配列を有するその相補鎖を包含する。相補鎖はまた、アンチセンス療法、ハイブリダイゼーションプローブおよびPCRプライマーにも有用である。

分子生物学の分野において、研究者は、用語「配列同一性パーセント」、「配列類似性パーセント」および「配列相同性パーセント」を同様の意味に用いている。本出願においては、これらの用語は、核酸配列のみに関して、同様の意味を有するものとする。

用語「実質的な類似性」または「実質的な配列類似性」は、核酸またはその断片を指す場合、適切なヌクレオチドの挿入または欠失とともに、他の核酸（またはその相補鎖）と最適に整列した場合、上記で論じたFASTA、BLASTまたはGapなどの、任意の周知の配列同一性のアルゴリズムにより測定したときに、ヌクレオチド塩基の少なくとも約50%、より好ましくは約60%、普通にはヌクレオチド塩基の少なくとも約70%、より普通には少なくとも約80%、好ましくは少なくとも約90%、そしてより好ましくは

少なくとも約 95～98% に、ヌクレオチド配列同一性があることを示す。

あるいは、実質的な類似性は、選択的なハイブリダイゼーション条件下で、核酸またはその断片が、他の核酸、他の核酸の鎖、またはその相補鎖にハイブリダイズする場合に存する。典型的には、選択的なハイブリダイゼーションは、少なくとも約 14 ヌクレオチド、より好ましくは少なくとも 17 ヌクレオチド、さらにより好ましくは少なくとも 20、25、30、35、40、50、60、70、80、90 または 100 ヌクレオチドの範囲にわたり、少なくとも約 55% の配列同一性、好ましくは少なくとも約 65%、より好ましくは少なくとも約 75%、そして最も好ましくは少なくとも約 90% の配列同一性がある場合に生じる。

【0022】

核酸ハイブリダイゼーションは、塩濃度、温度、溶媒、ハイブリダイズする種の塩基組成、相補領域の長さ、およびハイブリダイズする核酸同士のヌクレオチド塩基のミスマッチなどの条件により影響を受け、これらは当業者によって容易に評価される。核酸ハイブリダイゼーション実験の場面における「ストリンジェントなハイブリダイゼーション条件」および「ストリンジェントな洗浄条件」は、多数の異なる物理学的パラメータに依存する。最も重要なパラメータは、ハイブリダイゼーションの温度、核酸の塩基組成、塩濃度および核酸の長さを含む。当業者は、いかにしてこれらのパラメータを変化させて特定のハイブリダイゼーションのストリンジェンシーを達成するかを知っている。一般的に、「ストリンジェントなハイブリダイゼーション」は、具体的な DNA ハイブリッドの融解温度 (T_m) より約 25℃ 下で、特定の条件のセットの下で行われる。「ストリンジェントな洗浄」は、具体的な DNA ハイブリッドの T_m より約 5℃ 低い温度で、特定の条件のセットの下で行われる。 T_m は、標的配列の 50% が、完全にマッチしたプローブにハイブリダイズする温度である。本明細書中に参照により組み込まれる Sambrook (1989)、前出、p. 9.51 を参照。

特定の DNA-DNA ハイブリッドの T_m は、式：

$$T_m = 81.5^{\circ}\text{C} + 16.6 (\log_{10} [\text{Na}^+]) + 0.41 (\text{フラクション G} + \text{C}) - 0.63 (\% \text{ホルムアミド}) - (600/l)$$

式中、 l は塩基対で表したハイブリッドの長さである、により予測することができる。

特定の RNA-RNA ハイブリッドの T_m は、式：

$$T_m = 79.8^{\circ}\text{C} + 18.5 (\log_{10} [\text{Na}^+]) + 0.58 (\text{フラクション G} + \text{C}) + 11.8 (\text{フラクション G} + \text{C})^2 - 0.35$$

により予測することができる。

特定の RNA-DNA ハイブリッドの T_m は、式：

$$T_m = 79.8^{\circ}\text{C} + 18.5 (\log_{10} [\text{Na}^+]) + 0.58 (\text{フラクション G} + \text{C}) + 11.8 (\text{フラクション G} + \text{C})^2 - 0.50 (\% \text{ホルムアミド}) - (820/l)$$

により予測することができる。

一般的に、 T_m は、2つの核酸配列同士のミスマッチ 1% ごとに、1～1.5℃ ずつ低下する。従って、当業者は、標的核酸に対し、より高いまたは低い程度の配列同一性を得るために、ハイブリダイゼーションおよび／または洗浄条件を変えることができる。例えば、標的核酸配列に対し 10% までのミスマッチを含むハイブリダイズする核酸を得るには、完全にマッチするハイブリッドの計算された T_m から 10～15℃ を減じ、次にそれに従って、ハイブリダイゼーションおよび洗浄温度を調節する。プローブ配列もまた、ある条件の下で 2 量体 DNA に特異的にハイブリダイズし、3 量体または他の高次 DNA 複合体を形成し得る。かかるプローブの調製および好適なハイブリダイゼーション条件は当該技術分野でよく知られている。

【0023】

サザンもしくはノザンプロットにおける、またはライブラリーをスクリーニングするための、100 を超える相補的な残基を有する相補核酸配列のフィルタ上でのハイブリダイゼ

10

20

30

40

50

ーションのための、ストリンジェントなハイブリダイゼーション条件の例は、42℃での、少なくとも10時間、好ましくは1晩（約16時間）の50%ホルムアミド/6×SSCである。ストリンジェントなハイブリダイゼーション条件の他の例は、68℃、ホルムアミドなしでの、少なくとも10時間および好ましくは1晩の6×SSCである。中程度のストリンジェンシーのハイブリダイゼーション条件の例は、55℃、ホルムアミドなしでの、少なくとも10時間および好ましくは1晩の6×SSCである。サザンもしくはノザンブロットにおける、またはライブラリーをスクリーニングするための、100を超える相補的な残基を有する相補核酸配列のフィルタ上でのハイブリダイゼーションのための、低いストリンジェンシーのハイブリダイゼーション条件の例は、42℃での、少なくとも10時間の6×SSCである。類似だが同一でない核酸配列を同定するためのハイブリダイゼーション条件は、実験的に、塩濃度を一定に（6×SSC）保ちながらハイブリダイゼーション温度を68℃から42℃に変化させること、または、ハイブリダイゼーション温度および塩濃度を一定に（例えば42℃および6×SSC）保ち、ホルムアミド濃度を50%から0%に変化させることによって同定することができる。ハイブリダイゼーションバッファーはまた、バックグラウンドを下げるためのブロック剤を含み得る。これらの剤は、当該技術分野においてよく知られている。本明細書中に参照により組み込まれる、上記Sambrook et al. (1989), pages 8.46 and 9.46-9.58を参照。また、上記Ausubel (1992)、上記Ausubel (1999)、および上記Sambrook (2001)も参照。

洗浄条件もまた、ストリンジェンシー条件を変化させるために変えることができる。ストリンジェントな洗浄条件の例は、65℃で15分間の0.2×SSC洗浄である（SSCバッファーについては、上記Sambrook et al. (1989)を参照）。しばしば、高いストリンジェンシーの洗浄の前に、過剰なプローブを除去するために、低いストリンジェンシーの洗浄がなされる。100塩基対より多い2量体DNAのための中程度のストリンジェンシーの洗浄の例は、45℃での15分間の1×SSCである。かかる2量体のための低ストリンジェンシーの洗浄の例は、40℃での15分間の4×SSCである。一般的に、特定のハイブリダイゼーションアッセイにおける無関係なプローブについて観察されたものより2倍以上高いシグナル対ノイズ比は、特異的ハイブリダイゼーションの検出を示す。

本明細書で定義するとおり、ストリンジェントな条件下で互いにハイブリダイズしない核酸分子は、互いに実質的に同一のポリペプチドをコードする場合は、依然として互いに実質的に類似である。これは、例えば、遺伝子コードの冗長性により許容される高度のコードンの縮重を用い、合成的にまたは組換え的に創出された場合に生じる。

【0024】

100ヌクレオチド長より短い核酸分子のための（例えばオリゴヌクレオチドのための）ハイブリダイゼーション条件は、式：

$$T_m = 81.5^{\circ}\text{C} + 16.6 (\log_{10} [\text{Na}^+]) + 0.41 (\text{フラクション } G+C) - (600/N)$$

式中、Nは長さの変化であり、 $[\text{Na}^+]$ は1M以下である、

により計算することができる。上記Sambrook et al. (1989), p. 40

11.46を参照。100ヌクレオチドより短いプローブのハイブリダイゼーションに関しては、ハイブリダイゼーションは、通常、ストリンジェントな条件（ T_m より5～10℃下）のもと、高濃度（0.1～1.0 pmol/ml）のプローブを用いて行う。同書、p. 11.45。ミスマッチプローブ、縮重プローブのプールまたは「ゲスマー（guesmer）」を用いたハイブリダイゼーションの決定、ならびにハイブリダイゼーション溶液および実験的にハイブリダイゼーション条件を決定する方法は、当該技術分野でよく知られている。例えば、上記Ausubel (1999)、上記Sambrook (1989), pp. 11.45-11.57参照。

用語「消化」または「DNAの消化」は、DNA中のある配列にのみ作用する制限酵素によるDNAの酵素的開裂を指す。本明細書中で参照する種々の制限酵素は、市販されてお

り、その反応条件、コファクターおよび使用上の他の必要事項は知られており、当業者には日常的である。分析の目的では、典型的には、1 μ g のプラスミドまたはDNA断片が、約20 μ l の反応バッファー中の約2単位の酵素で消化される。プラスミドの構築のためにDNA断片を単離する目的では、典型的に5～50 μ g のDNAが、比例した大きな容積の20～250単位の酵素で消化される。特定の制限酵素のための適切なバッファーおよび基質量は、下記に参照するものなどの標準的な実験室マニュアルに記載されており、それらは販売者によって特定されている。37℃での約1時間のインキュベーション時間が通常用いられるが、条件は、標準的な手順、供給者の説明および反応の細目に応じて変わることがある。消化の後、反応を分析することができ、断片は、アガロースまたはポリアクリルアミドゲルを介した電気泳動により、当業者に日常的なよく知られた方法を用いて、精製することができる。

10

用語「連結」は、2または3以上のポリヌクレオチド（最も頻繁には2本鎖DNA）の間にホスホジエステル結合を形成する過程を指す。連結のための技術は当該技術分野でよく知られており、連結のためのプロトコルは標準的な実験室マニュアルおよび上記Sambrook（1989）などの参考文献に記載されている。

【0025】

ゲノム由来の「シングルエクソンプローブ」は、少なくとも1つのエクソン（「参照エクソン」）を含み、高ストリンジェンシーの条件下、参照エクソンを含む転写物由来の核酸に検出可能にハイブリダイズすることができるが、参照エクソンを欠く核酸とは高ストリンジェンシーの条件下で検出可能にハイブリダイズしないプローブである。シングルエクソンプローブは、典型的には、さらに、エクソン部分の最初の末端に隣接した、ゲノム中のエクソンに同様に隣接する第1のイントロン配列および／または遺伝子間配列を含み、ゲノム中のエクソンに同様に隣接する第2のイントロン配列および／または遺伝子間配列を含み得る。ゲノム由来のシングルエクソンプローブの最小長は、上記で論じたとおり、エクソン部分が、転写物由来核酸に高ストリンジェンシーの条件下でハイブリダイズするのに十分な長さであるための要件により決められる。ゲノム由来のシングルエクソンプローブの最大長は、プローブが、1エクソン以下の部分を含むための要件により決められる。シングルエクソンプローブは、ゲノム中のプローブ配列の残りに連続して見出されないプライミング配列であって、PCRおよび他の増幅ベースの技術に有用なプライミング配列を含んでもよい。

20

30

用語「マイクロアレイ」または「核酸マイクロアレイ」は、基体に結合した多数の核酸の集合を指し、結合した多数の核酸のそれぞれへのハイブリダイゼーションは、個別に検出可能である。基体は非孔質または多孔質でも、平坦または非平坦でも、ユニット型（unitary）または分散型（distributed）でもよい。マイクロアレイまたは核酸マイクロアレイは、Schena（ed.），DNA Microarrays：A Practical Approach（Practical Approach Series），Oxford University Press（1999）；Nature Genet. 21（1）（suppl.）：1－60（1999）；Schena（ed.），Microarray Biochip：Tools and Techniques，Eaton Publishing Company／BioTechniques Books Division（2000）でそのように呼ばれているすべてのデバイスを含む。これらのマイクロアレイは、基体に結合した多数の核酸の集合体であって、該多数の核酸が、とりわけBrenner et al.，Proc. Natl. Acad. Sci. USA 97（4）：1665－1670（2000）に記載のとおり、単一の平坦な基体上ではなく、むしろ多数のビーズ上に配置されているものを含む。

40

用語「変異した」は、核酸分子に適用される場合、核酸分子の核酸配列中のヌクレオチドが、参照核酸配列に比べ、挿入、欠失または変化を受け得ることを意味する。1つの座に単一の変化が生じててもよく（点突然変異）、または複数のヌクレオチドが、単一の座において挿入され、欠失しまたは変更されてもよい。さらに、1または2以上の変化が、核酸

50

配列中の任意の数の座で生じてよい。好ましい態様においては、核酸分子は、C S Pをコードする野生型の核酸配列を含むか、またはC S N Aである。核酸分子は、下記に記載の変異導入技術の方法を含む、当該技術分野で知られた任意の方法によって変異させることができる。

【0026】

用語「エラープローンPCR」は、DNAポリメラーゼのコピー信頼性が低い、例えばPCR産物の全長にわたり高率の点突然変異が得られるような条件下でPCRを行う方法を指す。例えば、Leung et al., Technique 1: 11-15 (1989) およ Caldwell et al., PCR Methods Applic. 2: 28-33 (1992) を参照。

10

用語「オリゴヌクレオチドを利用した突然変異導入」は、対象となる任意のクローンDNAセグメントにおける部位特異的突然変異の創出を可能にする方法を指す。例えば、Reidhaar-Olson et al., Science 241: 53-57 (1988) を参照。

用語「アセンブリPCR」は、小DNA断片の混合物からのPCR産物のアセンブリを伴う方法を指す。多数の異なるPCR反応が、同一のバイアル中で平行して生じ、1つの反応の産物が、他の反応の産物をプライミングする。

用語「セクシャルPCR変異導入」または「DNAシャッフリング」は、異なるが高度に関連したDNA配列のDNA分子同士のin vitroでの強制的な相同組換えと組合わされたエラープローンPCR法を指し、これは、配列類似性に基づくDNA分子のランダムな断片化と、それに引き続くエラープローンPCR反応におけるプライマー伸長による交差の固定によりもたらされる。例えばStemmer, Proc. Natl. Acad. Sci. U. S. A. 91: 10747-10751 (1994) を参照。DNAシャッフリングは、いくつかの関連遺伝子の間で行うことができる（「ファミリーシャッフリング」）。

20

用語「in vivo変異導入」は、対象となる任意のクローンDNAにおけるランダム突然変異を創出する方法を指し、これは、1または2以上のDNA修復経路に突然変異を有する大腸菌などの細菌株におけるDNAの増殖を伴う。これらの「ミューテーター」株は、野生型の親よりも高いランダム変異率を有する。ミューテーター株でのDNAの増殖は、最終的にDNA中にランダム変異を生じさせる。

30

用語「カセット変異導入」は、二本鎖DNA分子の小領域を、天然配列とは異なる合成オリゴヌクレオチド「カセット」で置き換える任意の方法を指す。オリゴヌクレオチドは、しばしば完全におよび／または部分的にランダム化された天然配列を含む。

【0027】

用語「リカーシブアンサンブル変異導入」は、フェノタイプが関連する変異体の多様な集団であって、その構成要素のアミノ酸配列が異なるものを作出するために開発された、タンパク質操作（タンパク質変異導入）用のアルゴリズムを指す。この方法は、連続する一連のコンビナトリアルカセット変異導入を制御するために、フィードバック機構を用いる。例えば、Arkin et al., Proc. Natl. Acad. Sci. U. S. A. 89: 7811-7815 (1992) を参照。

40

用語「エクスポネンシャルアンサンブル変異導入」は、ユニークで機能的な突然変異体の高いパーセンテージを有するコンビナトリアルライブラリーを創出するための方法を指し、ここでは、残基の小さな群が並行してランダム化され、そのそれぞれの変化位置において、機能的タンパク質をもたらすアミノ酸を同定する。例えば、Delegrave et al., Biotechnology Research 11: 1548-1552 (1993); Arnold, Current Opinion in Biotechnology 4: 450-455 (1993) を参照。上記に記載した各参考文献は、その全体が参照により本明細書に組み込まれる。

「作動可能に連結された」発現制御配列は、発現制御配列が、対象遺伝子を制御するために対象遺伝子に隣接している結合、ならびに、対象遺伝子を制御するためにtransで

50

、または、離れて作用する発現制御配列を指す。

本明細書中で用いる用語「発現制御配列」は、それが作動可能に連結しているコード配列の発現に影響するのに必要なポリヌクレオチド配列を指す。発現制御配列は、核酸配列の転写、転写後イベントおよび翻訳を制御する配列である。発現制御配列は、適切な転写開始配列、終止配列、プロモーター配列およびエンハンサー配列；スプライシングおよびポリアダニル化シグナルなどの有効なRNAプロセッシングシグナル；細胞質mRNAを安定化させる配列；翻訳効率を増強する配列（例えば、リボソーム結合部位）；タンパク質の安定性を増強する配列；および所望により、タンパク質分泌を増強する配列を含む。かかる制御配列の性質は、宿主生物に依存して異なる。原核生物においては、かかる制御配列は、一般的にプロモーター、リボソーム結合部位、および転写終止配列を含む。用語「制御配列」は、最小限、その存在が発現に必須なすべての要素を含むこととし、また、その存在が有利な追加の要素、例えば、リーダー配列および融合パートナー配列を含むことができる。

10

【0028】

本明細書中で用いる用語「ベクター」は、それが連結した他の核酸を輸送することが可能な核酸分子を指すものとする。ベクターの1つのタイプは「プラスミド」であり、これは、追加のDNAセグメントを連結することができる環状の二本鎖DNAループを指す。他のベクターは、コスミド、細菌人工染色体（BAC）および酵母人工染色体（YAC）を含む。他のタイプのベクターは、ウイルスベクターであり、ここでは、追加のDNAセグメントをウイルスゲノム中に連結することができる。細菌細胞に感染するウイルスベクターをバクテリオファージと呼ぶ。ある種のベクターは、導入された宿主細胞中で自律的に複製することが可能である（例えば、細菌由来の複製を有する細菌ベクター）。他のベクターは、宿主細胞への導入によって、宿主細胞のゲノムに組み込まれることができ、それによって宿主ゲノムとともに複製する。さらにまた、ある種のベクターは、それが作動可能に連結している遺伝子の発現を導くことができる。かかるベクターを本明細書中では、「組換え発現ベクター」（または単に「発現ベクター」）と呼ぶ。一般的に、組換えDNA技術で有用な発現ベクターは、しばしばプラスミドの形態である。本明細書中では、「プラスミド」および「ベクター」は、プラスミドが、最も多く用いられるベクター形態であることから、同じ意味に用いることができる。しかしながら、本発明は、同等の機能を果たす他の形態の発現ベクターを含むものとする。

20

30

本明細書中で用いる用語「組換え宿主細胞」（または単に「宿主細胞」）は、その中に発現ベクターが導入された細胞を指すものとする。かかる用語は、特定の対象細胞だけでなく、かかる細胞の子孫を指すものと理解されるべきである。突然変異かまたは環境的な影響によって、後続の世代にある種の変更が生じ得るため、かかる子孫は、実際のところ、親細胞と同一でない場合があり得るが、それでも依然として本明細書中で用いる用語「宿主細胞」の範囲に含まれる。

本明細書中で用いる熟語「オープンリーディングフレーム」およびその同等の略語である「ORF」は、その全体を連続したアミノ酸配列に翻訳することができる、転写物由来の核酸の部分の指す。そのように定義されているので、ORFは、ヌクレオチドで測った場合、正確に3で割り切れる長さを有する。そのように定義されているので、ORFは、天然タンパク質の全体をコードする必要はない。

40

本明細書中で用いられている熟語「ORFコードペプチド」は、ORFの予想されたまたは実際の翻訳物である。

本明細書中で用いる熟語である、参照核酸配列の「縮重変異体」は、標準的な遺伝子コードを用いて直接翻訳され、参照核酸配列から翻訳されたのと同じのアミノ酸配列を提供できる、すべての核酸配列を意味する。

【0029】

用語「ポリペプチド」は、天然に存在するおよび天然に存在しないタンパク質およびポリペプチド、ポリペプチド断片およびポリペプチドの突然変異体、誘導体および類縁体を包含する。ポリペプチドは、モノマーであってもポリマーであってもよい。さらに、ポリペ

50

プチドは単一のポリペプチド中に、そのそれぞれが1または2以上の異なる活性を有する、多数の異なるモジュールを含んでもよい。本発明の好ましいポリペプチドは、本発明の核酸分子によってコードされたCSP、ならびにその断片、突然変異体、類縁体および誘導体を含む。

用語「単離タンパク質」または「単離ポリペプチド」は、その起源または由来する給源に起因して、(1) ネイティブな状態でそれに随伴する天然に結びついた要素に結びついておらず、(2) 同一の種からの他のタンパク質を含まず、(3) 異なる種からの細胞によって発現されており、または(4) 天然に存在しない、タンパク質またはポリペプチドである。従って、化学的に合成された、またはそれが天然に由来する細胞とは異なる細胞系で合成されたポリペプチドは、その天然に結びついた要素から「単離」されている。ポリペプチドまたはタンパク質はまた、当該技術分野でよく知られているタンパク質精製技術を用いて、単離によって、天然に結びついた要素から実質的にフリーにすることができる。

タンパク質またはポリペプチドは、試料の少なくとも約60%~75%が、単一の種類のポリペプチドを示す場合は、「実質的に純粋」であり、「実質的に均一」であり、または「実質的に純化」されている。ポリペプチドまたはタンパク質は、単量体であっても多量体であってもよい。実質的に純粋なポリペプチドまたはタンパク質は、典型的には、約50%、60%、70%、80%または90% W/W、より普通には約95%のタンパク質試料を含み、そして好ましくは、99%を超えて純粋である。タンパク質の純度または均一性は、当該技術分野でよく知られている多くの手段、例えば、タンパク質試料のポリ

アクリルアミドゲル電気泳動と、それに引き続く当該技術分野でよく知られている染色でゲルを染色することによる、単一のポリペプチドバンドの視覚化などによって示すことができる。ある目的のためには、HPLCまたは精製のための当該技術分野でよく知られた他の手段を用いることによって、より高い解像度を提供することができる。本明細書中で用いる用語「ポリペプチド断片」は、全長ポリペプチドに比較して、アミノ末端および/またはカルボキシ末端の欠失を有する本発明のポリペプチドを指す。好ましい態様では、ポリペプチド断片は、断片のアミノ酸が、天然に存在する配列中の対応する位置と同一である、コンティグ配列である。断片は、典型的には、少なくとも5、6、7、8、9または10アミノ酸長、好ましくは少なくとも12、14、16または18アミノ酸長、より好ましくは少なくとも20アミノ酸長、より好ましくは少なくとも25、30、35、40または45アミノ酸、さらにより好ましくは少なくとも50または60アミノ酸長、そしてさらにより好ましくは少なくとも70アミノ酸長である。

【0030】

「誘導体」は、一次構造配列では実質的に類似しているが、例えばネイティブなポリペプチドでは見出されない *in vivo* または *in vitro* の化学的および生化学的修飾を含むポリペプチドまたはその断片を指す。かかる修飾としては、例えば、アセチル化、アシル化、ADP-リボシル化、アミド化、フラビンの共有結合的な付着、ヘム部分の共有結合的な付着、ヌクレオチドまたはヌクレオチド誘導体の共有結合的な付着、脂質または脂質誘導体の共有結合的な付着、ホスファチジルイノシトールの共有結合的な付着、架橋、環化、ジスルフィド結合形成、脱メチル化、共有結合架橋の形成、シスチンの形成、ピログルタメートの形成、ホルミル化、ガンマーカルボキシル化、グリコシル化、GPIアンカー形成、ヒドロキシル化、ヨード化、メチル化、ミリストイル化、酸化、タンパク質分解プロセッシング、リン酸化、プレニル化、ラセミ化、セレノイル化、硫化、アルギニン化などの転移RNA誘導性のタンパク質へのアミノ酸付加、およびユビキチン化が挙げられる。他の修飾としては、例えば、放射性核種による標識、および種々の酵素的修飾が挙げられ、これらは当業者によって容易に認識できる。ポリペプチドを標識するための多様な方法、および、かかる目的に有用な多様な置換基または標識は当該技術分野で知られており、¹²⁵I、³²P、³⁵S および ³H などの放射性同位体、標識化抗リガンド (antiligand) に結合するリガンド (例えば抗体)、フルオロフォア、化学発光剤、酵素および標識リガンドの特異的結合対の構成要素として機能することができる抗リ

ガンドを含む。標識の選択は、要求される感度、プライマーとの結合の容易性、安定性の要件、および利用可能な機材に依存する。ポリペプチドを標識するための方法は当該技術分野でよく知られている。本明細書中に参照により組み込まれる、上記 Ausubel (1992)、上記 Ausubel (1999) を参照。

【0031】

用語「融合タンパク質」は、異種のアミノ酸配列に結合したポリペプチドまたは断片を含む、本発明のポリペプチドを指す。融合タンパク質は、それらが、2または3以上の異なるタンパク質から2または3以上の所望の機能的要素を含むように構築することができるため、有用である。融合タンパク質は、対象となるポリペプチドからの少なくとも10の連続したアミノ酸、より好ましくは20または30のアミノ酸、さらにより好ましくは少なくとも40、50または60のアミノ酸、より一層好ましくは少なくとも75、100または125のアミノ酸を含む。融合タンパク質は、異なるタンパク質またはペプチドをコードする核酸配列とフレームになったポリペプチドまたはその断片をコードする核酸配列を構築し、その後融合タンパク質を発現することにより、組換え的に製造することができる。あるいは、融合タンパク質は、ポリペプチドまたはその断片を他のタンパク質と架橋することにより、化学的に製造することができる。

用語「類縁体」は、ポリペプチド類縁体および非ペプチド類縁体の両方を指す。本明細書中で用いる用語「ポリペプチド類縁体」は、アミノ酸配列の部分と実質的な同一性を有するが、非天然アミノ酸または非天然残基間結合を含む、少なくとも25アミノ酸のセグメントを含む、本発明のポリペプチドを指す。好ましい態様では、類縁体は、天然のポリペプチドと同じかまたは類似の生物活性を有する。典型的には、ポリペプチド類縁体は、天然に存在する配列に対して保存的アミノ酸置換（または挿入または欠失）を含む。類縁体は、典型的には、少なくとも20アミノ酸長、好ましくは少なくとも50アミノ酸長以上であり、しばしば天然に存在する全長ポリペプチドと同じ長さであることができる。

用語「非ペプチド類縁体」は、本発明の参照ポリペプチドと類似の特性の化合物を指す。非ペプチド化合物はまた、「ペプチドの模倣体」または「ペプチド模倣体」と呼ばれることもある。かかる化合物は、しばしばコンピュータ化された分子モデリングによって開発される。有用なペプチドと構造的に類似しているペプチドの模倣体は、同等の効果をもたらすために用いることができる。一般的に、ペプチド模倣体は、模範となる (paradigm) ポリペプチド（つまり、所望の生化学的特性または薬理学的活性を有するポリペプチド）と構造的に類似であるが、当該技術分野でよく知られている方法により、 $-CH_2NH-$ 、 $-CH_2S-$ 、 $-CH_2-CH_2-$ 、 $-CH=CH-$ （シスおよびトランス）、 $-COCH_2-$ 、 $-CH(OH)CH_2-$ 、および $-CH_2SO-$ からなる群から選択される結合で任意に置換されている、1または2以上のペプチド結合を有する。コンセンサス配列の1または2以上のアミノ酸の、同タイプのD-アミノ酸（例えば、L-リジンの代わりにD-リジン）での系統的な置換はまた、より安定なペプチドを創出するのに用いることもできる。さらに、コンセンサス配列または実質的に同一なコンセンサス配列の変種を含む制限された (constrained) ペプチドは、当該技術分野で知られた方法によって創出することができる（本明細書中に参照により組み込まれる Rizo et al., Ann. Rev. Biochem. 61:387-418 (1992)）。例えば、ペプチドを環化する分子内ジスルフィド架橋を形成することが可能な内部システイン残基を付加することができる。

【0032】

「ポリペプチド突然変異体」または「ミューテイン」は、配列が、天然または野生型タンパク質のアミノ酸配列に比べ、1または2以上のアミノ酸の置換、挿入または欠失を含む、本発明のポリペプチドを指す。ミューテインは、ある位置の単一のアミノ酸が別のアミノ酸に変わっている、1または2以上のアミノ酸点置換、1または2以上のアミノ酸が、天然に存在するタンパク質の配列中で、それぞれ挿入または欠失している、1または2以上の挿入および／または欠失、および／またはアミノ末端またはカルボキシ末端の一方または両方におけるアミノ酸配列の切断を有し得る。さらに、ミューテインは、天然に存在

するタンパク質と同一または異なる生物活性を有し得る。例えば、ミューテインは、増大したまたは低減した生物活性を有し得る。ミューテインは野生型タンパク質に対し少なくとも50%の配列類似性を有し、好ましいのは、60%の配列類似性、より好ましいのは70%の配列類似性である。さらにまた好ましいのは、野生型タンパク質に対し、80%、85%または90%の配列類似性を有するミューテインである。さらにより好ましい態様においては、ミューテインは95%の配列同一性を示し、さらにより好ましくは97%、さらにより好ましくは98%、さらにより好ましくは99%である。配列類似性は、GapまたはBest fitなどの任意の一般的な配列分析アルゴリズムによって測定することができる。

好ましいアミノ酸置換は：(1) タンパク質分解への感受性を減らすもの、(2) 酸化への感受性を減らすもの、(3) タンパク質複合体を形成するための結合親和性を変化させるもの、(4) 結合親和性または酵素活性を変化させるもの、および(5) かかる類縁体の別の物理化学的または機能的特性を付与しまたは変更するものである。例えば、単一または多数のアミノ酸置換(好ましくは保存的アミノ酸置換)が、天然に存在する配列(好ましくは、分子間接触を形成するドメインの外部のポリペプチド部分)になされ得る。好ましい態様では、アミノ酸置換は中程度の保存的置換または保存的置換である。より好ましい態様では、アミノ酸置換は保存的置換である。保存的アミノ酸置換は、親配列の構造的特徴を実質的に変えてはならない(例えば、代替りのアミノ酸は、親配列に存在するヘリックスを乱し、または、親配列を特徴付ける別のタイプの二次構造を乱す傾向があってはならない)。当該技術分野で認識されているポリペプチドの二次および三次構造の例は、そのそれぞれが参照によって本明細書に組み込まれる、Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden et al. (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton et al., Nature 354:105-106 (1991)に記載されている。

【0033】

本明細書で用いる場合、20種の通常アミノ酸およびその略称は、従来の用法に従う。参照によって本明細書に組み込まれるGolub et al. (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991)を参照。20種の通常アミノ酸の立体異性体(例えばD-アミノ酸)、 α -、 β -二置換アミノ酸、N-アルキルアミノ酸などの非天然アミノ酸、および他の非通常アミノ酸もまた、本発明のポリペプチドの好適な構成要素である。非通常アミノ酸の例としては、4-ヒドロキシプロリン、 γ -カルボキシグルタメート、 α -N, N-トリメチルリジン、 β -N-アセチルリジン、O-ホスホセリン、N-アセチルセリン、N-ホルミルメチオニン、3-メチルヒスチジン、5-ヒドロキシリジン、s-N-メチルアルギニン、および他の類似のアミノ酸およびイミノ酸(例えば4-ヒドロキシプロリン)が挙げられる。本明細書で用いられるポリペプチド表記法では、標準的な用法および慣習に従い、左手方向はアミノ末端方向であり、右手方向はカルボキシ末端方向である。

タンパク質は、コードされたタンパク質のアミノ酸配列が、異なる生物のタンパク質のコードされたアミノ酸配列と類似した配列を有し、類似の生物活性または機能を有する場合、別の生物のタンパク質に対し「相同性」を有し、または「相同」である。あるいは、タンパク質は、2つのタンパク質が類似のアミノ酸配列を有し類似の生物活性または機能を有する場合、他方のタンパク質に対し相同性を有しまたは相同である。2つのタンパク質は「相同」であるといわれるが、これは、タンパク質同士の進化論上の関連性を必ずしも暗示しない。その代わり、用語「相同」は、2つのタンパク質が類似のアミノ酸配列および類似の生物活性または機能を有することを意味すると定義される。好ましい態様では、相同なタンパク質は、野生型タンパク質に対し50%の配列類似性を示すものであり、好

ましいのは60%の配列類似性、より好ましいのは70%の配列類似性である。さらにより好ましいのは、野生型タンパク質に対し80%、85%または90%の配列類似性を示す相同タンパク質である。より一層好ましい態様では、相同タンパク質は95%、97%、98%または99%の配列類似性を示す。

【0034】

「配列類似性」が、タンパク質またはペプチドに関して用いられる場合、同一でない残基の位置は、しばしば保存的アミノ酸置換によって異なることが認められている。好ましい態様では、「配列類似性」を有するポリペプチドは、保存的または中程度に保存的なアミノ酸置換を含む。「保存的アミノ酸置換」は、アミノ酸残基が、類似の化学的特性（たとえば電荷または疎水性）の側鎖（R基）を有する他のアミノ酸残基で置換されたものである。一般的に、保存的アミノ酸置換は、タンパク質の機能的特性を実質的には変化させないだろう。2または3以上のアミノ酸配列が、保存的置換によって互いに異なる場合、置換の保存的性質について修正するために、配列同一性パーセントまたは類似の程度を、上方修正できる。この修正をマークする手段は、当業者によく知られている。本明細書に参照によって組み込まれる、例えばPearson, Methods Mol. Biol. 24: 307-31 (1994)を参照。

例えば、以下の6の群は、それぞれ互いに保存的置換であるアミノ酸を含んでいる：

- 1) セリン（S）、スレオニン（T）；
- 2) アスパラギン酸（D）、グルタミン酸（E）；
- 3) アスパラギン（N）、グルタミン（Q）；
- 4) アルギニン（R）、リジン（K）；
- 5) イソロイシン（I）、ロイシン（L）、メチオニン（M）、アラニン（A）、バリン（V）、および
- 6) フェニルアラニン（F）、チロシン（Y）、トリプトファン（W）。

あるいは、保存的置換は、本明細書に参照によって組み込まれるGonnet et al., Science 256: 1443-45 (1992)に開示されたPAM250対数尤度マトリクス中で正の値を有する任意の変化である。「中程度に保存的な」置換は、PAM250対数尤度マトリクス中で負でない値を有する任意の変化である。

配列同一性とも呼ばれる、ポリペプチドについての配列類似性は、典型的には、配列分析ソフトを用いて測定する。タンパク質分析ソフトは、保存的アミノ酸置換を含む、種々の置換、欠失および他の変更に割当てた類似性尺度を用いて、類似配列をマッチングする。例えば、GCGは、異なる種の生物からの相同ポリペプチドなどの関連性の深いポリペプチド同士、または野生型タンパク質とそのミューテインとの間の配列相同性または配列同一性を決定するために、デフォルトパラメータで用いることができる「Gap」および「Best fit」などのプログラムを備えている。例えば、GCG Version 6.1を参照。他のプログラムとしては、上記で論じたFASTAが挙げられる。

【0035】

本発明の配列を他の生物からの多数の配列を含むデータベースと比較する場合に好ましいアルゴリズムは、コンピュータプログラムBLAST、特にblastpまたはtblastnである。本明細書に参照により組み込まれる、例えば、Altschul et al., J. Mol. Biol. 215: 403-410 (1990)；Altschul et al., Nucleic Acids Res. 25: 3389-402 (1997)を参照。blastp用の好ましいパラメータは：

期待値：10（デフォルト）

フィルター：seg（デフォルト）

ギャップを開くための負担（cost）：11（デフォルト）

ギャップを広げるための負担：1（デフォルト）

最大アラインメント：100（デフォルト）

文字サイズ：11（デフォルト）

表示数（No. of descriptions）：100（デフォルト）

10

20

30

40

50

ペナルティマトリクス：B L O S U M 6 2

相同性のために比較されるポリペプチド配列の長さは、一般的には少なくとも約 16 アミノ酸残基、普通は少なくとも約 20 残基、より普通には少なくとも約 24 残基、典型的には少なくとも約 28 残基であり、そして好ましくは約 35 残基より多い。多数の異なる生物からの配列を含むデータベースを検索する場合、アミノ酸配列を比較するのが好ましい。

b l a s t p 以外のアルゴリズムによって測定することができるアミノ酸配列を用いたデータベース検索は、当該技術分野で知られている。例えば、ポリペプチド配列は、G C G V e r s i o n 6 . 1 中のプログラムである F A S T A を用いて比較することができる。F A S T A (例えば F A S T A 2 および F A S T A 3) は、照会配列と検索配列との間の最良のオーバーラップ領域のアラインメントと配列同一性パーセントを提供する。例えば、アミノ酸配列同士の配列同一性パーセントは、F A S T A を用いて、参照により本明細書に組み込む G C G V e r s i o n 6 . 1 で提供されるそのデフォルトまたは推奨パラメータ (文字サイズ 2 および P A M 2 5 0 スコアリングマトリクス) で決定することができる。

【0036】

「抗体」は、インタクトな免疫グロブリン、または、分子種、例えば本発明のポリペプチドへの特異的結合についてインタクトな抗体と競合するその抗原結合部分を指す。抗原結合部分は、組換え DNA 技術により、またはインタクト抗体の酵素的または化学的開裂により作製することができる。抗原結合部分は、とりわけ、F a b、F a b & # 0 ;、F (a b & # 0 ;)₂、F v、d A b および相補性決定領域断片、一本鎖抗体、キメラ抗体、ダイアボディ (d i a b o d y)、および、ポリペプチドに特異的抗原結合を付与するのに十分な免疫グロブリンの部分を少なくとも含むポリペプチドを含む。F a b 断片は、V L、V H、C L、および C H 1 ドメインからなる単価断片であり、F (a b ')₂ 断片は、ヒンジ領域のジスルフィド結合によって架橋した 2 個の F a b 断片を含む二価の断片であり、F d 断片は、V H および C H 1 ドメインからなり、F v 断片は、抗体の短腕の V L および V H ドメインからなり、そして d A b 断片は、V H ドメインからなる。例えば、W a r d e t a l . , N a t u r e 3 4 1 : 5 4 4 - 5 4 6 (1 9 8 9) を参照。本明細書中で、「特異的に結合する」および「特異的結合」とは、抗体および第一の分子種とともに混合されている他の分子種への結合に優先して、第一の分子種に結合する抗体の能力を意味する。抗体は、それが第一の分子種に特異的に結合する場合、特別に、第一の分子種を「認識する」という。

一本鎖抗体 (s c F v) は、V L および V H 領域が組になり、これらを単一のタンパク質鎖として作出されること可能にする合成リンカーを介して、単価分子を形成している抗体である。例えば、B i r d e t a l . , S c i e n c e 2 4 2 : 4 2 3 - 4 2 6 (1 9 8 8) ; H u s t o n e t a l . , P r o c . N a t l . A c a d . S c i . U S A 8 5 : 5 8 7 9 - 5 8 8 3 (1 9 8 8) を参照。ダイアボディは、二価の両特異的抗体であって、V H および V L ドメインが単一のポリペプチド鎖上に発現されているが、同一鎖上の 2 つのドメイン間の対合を許容するには短すぎるリンカーを用いており、これによって、該ドメインを他の鎖の相補ドメインと対合することを余儀なくさせ、2 つの抗原結合部位が創出されるものである。例えば、H o l l i g e r e t a l . , P r o c . N a t l . A c a d . S c i . U S A 9 0 : 6 4 4 4 - 6 4 4 8 (1 9 9 3) ; P o l j a k e t a l . , S t r u c t u r e 2 : 1 1 2 1 - 1 1 2 3 (1 9 9 4) を参照。

1 または 2 以上の C D R は、共有結合的に、または非共有結合的に分子中に組み込まれ、それを免疫付着因子 (i m m u n o a d h e s i n) にすることができる。免疫付着因子は、C D R をより大きなポリペプチド鎖の部分として組み込んでもよく、C D R を他のポリペプチド鎖に共有結合させてもよく、または、C D R を非共有結合的に組み込んでもよい。C D R により、免疫付着因子は、対象となる特定の抗原に特異的に結合することができる。キメラ抗体は、1 の抗体からの 1 または 2 以上の領域と 1 または 2 以上の他の抗体の 1 または 2 以上の領域とを含む抗体である。

【0037】

抗体は、1または2以上の結合部位を有することができる。1以上の結合部位がある場合、該結合部位は互いに同一でもよく、または異なっているもよい。例えば、天然に存在する免疫グロブリンは、2つの同一の結合部位を有し、一本鎖抗体またはFab断片は、1つの結合部位を有するが、「二重特異性」または「二重機能性」抗体は、2つの異なる結合部位を有する。

「単離抗体」は、(1)その天然の状態で随伴している、他の天然に結び付いた抗体を含む、天然に結び付いた構成要素と結び付いていない抗体、(2)同一の種からの他のタンパク質からフリーな抗体、(3)異なる種からの細胞によって発現されている抗体、または(4)天然に存在しない抗体である。精製抗体を含む精製タンパク質を、天然には結び付いていない構成要素によって安定化できることが知られている。天然には結び付いていない構成要素は、アルブミン(例えばBSA)などのタンパク質、またはポリエチレングリコール(PEG)などの化学物質であってもよい。

「中和抗体」または「阻害抗体」は、ポリペプチドの活性を阻害し、または通常それに結合するリガンドへのポリペプチドの結合を遮断する抗体である。「活性化抗体」は、ポリペプチドの活性を増大させる抗体である。

用語「エピトープ」は、免疫グロブリンまたはT細胞レセプターに特異的に結合することが可能な任意のタンパク質決定基を含む。エピトープ決定基は、通常アミノ酸または糖側鎖などの、化学的に活性な表面分子群からなり、通常は、特異的な三次元構造的な特性ならびに特異的な電荷特性を有する。抗体は、解離定数が $1\mu\text{M}$ より低く、好ましくは 100nM より低く、そして最も好ましくは 10nM より低い場合に抗原を特異的に結合するという。

本明細書中で用いる用語「患者」は、ヒトおよび動物対象を含む。

本明細書および特許請求の範囲を通して、単語「含む」、または「含み」または「含んでいる」などの変化形は、記載された整数または整数の群を包含することを意味するが、いずれの整数または整数の群の排除も意味しないと理解される。

用語「大腸特異的」は、体内の他の組織に比較して、大腸に優勢に発現されている核酸分子またはポリペプチドを指す。好ましい態様では、「大腸特異的」核酸分子またはポリペプチドは、体内のいずれの他の組織よりも5倍高いレベルで発現される。より好ましい態様では、「大腸特異的」核酸分子またはポリペプチドは、体内のいずれの他の組織よりも10倍高いレベルで、より好ましくは体内のいずれの他の組織よりも少なくとも15倍、20倍、25倍、50倍または100倍高く発現される。核酸分子レベルは、ノザンブロットハイブリダイゼーションなどの核酸ハイブリダイゼーション、または定量的PCRによって測定することができる。ポリペプチドレベルは、ウェスタンブロット分析などの、タンパク質レベルを正確に定量することで知られている任意の方法によって測定することができる。

【0038】

核酸分子、調節配列、ベクター、宿主細胞およびポリペプチド作製のための組換え的方法
核酸分子

本発明の1つの側面は、大腸または大腸細胞もしくは組織に特異的な単離核酸分子、またはかかる核酸分子に由来する単離核酸分子を提供する。これらの単離された大腸特異的核酸(CSNA)は、cDNA、ゲノムDNA、RNA、またはこれらの核酸のうちの1つの断片を含んでもよく、または、天然に存在しない核酸分子であってもよい。好ましい態様では、核酸分子は、大腸に特異的なポリペプチドである、大腸特異的ポリペプチド(CSP)をコードする。より好ましい態様では、核酸分子は、配列番号148~250のアミノ酸配列を含むポリペプチドをコードする。他の極めて好ましい態様では、核酸分子は、配列番号1~147の核酸配列の1つを含む。

CSNAは、ヒトまたは他の動物に由来してもよい。好ましい態様では、CSNAは、ヒトまたは他の哺乳類に由来する。より好ましい態様では、CSNAはヒトまたは他の霊長類に由来する。さらにより好ましい態様では、CSNAはヒトに由来する。

本発明に関して、「核酸分子」はまた、C S N Aをコードする核酸分子またはその相補体と選択的にハイブリダイズする核酸配列を含むものとする。ハイブリダイズする核酸分子は、ポリペプチドをコードしてもよく、もしくはしなくてもよく、またはC S Pをコードしなくてもよい。しかしながら、好ましい態様では、ハイブリダイズする核酸分子はC S Pをコードする。より好ましい態様では、本発明は、配列番号148～250のアミノ酸配列を含むポリペプチドをコードする核酸分子と選択的にハイブリダイズする核酸分子を提供する。さらにより好ましい態様では、本発明は、配列番号1～147のアミノ酸配列を含む核酸分子と選択的にハイブリダイズする核酸分子を提供する。

【0039】

好ましい態様では、核酸分子は、C S Pをコードする核酸分子と、低いストリンジェンシーの条件下で選択的にハイブリダイズする。より好ましい態様では、核酸分子は、C S Pをコードする核酸分子と、中程度のストリンジェンシーの条件下で選択的にハイブリダイズする。より好ましい態様では、核酸分子は、C S Pをコードする核酸分子と、高いストリンジェンシーの条件下で選択的にハイブリダイズする。さらにより好ましい態様では、核酸分子は、低い、中程度の、または高いストリンジェンシーの条件下で、配列番号148～250のアミノ酸配列を含むポリペプチドをコードする核酸分子とハイブリダイズする。より一層好ましい態様では、核酸分子は、C S Pをコードする核酸分子と、高いストリンジェンシーの条件下で選択的にハイブリダイズする。さらにより好ましい態様では、核酸分子は、低い、中程度の、または高いストリンジェンシーの条件下で、配列番号1～147の核酸配列の1つを含む核酸分子とハイブリダイズする。本発明の好ましい態様では、ハイブリダイズする核酸分子は、本発明のポリペプチドを組換え的に発現するために用いることができる。

本明細書中で用いる「核酸分子」は、C S Pをコードする核酸または該コードする核酸分子の相補体と実質的な配列類似性を示す配列を含むものとする。好ましい態様では、核酸分子は、ヒトC S Pをコードする核酸分子と実質的な配列類似性を示す。より好ましい態様では、核酸分子は、配列番号148～250のアミノ酸配列を有するポリペプチドをコードする核酸分子と実質的な配列類似性を示す。好ましい態様では、類似の核酸分子は、配列番号148～250のアミノ酸配列を有するポリペプチドなどのC S Pをコードする核酸分子と少なくとも60%の配列同一性を有するものであり、より好ましくは少なくとも70%、さらにより好ましくは少なくとも80%、そしてさらにより好ましくは少なくとも85%である。より好ましい態様では、類似の核酸分子は、C S Pをコードする核酸分子と少なくとも90%の配列同一性を有するものであり、より好ましくは少なくとも95%、より好ましくは少なくとも97%、さらにより好ましくは少なくとも98%、そしてなお一層好ましくは少なくとも99%である。他の極めて好ましい態様では、核酸分子は、C S Pをコードする核酸分子と少なくとも99.5%、99.6%、99.7%、99.8%または99.9%の配列同一性を有するものである。

【0040】

他の好ましい態様では、核酸分子は、C S N Aまたはその相補体と実質的な配列類似性を示す。より好ましい態様では、核酸分子は、配列番号1～147の核酸配列を含む核酸分子と実質的な配列類似性を示す。好ましい態様では、核酸分子は、C S N A、例えば配列番号1～147の核酸配列を有するものと少なくとも60%、より好ましくは少なくとも70%、さらにより好ましくは少なくとも80%、そして、さらにより好ましくは85%の配列同一性を有するものである。より好ましい態様では、核酸分子は、C S N Aと少なくとも90%、より好ましくは少なくとも85%、より好ましくは少なくとも97%、さらにより好ましくは少なくとも98%、そしてより一層好ましくは少なくとも99%の配列同一性を有するものである。別の極めて好ましい態様では、核酸分子は、少なくとも99.5%、99.6%、99.7%、99.8%または99.9% C S N Aとの配列同一性を有するものである。

実質的な配列類似性を示す核酸分子は、C S N AまたはC S Pをコードする核酸分子と、その全長にわたり配列同一性を示すものであってもよく、または、その全長の一部分のみ

にわたって類似するものであってもよい。この場合、該部分は、C S N AまたはC S Pをコードする核酸分子の少なくとも50ヌクレオチド、好ましくは少なくとも100ヌクレオチド、より好ましくは少なくとも150または200ヌクレオチド、さらにより好ましくは少なくとも250または300ヌクレオチド、より一層好ましくは少なくとも400または500ヌクレオチドである。

【0041】

実質的に類似の核酸分子は、他の種に由来する天然に存在するもの、とりわけ他の霊長類に由来するものであって、該類似の核酸分子が、配列番号148～250と顕著な配列同一性を示すアミノ酸配列をコードする、または、配列番号1～147のヌクレオチド配列と顕著な配列同一性を示すものである。この類似の核酸分子はまた、C S N Aが遺伝子ファミリーの成員である場合、ヒトからの天然に存在する核酸分子であってもよい。この類似の核酸分子はまた、家畜化種、例えばイヌ、ネコ、マウス、ラット、ウサギ、ハムスター、ウシ、ウマおよびブタ、および、例えばサル、キツネ、ライオン、トラ、クマ、キリン、シマウマなどの野生動物を含む、霊長類ではない哺乳類種に由来する天然に存在する核酸分子であってもよい。実質的に類似の核酸分子はまた、鳥類または爬虫類などの、非哺乳類種に由来する天然に存在する核酸分子であってもよい。天然に存在する実質的に類似の核酸分子は、ヒトまたは他の種から直接単離することができる。別の態様では、実質的に類似の核酸分子は、核酸分子のランダム突然変異によって実験的に製造されたものであってもよい。別の態様では、実質的に類似の核酸分子は、C S N Aの指向性(directed)突然変異によって実験的に作製されたものであってもよい。さらに、実質的に類似の核酸分子は、C S N Aであってもなくてもよい。しかしながら、好ましい態様では、実質的に類似の核酸分子はC S N Aである。

「核酸分子」は、C S N AまたはC S Pをコードする核酸の対立遺伝子変異体を含むものとする。例えば、真核生物のゲノムにおいては、一塩基多型(SNPs)がしばしば起こる。実際、1400万のSNPsがすでにヒトゲノムにおいて同定されている(International Human Genome Sequencing Consortium, Nature 409: 860-921 (2001))。従って、1つの種の1個体から決定された配列は、集団内に存在する他の対立遺伝子形態と異なり得る。さらに、微小な欠失および挿入は、一般的な集団において一塩基多型よりもむしろ稀ではなく、そしてしばしばタンパク質の機能を変えない。さらに、アミノ酸置換は、天然の対立遺伝子変異体の中で頻繁に生じ、しばしばタンパク質機能を実質的に変化させない。

好ましい態様では、対立遺伝子変異体を含む核酸分子は、遺伝子の変異体であって、該遺伝子がC S PをコードするmRNAに転写されているものである。より好ましい態様では、遺伝子は、配列番号148～250のアミノ酸配列を含むC S PをコードするmRNAに転写されている。別の好ましい態様では、対立遺伝子変異体は、遺伝子の変異体であって、遺伝子が、C S N AであるmRNAに転写されているものである。より好ましい態様では、遺伝子は、配列番号1～147の核酸配列を含むmRNAに転写されている。好ましい態様では、対立遺伝子変異体は、対象となる種の天然に存在する対立遺伝子変異体である。より好ましい態様では、対象となる種はヒトである。

【0042】

「核酸分子」はまた、本発明の核酸配列の部分を含むものとする。該部分は、ポリペプチドをコードしてもしていなくてもよく、そして、C S Pであるポリペプチドをコードしてもしなくてもよい。しかしながら、好ましい態様では、該部分はC S Pをコードする。1つの側面では、本発明は、C S N Aの部分を含む。第2の側面では、本発明は、C S N Aにハイブリダイスするまたは、C S N Aに対し実質的な配列類似性を示す核酸分子の部分を含む。第3の側面では、本発明は、C S N Aの対立遺伝子である核酸分子の部分を含む。第4の側面では、C S Pをコードする核酸分子の部分を含む。1つの部分は、少なくとも10ヌクレオチド、より好ましくは少なくとも15、17、18、20、25、30、35、40、50、60、70、80、90、100、150、200、250、300、350、400または500ヌクレオチドを含む。核酸の部分の最大サイズは、全長タ

10

20

30

40

50

ンパク質をコードする核酸分子の配列より1ヌクレオチド短いものである。

「核酸分子」はまた、下記の融合タンパク質、相同タンパク質、タンパク質断片、ミューテイン、またはポリペプチド類縁体をコードする配列を含むものとする。

本明細書記載の核酸のヌクレオチド配列は、少なくとも1つの酵素的重合反応（例えば、逆転写および／またはポリメラーゼ連鎖反応）から、自動シーケンサー（MegaBACE（登録商標）1000, Molecular Dynamics, Sunnyvale, CA, USAなど）を用いて、直接または間接にもたらされるDNA分子をシーケンシングして決定した。さらに、本発明のポリペプチドのすべてのアミノ酸配列は、他に特記のない限り、そのように決定された核酸配列からの翻訳によって予測した。

本発明の好ましい態様では、核酸分子は天然核酸分子の修飾を含む。これらの修飾は、非天然のヌクレオシド間結合、合成後修飾、または変化したヌクレオチド類縁体を含む。当業者は、行うことができる修飾のタイプが、核酸分子の意図された使用に依存することを認識している。例えば、核酸分子がハイブリダイゼーションプローブとして用いられる場合、かかる修飾は、もたらされる核酸の配列を識別する塩基対合を許容するものに限定される。in vitroまたはin vivoで、RNAまたはタンパク質の発現を導入するのに用いる場合、かかる修飾の範囲は、核酸が、重合基質として正しく機能することを許容するものに限定される。単離核酸が治療剤として用いられる場合、修飾は単離核酸に毒性を付与しないものに限定される。

【0043】

好ましい態様では、単離核酸分子は、放射性標識またはフルオロフォアなどの直接検出可能な標識を組み込んだヌクレオチド類縁体、またはビオチンまたは種々のハプテンなどの、その後の反応によって視覚化できる標識を組み込んだヌクレオチド類縁体を含むことができる。より好ましい態様では、標識核酸分子をハイブリダイゼーションプローブとして用いてもよい。

一般的な放射性標識類縁体は、 α - ^{32}P -dATP、 α - ^{32}P -dCTP、 α - ^{32}P -dGTP、 α - ^{32}P -dTTP、 α - ^{32}P -3'-dATP、 α - ^{32}P -ATP、 α - ^{32}P -CTP、 α - ^{32}P -GTP、 α - ^{32}P -UTP、 α - ^{35}S -dATP、 α - ^{35}S -GTP、 α - ^{33}P -dATPなどのように、 ^{33}P 、 ^{32}P および ^{35}S で標識されたものを含む。

本発明の核酸に容易に組み込まれる市販の蛍光ヌクレオチド類縁体は、Cy3-dCTP、Cy3-dUTP、Cy5-dCTP、Cy3-dUTP（Amersham Pharmacia Biotech, Piscataway, New Jersey, USA）、フルオレセイン-12-dUTP、テトラメチルローダミン-6-dUTP、テキサスレッド（登録商標）-5-dUTP、カスケードブルー（登録商標）-7-dUTP、BODIPY（登録商標）FL-14-dUTP、BODIPY（登録商標）TMR-14-dUTP、BODIPY（登録商標）TR-14-dUTP、ローダミングリーン（登録商標）-5-dUTP、オレゴングリーン（登録商標）488-5-dUTP、テキサスレッド（登録商標）-12-dUTP、BODIPY（登録商標）630/650-14-dUTP、BODIPY（登録商標）650/665-14-dUTP、アレクサフルオール（Alexa Fluor）（登録商標）488-5-dUTP、アレクサフルオール（登録商標）532-5-dUTP、アレクサフルオール（登録商標）568-5-dUTP、アレクサフルオール（登録商標）594-5-dUTP、アレクサフルオール（登録商標）546-14-dUTP、フルオレセイン-12-UTP、テトラメチルローダミン-6-UTP、テキサスレッド（登録商標）-5-UTP、カスケードブルー（登録商標）-7-UTP、BODIPY（登録商標）FL-14-UTP、BODIPY（登録商標）TMR-14-UTP、BODIPY（登録商標）TR-14-UTP、ローダミングリーン（登録商標）-5-UTP、アレクサフルオール（登録商標）488-5-UTP、アレクサフルオール（登録商標）546-14-UTP（Molecular Probes, Inc. Eugene, OR, USA）を含む。また、他のフルオロフォアを有するヌクレオチドをカスタム合成することもできる。開示がそ

の全体として参照により本明細書に組み込まれる Henegariu et al., Nature Biotechnol. 18: 345-348 (2000) を参照。
その後の標識のためにヌクレオチドに一般的に結合されるハプテンは、ビオチン（ビオチン-11-dUTP、Molecular Probes, Inc., Eugene, OR, USA、ビオチン-21-UTP、ビオチン-21-dUTP、Clontech Laboratories, Inc., Palo Alto, CA, USA）、ジゴキシゲニン（DIG-11-dUTP、アルカリ不安定、DIG-11-UTP、Roche Diagnostics Corp., Indianapolis, IN, USA）、およびジニトロフェニル（ジニトロフェニル-11-dUTP、Molecular Probes, Inc., Eugene, OR, USA）を含む。

10

【0044】

核酸分子は、核酸中への標識ヌクレオチド類縁体の組み込みによって標識できる。かかる類縁体は、DNA分子については、ニックトランスレーション、ランダムプライミング、ポリメラーゼ連鎖反応（PCR）、末端トランスフェラーゼターリング、およびオーバーハングのエンドフィリング、ならびに、RNA分子については、例えばT7、T3およびSP6などのファージプロモーターから誘導される *in vitro* での転写などによる、酵素重合によって組み込むことができる。このような各標識手法のために、市販のキットが容易に入手可能である。類縁体は、自動化された固相化学合成の最中に組み込むこともできる。標識はまた、核酸合成後に組み込むこともでき、5'リン酸塩および3'ヒドロキシルは検出可能な標識の合成後の共有結合による付着に好都合な部位を提供する。
他の合成後の手法はまた、核酸の内部標識を可能にする。例えば、DNA、RNAおよびPNA中のグアニン残基のN7（および、より少ない程度ではあるが、アデニン塩基）と反応するシスプラチン試薬を用い、核酸とフルオロフォア標識間に安定した配位複合体を提供することで、フルオロフォアを付着させることができる（Universal Linkage System）（Molecular Probes, Inc., Eugene, OR, USA および Amersham Pharmacia Biotech, Piscataway, NJ, USA から入手可能）。本明細書中に参照によって組み込まれる Alers et al., Genes, Chromosomes & Cancer 25: 301-305 (1999); Jelsma et al., J. NIH Res. 5: 82 (1994); Van Belkum et al., BioTechniques 16: 148-153 (1994) を参照。他の例としては、核酸は、アリールアザイド化学を用いて、標識核酸に光学的または熱的に結合しているジスルフィド含有リンカー（Fast Tag（登録商標）試薬、Vector Laboratories, Inc., Burlingame, CA, USA）を用いて標識することができる。還元後、遊離チオールが、ハプテン、フルオロフォア、糖、アフィニティリガンドまたは他のマーカーとの結合に利用できる。

20

30

1または2以上の独立のまたは相互作用性の標識を、本発明の核酸分子に組み込むことができる。たとえば、フルオロフォア、およびその近くでは蛍光を消光するように作用する部分の両方を組み込み、蛍光消光のリリースを介して特異的ハイブリダイゼーションを報告し、またはヌクレオチド外（*exonucleotidic*）除去を報告することができる。その開示の全体が参照により本明細書中に組み込まれる、例えば Tyagi et al., Nature Biotechnol. 14: 303-308 (1996); Tyagi et al., Nature Biotechnol. 16: 49-53 (1998); Sokol et al., Proc. Natl. Acad. Sci. USA 95: 11538-11543 (1998); Kostrikis et al., Science 279: 1228-1229 (1998); Marras et al., Genet. Anal. 14: 151-156 (1999); 米国特許第5,846,726号; 第5,925,517号; 第5,723,591号および第5,538,848号; Holland et al., Proc. Natl. Acad. Sci. USA 88: 7276-7280 (1991); Heid et

40

50

al., Genome Res. 6 (10): 986-94 (1996); Kuimelis et al., Nucleic Acids Symp. Ser. (37): 255-6 (1997) を参照。

【0045】

本発明の核酸分子は、1または2以上の天然のホスホジエステルヌクレオシド間結合を、よりヌクレアーゼ耐性のヌクレオシド間結合に変更することによって修飾することができる。その開示の全体が参照により本明細書中に組み込まれる、Hartmann et al. (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein et al. (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick et al. (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997) を参照。かかる変更されたヌクレオシド間結合は、しばしば、アンチセンス技術または標的遺伝子修正のために望まれる。その開示の全体が参照により本明細書中に組み込まれる、Gamper et al., Nucl. Acids Res. 28 (21): 4332-4339 (2000) を参照。

修飾オリゴヌクレオチド主鎖としては、限定されることなく、ホスホロチオエート、キラルホスホロチオエート、ホスホロジチオエート、ホスホトリエステル、アミノアルキルホスホトリエステル、3'-アルキレンホスホネートおよびキラルホスホネートを含むメチルホスホネートおよび他のアルキルホスホネート、ホスフィネート、3'-アミノホスホルアミデートおよびアミノアルキルホスホルアミデートを含むホスホルアミデート、チオノホスホルアミデート、チオノアルキルホスホネート、チオノアルキルホスホトリエステルならびにボラノホスフェートであって、通常の3'-5'結合を有するもの、これらの2'-5'結合類縁体、およびヌクレオシド単位の隣接した組が3'-5'から5'-3'または2'-5'から5'-2'に結合したものがあげられる。上記のリン含有結合の調製を教示する代表的な米国特許としては、限定されることなく、米国特許第3,687,808号、第4,469,863号、第4,476,301号、第5,023,243号、第5,177,196号、第5,188,897号、第5,264,423号、第5,276,019号、第5,278,302号、第5,286,717号、第5,321,131号、第5,399,676号、第5,405,939号、第5,453,496号、第5,455,233号、第5,466,677号、第5,476,925号、第5,519,126号、第5,536,821号、第5,541,306号、第5,550,111号、第5,563,253号、第5,571,799号、第5,587,361号、第5,625,050号が挙げられ、その開示の全体は、参照により本明細書中に組み込まれる。好ましい態様では、修飾ヌクレオシド間結合は、アンチセンス技術用に用いることができる。

【0046】

他の修飾オリゴヌクレオチド主鎖は、リン原子は含まないが、短鎖アルキルまたはシクロアルキルヌクレオシド間結合、混合されたヘテロ原子およびアルキルまたはシクロアルキルヌクレオシド間結合、または1または2以上の短鎖ヘテロ原子または複素環式ヌクレオシド間結合によって形成される主鎖を有する。これらは、モルホリノ結合（一部ヌクレオシドの糖部分から形成されている）を有するもの、シロキサン主鎖、スルフィド、スルホキシドおよびスルホン主鎖、ホルムアセチルおよびチオホルムアセチル主鎖、メチレンホルムアセチルおよびチオホルムアセチル主鎖、アルケン含有主鎖、スルファメート主鎖、メチレンイミノおよびメチレンヒドラジノ主鎖、スルホネートおよびスルホンアミド主鎖、アミド主鎖、および混合されたN、O、SおよびCH₂の構成部分を有するその他のものを含む。上記主鎖の調製を教示する代表的な米国特許としては、限定されることなく、米国特許第5,034,506号、第5,166,315号、第5,185,444号、

第5, 214, 134号、第5, 216, 141号、第5, 235, 033号、第5, 264, 562号、第5, 264, 564号、第5, 405, 938号、第5, 434, 257号、第5, 466, 677号、第5, 470, 967号、第5, 489, 677号、第5, 541, 307号、第5, 561, 225号、第5, 596, 086号、第5, 602, 240号、第5, 610, 289号、第5, 602, 240号、第5, 608, 046号、第5, 610, 289号、第5, 618, 704号、第5, 623, 070号、第5, 663, 312号、第5, 633, 360号、第5, 677, 437号および第5, 677, 439号が挙げられ、その開示の全体は、参照により本明細書中に組み込まれる。

他の好ましいオリゴヌクレオチド模倣体においては、糖およびヌクレオシド間結合の両方が、ペプチド核酸 (PNA) などの新規な基で置換されている。PNA化合物では、核酸のホスホジエステル主鎖は、アミド含有主鎖で、特にアミド結合で連結された反復するN-(2-アミノエチル)グリシンユニットによって置換されている。核酸塩基は、主鎖のアミド部分のアザ窒素原子に直接または間接的に、典型的には、メチレンカルボニル結合で結合している。PNAは、ペプチド合成プロトコル変法を用いて合成することができる。PNAオリゴマーは、FmocおよびtBoc法の両方によって合成することができる。PNA化合物の調製を教示する代表的な米国特許としては、限定されることなく、米国特許第5, 539, 082、5, 714, 331および5, 719, 262号が挙げられ、その開示の全体は、参照により本明細書中に組み込まれる。自動PNA合成は、市販の合成機で容易に達成できる (例えば、"PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA参照)。

【0047】

PNA分子は、多数の理由で有利である。第1に、PNA主鎖は非荷電なので、PNA/DNAおよびPNA/RNA 2量体は、DNA/DNAおよびDNA/RNA 2量体でみられるより、高い熱安定性を有する。PNA/DNAまたはPNA/RNA 2量体のT_mは、一般的に、対応するDNA/DNAまたはDNA/RNA 2量体のT_mよりも1塩基対当たり1℃高い (100mMのNaCl中で)。第2に、PNA分子はまた、DNA/DNA 2量体形成が起こらない条件下、低いイオン強度で、安定なPNA/DNA複合体を形成することができる。第3に、PNAはまた、PNA/DNAのミスマッチが、DNA/DNAのミスマッチより不安定であるため、相補DNAへの結合における高い特異性を示す。混合したPNA/DNAの15マーにおける単一のミスマッチは、T_mを8~20℃ (平均15℃) 下げる。対応するDNA/DNA 2量体では、単一のミスマッチは、T_mを4~16℃ (平均11℃) 下げる。PNAプローブは、DNAプローブより顕著に短くすることができるので、その特異性は高い。第4に、ヌクレアーゼおよびプロテアーゼは核酸塩基側鎖を有するPNAポリアミド主鎖を認識しないので、PNAオリゴマーは、酵素による分解に抵抗性であり、これらの化合物の半減期は*in vivo*および*in vitro*で長くなる。その開示の全体が参照によって本明細書に組み込まれる、例えば、Ray et al., FASEB J. 14 (9): 1041-60 (2000); Nielsen et al., Pharmacol Toxicol. 86 (1): 3-7 (2000); Larsen et al., Biochim Biophys Acta. 1489 (1): 159-66 (1999); Nielsen, Curr. Opin. Struct. Biol. 9 (3): 353-7 (1999), and Nielsen, Curr. Opin. Biotechnol. 10 (1): 71-5 (1999)を参照。

核酸分子は、核酸分子の全長にわたってその天然の構造に比べて修飾されていてもよく、または、その不連続な部分に限局されていてもよい。後者の例としては、不連続なDNAおよびRNAドメインを有し、標的遺伝子修復およびPCR反応変法に用いることができるキメラ核酸を合成することができ、これは、その開示の全体が参照によって本明細書中

に組み込まれる米国特許第 5, 760, 012 号および第 5, 731, 181 号、Misra et al., Biochem. 37: 1917-1925 (1998) および Finn et al., Nucl. Acids Res. 24: 3357-3363 (1996) にさらに記載されている。

他に特記のない限り、本発明の核酸は、所望の用途に適切な任意の位相構造を含むことができる。この用語は、従って、とりわけ、1 本鎖、2 本鎖、3 量体、4 量体、部分的 2 本鎖、部分的 3 量体、部分的 4 量体、分枝状、ヘアピン状、環状およびパドロック状構造を明白に包含する。パドロック構造およびその有用性は、その開示の全体が参照によって本明細書中に組み込まれる Baner et al., Curr. Opin. Biotechnol. 12: 11-15 (2001); Escude et al., Proc. Natl. Acad. Sci. USA 14: 96 (19): 10603-7 (1999); Nilsson et al., Science 265 (5181): 2085-8 (1994) にさらに記載されている。3 量体および 4 量体構造、およびその有用性は、その開示の全体が参照によって本明細書中に組み込まれる Praseuth et al., Biochim. Biophys. Acta. 1489 (1): 181-206 (1999); Fox, Curr. Med. Chem. 7 (1): 17-37 (2000); Kochetkova et al., Methods Mol. Biol. 130: 189-201 (2000); Chan et al., J. Mol. Med. 75 (4): 267-82 (1997) に総括されている。

10

20

【0048】

核酸分子をプローブおよびプライマーとして用いる方法

本発明の単離核酸分子は、ゲノムおよび転写物由来核酸試料の両方において、ハイブリダイズする核酸を検出、特徴付けおよび定量し、そこからハイブリダイズする核酸を単離するためのハイブリダイゼーションプローブとして用いることができる。溶液中で遊離している場合、かかるプローブは、典型的には、常にではないが、検出可能に標識されている。マイクロアレイにおけるように、基体に結合している場合、かかるプローブは、典型的には、常にではないが、標識されていない。

1 つの態様では、本発明の単離核酸は、C S N A の遺伝子中の総変化、例えば欠失、挿入、転座および C S N A 遺伝子座の重複を、染色体のスプレッド (chromosome spreads) に対する蛍光ハイブリダイゼーション (FISH) を介して、検出および特徴付けするためのプローブとして用いることができる。その開示の全体が参照によって本明細書中に組み込まれる、例えば Andreeff et al. (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999) を参照。本発明の単離核酸は、例えば制限断片長多型のサザンブロットによる検出を用いて、微小なゲノムの変化を評価するためのプローブとして用いることができる。本発明の単離核酸分子は、本発明の核酸分子を含むゲノムクローンを単離するためのプローブとして用いることができ、それをその後配列レベルで欠失、挿入、転座および置換 (1 塩基多型、SNPs) を同定するために制限マッピングおよびシーケンシングすることができる。

30

40

他の態様では、本発明の単離核酸分子は、転写由来核酸試料において C S N A を検出、特徴付けおよび定量し、およびそこから C S N A を単離するためのプローブとして用いることができる。1 つの側面では、本発明の単離核酸分子は、全 RNA 試料またはポリ A + 選択 RNA 試料のノザンブロットによって、mRNA を検出し、長さによって特徴付けし、そして定量するためのハイブリダイゼーションプローブとして用いることができる。他の側面では、本発明の単離核酸分子は、組織切片に対する in situ ハイブリダイゼーションによって、mRNA を検出し、位置によって特徴付けし、そして定量化するためのハイブリダイゼーションプローブとして用いることができる。その開示の全体が参照によって本明細書に組み込まれる、例えば, Schwarczacher et al., In Situ Hybridization, Springer-Verlag Ne

50

w York (2000)を参照。他の好ましい態様では、本発明の単離核酸分子は、cDNAライブラリー中のクローンの発現を測定するための、またはcDNAライブラリーからハイブリダイズする核酸分子を単離するためのハイブリダイゼーションプローブとして用いることができ、これにより、限定されることなく、欠失、挿入、置換、切断、代替的なスプライス形態および1塩基多型の同定を含む、CSNAにハイブリダイズするmRNAの配列レベルでの特徴付けが可能になる。さらに他の好ましい態様では、本発明の核酸分子は、マイクロアレイに用いることができる。

【0049】

上述のすべてのプローブ技術は、当業者の間で好適であり、その開示の全体が参照によって本明細書に組み込まれる、Sambrook (2001), *supra*; Ausubel (1999), *supra*; および Walker et al. (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000)などの標準的なテキストに、より詳細に記載されている。

従って、1つの態様では、本発明の核酸分子は、本発明の核酸分子に選択的にハイブリダイズする第2の核酸分子を同定または増幅するためのプローブまたはプライマーとして用いることができる。好ましい態様では、プローブまたはプライマーはCSPをコードする核酸分子に由来する。より好ましい態様では、プローブまたはプライマーは、配列番号148～250のアミノ酸配列を有するポリペプチドをコードする核酸分子に由来する。他の好ましい態様では、プローブまたはプライマーは、CSNAに由来する。より好ましい態様では、プローブまたはプライマーは、配列番号1～147の核酸配列の1つを有する核酸分子に由来する。

一般的に、プローブまたはプライマーは、少なくとも10ヌクレオチド長、より好ましくは少なくとも12、より好ましくは少なくとも14、そしてより一層好ましくは少なくとも16または17ヌクレオチド長である。より一層好ましい態様では、プローブまたはプライマーは、少なくとも18ヌクレオチド長、より一層好ましくは少なくとも20ヌクレオチド、そしてより一層好ましくは少なくとも22ヌクレオチド長である。プライマーおよびプローブはまた、長さにおいてより長くてもよい。例えば、プローブまたはプライマーは、25ヌクレオチド長、または30、40または50ヌクレオチド長であってもよい。オリゴヌクレオチドプローブを用いて核酸ハイブリダイゼーションを達成する方法は、当該技術分野でよく知られている。例えば、上記Sambrook et al., 1989中、短いプローブの放射性標識を説明した第11章、およびpp. 11.31-11.32および11.40-11.44、ならびにプローブハイブリダイゼーションのための具体的な条件(pp. 11.50-11.51)を含む、オリゴヌクレオチドプローブのハイブリダイゼーション条件を説明したpp. 11.45-11.53を参照。

【0050】

プライマーを用いた増幅を達成する方法もまた、当該技術分野でよく知られている。ポリメラーゼ連鎖反応(PCR)を行う方法は、とりわけ、その開示の全体が参照によって本明細書に組み込まれる、McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis et al. (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand et al. (eds.), PCR Strategies, Academic Press (1998); Newton et al., PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson et al. (eds.), PCR 2: A Prac

tical Approach, Oxford University Press, Inc. (1995) にまとめてある。RT-PCRを行うための方法は、例えば、その開示の全体が参照によって本明細書に組み込まれる Siebert et al. (ed.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/Bio Techniques Books (1995) に集められている。

PCRおよびハイブリダイゼーション法は、本発明の核酸分子の対立遺伝子変異体、相同核酸分子および断片を同定および／または単離するのに用いることができる。PCRおよびハイブリダイゼーション法はまた、本発明の相同タンパク質、類縁体、融合タンパク質またはミューテインをコードする核酸分子を同定、増幅および／または単離するために用いることができる。本発明の核酸プライマーは、転写由来またはゲノムDNAをテンプレートとして用い、本発明の核酸分子の増幅をプライミングするのに用いることができる。本発明の核酸プライマーはまた、例えば、SNP検出のための1塩基伸長(SBE)をプライミングするのに用いることができる(例えば、その開示の全体が参照によって本明細書に組み込まれる米国特許第6,004,744号を参照)。

ローリングサークル増幅などの等温増幅法もまた、現在ではよく記載されている。例えば、その開示の全体が参照によって本明細書に組み込まれる Schweitzer et al., Curr. Opin. Biotechnol. 12(1): 21-7 (2001)、米国特許第5,854,033号および第5,714,320号、ならびに国際公開WO 97/19193およびWO 00/15779を参照。ローリングサークル増幅は、SNP検出を容易にするために、他の技術と組み合わせることができる。例えば、Lizardi et al., Nature Genet. 19(3): 225-32 (1998) を参照。

【0051】

本発明の核酸分子は、基体と共有結合で、または非共有結合で結合していてもよい。基体は非孔質または多孔質でも、平坦または非平坦でも、ユニット型または分散型でもよい。結合した核酸分子は、ハイブリダイゼーションプローブとして用いることができ、標識されていても非標識でもよい。好ましい態様では、結合した核酸分子は非標識である。

1つの態様では、本発明の核酸分子は、多孔質の基体、例えば、典型的には、ニトロセルロース、ナイロンまたは正電荷誘導体化ナイロン(positively-charged derivatized nylon)を含む膜に結合している。本発明の核酸分子は、標識された核酸試料、例えば、転写物由来の核酸試料中に存在するハイブリダイズする核酸分子を検出するのに用いることができる。他の態様では、核酸分子は、限定されることなくガラス、アモルファスシリコン、結晶シリコンまたはプラスチックを含む、固体基体に結合している。プラスチックの例としては、限定されることなく、ポリメチルアクリル、ポリエチレン、ポリプロピレン、ポリアクリレート、ポリメチルメタクリレート、ポリ塩化ビニル、ポリテトラフルオロエチレン、ポリスチレン、ポリカーボネート、ポリアセタール、ポリスルホン、酢酸セルロース、硝酸セルロース、ニトロセルロース、またはこれらの混合物が挙げられる。固体基体は、長方形、円盤状および球状を含む、任意の形態であることができる。

本発明の核酸分子は、支持基体の表面に共有結合で付着させ、または、予想される非共有結合相互作用またはその組み合わせによって、変性および接着を容易にするカオトロピック剤中で、誘導体化した表面に適用することができる。本発明の核酸分子は、他の核酸が競合的に結合した基体に結合することができ、多数の結合核酸のそれぞれへのハイブリダイゼーションを個別に検出できる。低い密度、例えば、多孔質の膜では、これらの基体に結合した集合体は、典型的にはマイクロアレイと呼ばれる。高い密度、典型的には、ガラスなどの固体支持体では、これらの基体に結合した多数の核酸の集合体は、口語的にマイクロアレイと呼ばれる。本明細書で用いる場合、用語マイクロアレイは、すべての密度の

アレイを含む。従って、本発明の他の側面は、本発明の核酸を含むマイクロアレイを提供することである。

【0052】

発現ベクター、宿主細胞およびポリペプチドを作製するための組換え的方法

本発明の他の側面は、1または2以上の本発明の単離核酸分子を含むベクター、およびかかるベクターが導入された宿主細胞に関する。

ベクターは、とりわけ、本発明の核酸を宿主細胞内で増殖するため（クローニングベクター）、本発明の核酸を異種生物に由来する宿主細胞間で受け渡すため（シャトルベクター）、本発明の核酸を宿主細胞の染色体中に挿入するため（挿入ベクター）、本発明の核酸のセンスまたはアンチセンスRNA転写物を *in vitro* または宿主細胞内で発現させるため、および本発明の核酸によってコードされるポリペプチドを、単独で、またはポリペプチドとの融合体として発現させるため（発現ベクター）に用いることができる。本発明のベクターは、しばしば数種のかかる使用に好適だろう。

ベクターはこれまでに当該技術分野でよく知られており、とりわけ、その開示の全体が参照によって本明細書に組み込まれる Jones et al. (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones et al. (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa et al., Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); 上記 Sambrook (2001); 上記 Ausubel (1999) に記載されている。さらにまた、膨大な種類のベクターが市販されている。既存のベクターの使用およびその改変は当業者の間で好適であり、本明細書では、基本的な特徴しか説明する必要はない。

核酸配列は、それを適切な発現ベクター中の発現制御配列に作動可能に連結し、この発現ベクターを、適切な単細胞宿主を形質転換するために用いることによって、発現させることができる。発現制御配列は、核酸配列の転写、転写後イベントおよび翻訳を制御する配列である。この発明の核酸配列の、発現制御配列へのかかる作動可能な連結は、勿論、すでに該核酸配列の一部とはなっていない場合は、翻訳開始コドン、ATGまたはGTGの、該核酸配列の上流の正しいリーディングフレームにおける提供を含む。

多様な宿主／発現ベクターの組み合わせを、この発明の核酸配列の発現に用いることができる。有用な発現ベクターは、例えば、染色体性、非染色体性、および合成核酸配列のセグメントからなってもよい。

【0053】

1つの態様では、原核細胞を、適切なベクターとともに用いることができる。原核宿主細胞は、しばしば、クローニングおよび発現に用いられる。好ましい態様では、原核宿主細胞は、大腸菌、シュードモナス属、バチルス属およびストレプトマイセス属を含む。好ましい態様では、本発明の核酸分子を発現させるために、細菌性宿主細胞を用いる。細菌性宿主用の有用な発現ベクターは、大腸菌、バチルス属またはストレプトマイセス属からのものなどの、細菌性プラスミド、例えば、p Bluescript、p GEX-2T、p UCベクター、colE1、p CR1、p BR322、p MB9およびそれらの派生物、広宿主域プラスミド、例えば RP4、ファージDNA、例えば、NM989、λGT10およびλGT11などのラムダファージの多数の派生物、および例えばM13および線状1本鎖ファージDNAなどを含む。大腸菌が宿主として用いられる場合、選択マーカーは、同様に、グラム陰性細菌での選択性について選ばれる。例えば、典型的なマーカーは、アンピシリン、テトラサイクリン、クロラムフェニコール、カナマイシン、ストレプト

10

20

30

40

50

マイシンおよびゼオシンなどの抗生剤への耐性を付与する。栄養要求性のマーカーもまた用いることができる。

他の態様では、酵母、昆虫、哺乳類または植物細胞などの真核宿主細胞を用いることができる。酵母細胞、典型的には S. cerevisiae は、相同性組換えによる遺伝子変化のターゲティングの容易性、および組換え的に発現したタンパク質を用いて遺伝的欠陥を容易に補充する能力のために、真核生物の遺伝子研究に有用である。酵母細胞は、相互作用するタンパク質成分を、例えば、ツーハイブリッドシステムの使用を介して、同定するのに有用である。好ましい態様では、酵母細胞は、タンパク質発現に有用である。酵母に用いるための本発明のベクターは、常にではないが、典型的には、酵母で用いるのに好適な複製起点、および酵母で機能する選択マーカーを含む。酵母ベクターは、酵母組み込みプラスミド（例えば YIp5）および酵母複製プラスミド（YRp および YE p 型プラスミド）、酵母中心体プラスミド（YCp 型プラスミド）、Ylp と呼ばれる酵母線状プラスミド、pGPD-2、2 μ プラスミドおよびこれらの派生物、および Gietz et al., Gene, 74: 527-34 (1988) に記載されているような（YIplac、YEplac および YCplac）改善されたシャトルベクターを含む。酵母ベクターにおける選択マーカーは、種々の栄養要求性マーカーを含み、その最も一般的なものは、（Saccharomyces cerevisiae において）URA3、HIS3、LEU2、TRP1 および LYS2 であり、これは、ura3-52、his3-D1、leu2-D1、trp1-D1 および lys2-201 などの特異的な栄養要求性の突然変異を補充する。

10

20

【0054】

昆虫細胞は、しばしば高効率のタンパク質発現用に使われる。宿主細胞が Spodoptera frugiperda からのもの、例えば、Sf9 および Sf21 細胞系、および expresSF（登録商標）細胞（Protein Sciences Corp., Meriden, CT, USA）である場合、ベクター複製方法は、典型的にはバキュロウイルスの生活環に基づく。典型的には、バキュロウイルストランスファーベクターは、野生型 AcMNPV ポリヘドリン遺伝子を、対象となる異種遺伝子で置換するために用いられる。野生型ゲノムにおいてポリヘドリン遺伝子に隣接する配列は、トランスファーベクターの発現カセットの 5' および 3' に位置している。AcMNPV DNA でのトランスフェクションの後、相同性組換えイベントがこれらの配列間で生じ、対象となる遺伝子およびポリヘドリンまたは p10 プロモータを担持する組換えウイルスがもたらされる。選択は、lacZ 融合活性の視覚的なスクリーニングに基づくことができる。他の態様では、宿主細胞は哺乳類細胞であってもよく、これらは、薬剤として意図されているタンパク質の発現のため、およびタンパク質または生理学的経路の潜在的なアゴニストおよびアンタゴニストのスクリーニングのために特に有用である。自律的な染色体外での複製を対象とする哺乳類ベクターは、典型的には、ウイルス性の起源、例えば、SV40 起源（広い T 抗原を発現している、例えば COS1 および COS7 細胞などの細胞系での複製用）、パピローマウイルス起源、または長期エピソーム性の複製のための EBV 起源（例えば、EBV EBNA-1 遺伝子産物およびアデノウイルス E1A を構成的に発現している 293-EBNA 細胞で用いるため）を含むだろう。組込み、および従って哺乳類染色体の部分としての複製を対象とするベクターは、SV40 起源などの、哺乳類細胞で機能的な複製起点を含むことができるが、必須ではない。アデノウイルス、アデノ随伴ウイルス、ワクシニアウイルス、および種々の哺乳類レトロウイルスなどのウイルスに基づくベクターは、典型的には、ウイルスの複製方法に従って複製するだろう。哺乳類細胞で用いるための選択マーカーは、ネオマイシン、ブラスチシジン、ヒグロマイシンおよびゼオシンへの耐性、および HAT 培地を用いたプリンサルベージ経路に基づく選択を含む。

30

40

哺乳類細胞での発現は、種々のプラスミド、例えば pSV2、pBC12BI および p91023 など、ならびに溶菌性ウイルスベクター（例えば、ワクシニアウイルス、アデノウイルスおよびバキュロウイルス）、エピソームウイルスベクター（例えば、ウシパピロ

50

ーマウイルス)、およびレトロウイルスベクター(例えば、マウスのレトロウイルス)を用いて達成することができる。昆虫細胞に有用なベクターは、バキュロウイルスベクターおよび pVL 941 を含む。

植物細胞もまた、典型的には植物ウイルス(例えば、カリフラワーモザイクウイルス、CaMV、タバコモザイクウイルス、TMV)由来のベクターレプリコン、および植物における好適性について選ばれた選択マーカーによる発現に用いることができる。

【0055】

異なる宿主細胞のコドンの使用は異なり得ることが知られている。例えば、植物細胞およびヒト細胞は、特定のアミノ酸をコードするためのコドンの選好において異なり得る。結果として、ヒト mRNA は、植物、細菌または昆虫宿主細胞に、有効に翻訳されないことがある。従って、本発明の他の態様は、コドンの最適化を対象とする。本発明の核酸分子のコドンは、該核酸分子でコードされたアミノ酸配列を変化させることなく、可能な限り、宿主細胞内に天然に含まれている遺伝子に類似するように、変更することができる。本発明の DNA 配列を発現させるために、多種多様な発現制御配列のいずれをも、これらのベクターに使用することができる。かかる有益な発現制御配列は、上述の発現ベクターの構造遺伝子に結びついた発現制御配列を含む。転写を制御する発現制御配列は、例えば、プロモーター、エンハンサーおよび転写終止部位を含む。真核細胞において、転写後イベントを制御する発現制御配列は、スプライシング供与および受容部位、および転写された RNA の半減期を変更する配列、例えば、ポリ(A)付加を導く配列または RNA 結合タンパク質のための結合部位を含む。翻訳を制御する発現制御配列は、リボソーム結合部位、特定の細胞区画への、またはその内部でのポリペプチドの標的化された発現を導く配列、および翻訳の速度または効率を変更する 5' および 3' 非翻訳領域中の配列を含む。原核生物、例えば大腸菌のための有用な発現制御配列の例は、プロモーター、しばしば、ファージプロモーター、例えば、ラムダファージ pL プロモーター、trc プロモーター、trp および lac プロモーターに由来するハイブリッド、バクテリオファージ T7 プロモーター(T7 ポリメラーゼを発現するために操作された大腸菌において)、TAC または TRC システム、ラムダファージの主要オペレーターおよびプロモーター領域、fd コートタンパク質の制御領域、または araBAD オペロンなどを含むだろう。原核生物発現ベクターは、ターミネーターなどの転写ターミネーター、コンセンサスリボソーム結合部位および翻訳終止コドンなどの、翻訳を促進する要素をさらに含み得る (Schom 30
er et al., Proc. Natl. Acad. Sci. USA 83: 85
06-8510 (1986))。

【0056】

酵母細胞、典型的には *S. cerevisiae* のための発現制御配列は、CYC1 プロモーター、GAL1 プロモーター、GAL10 プロモーター、ADH1 プロモーター、酵母 α 接合系のプロモーターまたは GPD プロモーターなどの酵母プロモーターを含み、典型的には、CYC1 または ADH1 遺伝子からの転写終止シグナルなどの、転写の終止を促進する要素を有する。

哺乳類細胞におけるタンパク質発現に有用な発現ベクターは、哺乳類細胞中で活性のあるプロモーターを含む。これらのプロモーターは、哺乳類ウイルスに由来するもの、例えば、ヒトサイトメガロウイルス(CMV)の前初期遺伝子からのエンハンサープロモーター配列、ラウス肉腫ウイルスの長い末端反復配列(RSVLTR)からのエンハンサープロモーター配列、SV40 からのエンハンサープロモーター、またはアデノウイルスの初期および後期プロモーターなどを含む。他の発現制御配列は、3-ホスホグリセレートキナーゼまたは他の解糖酵素のためのプロモーター、酸性ホスファターゼのプロモーターを含む。他の発現制御配列は、対象となる CSNA を含む遺伝子からのものを含む。しばしば、後期 SV40 ポリアデニル化部位、およびウシ成長ホルモン(BGH)遺伝子からのポリアデニル化シグナルおよび転写終止配列などのポリアデニル化部位、ならびにリボソーム結合部位の組込みによって増強される。さらにまた、ベクターはウサギ β -グロビン遺伝子のイントロン II などのイントロン、および SV40 スプライスエレメント 40
50

を含むことができる。

好ましい核酸ベクターはまた、選択的または増幅可能なマーカー遺伝子、および対象遺伝子のコピー数を増幅するための手段を含む。かかるマーカー遺伝子は、当該技術分野でよく知られている。核酸ベクターは、安定化配列（例えば、oriまたはARS様配列およびテロメア様配列）をまた含んでもよく、または、代替的に宿主細胞ゲノムへの指向性または非指向性組込み（directed or non-directed integration）を促進するように設計してもよい。好ましい態様では、この発明の核酸配列は、対象となるコードされた核酸配列を含むタンパク質をコードするRNAの高レベルの発現を可能にする発現ベクター中のフレームに挿入される。核酸クローニングおよびシーケンシング方法は、当業者によく知られており、上記Sambrook（1989）
10、上記Sambrook（2000）、および上記Ausubel（1992）、上記Ausubel（1999）を含む、各種の実験室マニュアルに記載されている。生物学的試薬、化学試薬および免疫学的試薬の製造者からの製品情報もまた、有用な情報を提供する。

【0057】

発現ベクターは、構成型（constitutive）または誘導型のいずれかであることができる。誘導型ベクターは、lacオペロンで調節されているtrcプロモーター、およびトリプトファンで調節されるpLプロモーター、デキサメタゾンで誘導されるMMTV-LTRプロモーターなどの天然誘導型（naturally inducible）
20のプロモーターを含み、または、合成プロモーター、および／または隣接プロモーターに誘導型の制御を付与する付加要素を含むことができる。誘導型の合成プロモーターの例は、ハイブリッドPlac/aralプロモーターおよびPLtetO-1プロモーターである。PLtetO-1プロモーターは、ラムダファージのPLプロモーターからの高い発現レベルを利用するが、ラムダ抑制部位を、Tn10テトラサイクリン耐性オペロンのオペレーター2の2つのコピーで置き換え、このプロモーターは、Tet抑制タンパク質によって厳密に抑制され、テトラサイクリン（Tc）および無水テトラサイクリンなどのTc誘導体に応答して誘導される。ベクターはまた、グルココルチコイド応答要素（GRE）およびエストロゲン応答要素（ERE）などの、ベクターがそれぞれのホルモンレセプターを有する細胞での発現のために用いられた場合にホルモン誘導性を付与すること
30ができる、ホルモン応答要素を含むことによって、誘導型となってもよい。発現のバックグラウンドのレベルを低減するために、昆虫ホルモンであるエクジソンに応答する要素を、エクジソンレセプターの共発現と共に、代わりに用いることができる。

本発明の1つの側面では、発現ベクターは、発現されたポリペプチドを、精製および／または視覚化を容易にする小タンパク質タグに融合するように設計することができる。精製を容易にするタグは、例えばNINTA樹脂（Qiagen Inc., Valencia, CA, USA）またはTALON（登録商標）樹脂（コバルト固定化アフィニティクロマトグラフィー溶媒、Clontech Labs, Palo Alto, CA, USA）を用いた、固定化金属アフィニティクロマトグラフィーによる融合タンパク質の精製を容易にするポリヒスチジンタグを含む。融合タンパク質は、キチン結合タグおよび自己切断性のインティンを含むことができ、これにより、融合タグの自己除去（self-removal）
40を伴うキチンベースの精製が可能になる（IMPACT（登録商標）システム、New England Biolabs, Inc., Beverly, MA, USA）。あるいは、融合タンパク質は、カルモジュリンアフィニティ樹脂による精製を可能にするカルモジュリン結合ペプチドタグ（Stratagene, La Jolla, CA, USA）、または、アビジン樹脂とその後のタグの除去によるin vivoでビオチン化されたタンパク質の精製を可能にするビオチンカルボキシラーゼキャリアタンパク質の特異的に切断可能な断片（Promega, Madison, WI, USA）を含むことができる。他の有用な代替物として、本発明のタンパク質は、グルタチオン-S-トランスフェラーゼとの融合タンパク質として発現されることができ、グルタチオンへの結合の親和性および特異性により、グルタチオン-スーパーフロー樹脂
50

(Clontech Laboratories, Palo Alto, CA, USA)などのグルタチオンアフィニティ樹脂と、その後の遊離グルタチオンの溶離による精製が可能になる。他のタグとしては、例えば、抗Xpress抗体で検出可能なXpressエピトープ(Invitrogen, Carlsbad, CA, USA)、抗mycタグ抗体で検出可能なmycタグ、抗V5抗体で検出可能なV5エピトープ(Invitrogen, Carlsbad, CA, USA)、抗FLAG(登録商標)抗体で検出可能なFLAG(登録商標)エピトープ(Stratagene, La Jolla, CA, USA)、およびHAエピトープが挙げられる。

【0058】

発現されたタンパク質の分泌のために、ベクターは、リーダーペプチドなどの、分泌シグナルをコードする適切な配列を含むことができる。例えば、pSecTag2ベクター(Invitrogen, Carlsbad, CA, USA)は、種々の哺乳類細胞系からの組換えタンパク質の有効な分泌のための、マウスIgカッパー鎖のV-J2-C領域からの分泌シグナルを担持する、5.2 kbの哺乳類発現ベクターである。

発現ベクターはまた、精製および／または同定タグよりも大きなポリペプチドへの異種核酸インサートによってコードされるタンパク質を融合するように設計することができる。有用な融合タンパク質としては、コードされたタンパク質をファージまたは細胞の表面への提示、緑色蛍光タンパク質(GFP)様クロモフォアを有するものなどの本来的な蛍光タンパク質への融合、IgGfc領域への融合を可能にするもの、およびツーハイブリッドシステムで用いるための融合タンパク質が挙げられる。

ファージディスプレイ用のベクターは、コードされたポリペプチドを、例えば、M13などの線状ファージの表面に提示するために、遺伝子IIITANパク質(pIIITAN)または遺伝子VIITANパク質(pVIITAN)に融合する。Barbas et al., Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay et al. (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson et al. (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996)を参照。酵母ディスプレイ用のベクター、例えばpYD1酵母ディスプレイベクター(Invitrogen, Carlsbad, CA, USA)は、α-アグルチニン酵母接着レセプターを用いてS. cerevisiaeの表面に組換えタンパク質を提示する。哺乳類ディスプレイ用のベクター、例えば、pDisplay(登録商標)ベクター(Invitrogen, Carlsbad, CA, USA)は、N-末端細胞表面標的化シグナル、および血小板由来増殖因子のC-末端膜貫通アンカリングドメインを用いて、組換えタンパク質を標的化する。

【0059】

異種核酸によってコードされたタンパク質を、基質非依存性の、本来的に蛍光な、Aequorea victoriaからの緑色蛍光タンパク質(「GFP」)およびその変異体に融合する、広範な種類のベクターが現在存在する。GFP様クロモフォアは、A. victoria GFP (GenBankアクセッション番号AAA27721、Renilla reniformis GFP、FP583 (GenBankアクセッション番号AF168419) (DsRed)、FP593 (AF272711)、FP483 (AF168420)、FP484 (AF168424)、FP595 (AF246709)、FP486 (AF168421)、FP538 (AF168423)およびFP506 (AF168422)などの天然に存在するタンパク質において見出されたGFP様クロモフォアから選択することができ、クロモフォアの内在的な蛍光を保つ必要があるために、必要とされるネイティブタンパク質は限られている。蛍光に必要な最小のドメインを決定する方法は、当該技術分野で知られている。Li et al., J.

B i o l . C h e m . 2 7 2 : 2 8 5 4 5 - 2 8 5 4 9 (1 9 9 7) を参照。あるいは、G F P 様クロモフォアは、天然で見出されたものから改変されたG F P 様クロモフォアから選択することができる。かかる改変G F P 様クロモフォアを作出し、単独でのおよび融合タンパク質の一部としての蛍光活性についてテストする方法は、当該技術分野でよく知られている。その全体が参照によって本明細書に組み込まれるH e i m e t a l . , C u r r . B i o l . 6 : 1 7 8 - 1 8 2 (1 9 9 6) およびP a l m e t a l . , M e t h o d s E n z y m o l . 3 0 2 : 3 7 8 - 3 9 4 (1 9 9 9) を参照。多様なかかる改変クロモフォアは、現在市販されており、本発明の融合タンパク質に容易に用いることができる。これらは、E G F P (「増強G F P」)、E B F P (「増強青色蛍光タンパク質」、10)、B F P 2、E Y F P (「増強黄色蛍光タンパク質」、20)、E C F P (「増強シアン蛍光タンパク質」) またはシトリンを含む。E G F P (例えば、C o r m a c k e t a l . , G e n e 1 7 3 : 3 3 - 3 8 (1 9 9 6)、米国特許第6, 0 9 0, 9 1 9 および5, 8 0 4, 3 8 7 号を参照) は、プラスミドおよびウイルス両方の種々のベクターで見出され、市販されている (C l o n t e c h L a b s , P a l o A l t o , C A , U S A) 。 E B F P は哺乳類細胞での発現用に最適化されている一方、本来のクラゲのコドンを保っているB F P 2 は、細菌で発現することができる (例えば、H e i m e t a l . , C u r r . B i o l . 6 : 1 7 8 - 1 8 2 (1 9 9 6) およびC o r m a c k e t a l . , G e n e 1 7 3 : 3 3 - 3 8 (1 9 9 6) を参照)。これらの青方シフト (b l u e - s h i f t e d) 変異体を含むベクターは、C l o n t e c h L a b s (P a l o A l t o , C A , U S A) から入手可能である。E Y F P、E C F P (例えばH e i m e t a l . , C u r r . B i o l . 6 : 1 7 8 - 1 8 2 (1 9 9 6)、M i y a w a k i e t a l . , N a t u r e 3 8 8 : 8 8 2 - 8 8 7 (1 9 9 7) を参照) およびシトリン (例えばH e i k a l e t a l . , P r o c . N a t l . A c a d . S c i . U S A 9 7 : 1 1 9 9 6 - 1 2 0 0 1 (2 0 0 0) を参照) もまたC l o n t e c h L a b s から入手可能である。G F P 用クロモフォアはまた、その開示の全体が参照によって本明細書に組み込まれる米国特許第6, 1 2 4, 1 2 8 号、第6, 0 9 6, 8 6 5 号、第6, 0 9 0, 9 1 9 号、第6, 0 6 6, 4 7 6 号、第6, 0 5 4, 3 2 1 号、第6, 0 2 7, 8 8 1 号、第5, 9 6 8, 7 5 0 号、第5, 8 7 4, 3 0 4 号、第5, 8 0 4, 3 8 7 号、第5, 7 7 7, 0 7 9 号、第5, 7 4 1, 6 6 8 号、および第5, 6 2 5, 0 4 8 号に記載されたものを含む、タン 30

パク質の改変G F P から得ることができる。また、C o n n (e d .) , G r e e n F l u o r e s c e n t P r o t e i n (M e t h o d s i n E n z y m o l o g y , V o l . 3 0 2) , A c a d e m i c P r e s s , I n c . (1 9 9 9) を参照。これらのG F P 変異体のそれぞれのG F P 様クロモフォアは、本発明の融合タンパク質に有効に含めることができる。

【0060】

I g G F c 領域への融合は、F c R n レセプター (F c R p レセプターおよびグランベルレセプター、F c R b と呼ばれる) との相互作用を介して、タンパク質医薬製品の血清半減期を増加させ、これは、国際特許出願第W O 9 7 / 4 3 3 1 6 号、第W O 9 7 / 3 4 6 3 1 号、第W O 9 6 / 3 2 4 7 8 号、第W O 9 6 / 1 8 4 1 2 にさらに記載されて 40

本発明のタンパク質、タンパク質融合体およびタンパク質断片を、長期間、高収率で組換えにより作製するためには、安定発現が好ましい。安定発現は、選択マーカーを有するベクターの宿主細胞への組込みと、それに引き続くこれら組込み体 (i n t e g r a n t s) の選択によって容易に達成できる。p U B 6 / V 5 - H i s A、B および C (I n v i t r o g e n , C a r l s b a d , C A , U S A) などのベクターは、広範な哺乳類組織タイプおよび細胞系での、異種タンパク質の高レベルの安定発現のために設計されている。p U B 6 / V 5 - H i s は、ヒトユビキチンC 遺伝子のプロモーター/エンハンサー配列を用い、組換えタンパク質の発現を導く。2 9 3、C H O、およびN I H 3 T 3 細胞での発現レベルは、C M V およびヒトE F - 1 a プロモーターのレベルに匹敵する。 50

b s d 遺伝子は、有効な抗生剤ブラスチジンによって、安定にトランスフェクトされた哺乳類細胞の迅速な選択を可能にする。典型的にはモロニー Maus 白血病ウイルスに由来する、複製能力のないレトロウイルスベクターもまた、組み込まれたプロウイルスを有する安定なトランスフェクタントを創出するのに有用である。レトロウイルスの高効率の形質導入機構は、R e t r o P a c k (登録商標) P T 67、E c o P a c k 2 (登録商標) 293、A m p h o P a c k - 293 および G P 2 - 293 細胞系 (すべて C l o n t e c h L a b o r a t o r i e s , P a l o A l t o , C A , U S A から入手可能) などの種々のパッケージング細胞系の利用可能性と相俟って、広い宿主域に高効率で感染することを可能にする。感染の多重度を変えることにより、組み込まれたプロウイルスのコピー数が容易に調節される。

10

【0061】

勿論、すべてのベクターおよび発現制御配列が、この発明の核酸配列を発現するために、等しく良好に機能するわけではない。また、すべての宿主が同じ発現系で、等しく良好に機能するわけでもない。しかしながら、当業者は、過度の実験なしに、そして、この発明の範囲から離れることなく、これらのベクター、発現制御配列および宿主の中から、選択をなしえる。例えば、ベクターの選択においては、宿主を考慮に入れなければならない。なぜなら、ベクターはその中で複製する必要があるからである。ベクターのコピー数、このコピー数を制御する能力、もしある場合は、組み込みを制御する能力、および、ベクターによってコードされる、抗生剤または他の選択マーカーなどの任意の他のタンパク質の発現もまた考慮すべきである。本発明は、本発明のベクターを含む宿主細胞を包含し、該ベクターは細胞中にエピソームで存在するか、または宿主細胞の染色体に全部または一部が組み込まれる。他の考慮すべき事項の中で、そのいくつかは上記されているが、宿主細胞株は、発現されたタンパク質を所望の様式でプロセッシングするその能力について選択することができる。かかるポリペプチドの翻訳後修飾としては、限定されることなく、アセチル化、カルボキシル化、グリコシル化、リン酸化、脂質化およびアシル化が挙げられ、そして、かかる翻訳後修飾を伴う C S P を提供するものが、本発明の1つの側面である。本発明のポリペプチドは、翻訳後修飾を受けていてもよい。翻訳後修飾としては、アミノ酸残基セリン、トレオニンおよび/またはチロシンのリン酸化、N-結合型および/またはO-結合型グリコシル化、メチル化、アセチル化、プレニル化、アルギニル化、ユビキチン化およびラセミ化が挙げられる。翻訳後修飾のための部位を表すペプチドモチーフがあるかどうかを決定するために、ポリペプチドの配列を分析することによって、本発明のポリペプチドが翻訳後修飾を受けようかどうかを決定することができる。翻訳後修飾の予測を可能にする多数のコンピュータープログラムが存在する。例えば、タンパク質ソーティングシグナルおよび局在化部位の予測のための P S O R T、シグナルペプチド開裂部位の予測のための S i g n a l P、ミトコンドリア移行配列の予測のための M I T O P R O T および P r e d o t a r、哺乳類タンパク質におけるO型グリコシル化部位の予測のための N e t O G l y c、プレニル化アンカーおよび開裂部位の予測のための b i g - P I P r e d i c t o r および D G P I、および真核生物タンパク質における S e r、T h r および T y r リン酸化部位の予測のための N e t P h o s を含む、www. expasy. org (2001年8月31にアクセス) を参照。他のコンピュータープログラム、例えば、G C G に包含されるものなども、翻訳後修飾ペプチドモチーフを決定するのに用いることができる。

20

30

40

【0062】

翻訳後修飾のタイプの一般的な例は、the Delta Mass database <http://www.abrf.org/ABRF/Research/Committees/deltamass/deltamass.html> (2001年10月19日にアクセス)、"GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" C o o p e r e t a l. N u c l e i c A c i d s R e s. 29; 332-335 (2001

50

1) および <http://www.glycosuite.com/> (2001年10月19日にアクセス)、"O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999) および <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (2001年10月19日にアクセス)、"PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1): 237-239 (1999) および <http://www.cbs.dtu.dk/databases/PhosphoBase/> (2001年10月19日にアクセス) または <http://pir.georgetown.edu/pirwww/search/textresid.html> (2001年10月19日にアクセス) などのウェブサイトで見出すことができる。

腫瘍原性は、しばしば、タンパク質の翻訳後修飾の変化を伴う。従って、他の態様では、本発明は、正常細胞または組織からのポリペプチドの翻訳後修飾に比べて変化した翻訳後修飾を有する、癌性細胞または組織からのポリペプチドを提供する。多数の変化した翻訳後修飾が知られている。1つのよく見られる変化はリン酸化状態の変化であり、癌性細胞または組織からのポリペプチドは、正常組織からのポリペプチドに比べ過リン酸化または低リン酸化されているか、または、該ポリペプチドは、正常細胞からのポリペプチドとは異なる残基がリン酸化されている。他のよく見られる変化は、グリコシル化状態の変化であり、ここで、癌性細胞または組織からのポリペプチドは、正常組織からのポリペプチドに対しより多くのまたはより少ないグリコシル化を有し、および/または癌性細胞または組織からのポリペプチドは、非癌性細胞または組織からのポリペプチドとは異なるタイプのグリコシル化を有する。グリコシル化の変化は重大である。なぜなら、炭水化物-タンパク質または炭水化物-炭水化物相互作用が、ガン細胞の進行、播種および浸潤に重要だからである。例えば、Barchi, Curr. Pharm. Des. 6: 485-501 (2000)、Verma, Cancer Biochem. Biophys. 14: 151-162 (1994) および Dennis et al., Bioessays 5: 412-421 (1999) を参照。

ガン細胞で変化する他の翻訳後修飾は、プレニル化である。プレニル化は、疎水性プレニル基（ファルネシルまたはゲラニルゲラニル）のポリペプチドへの共有結合による付着である。プレニル化は、タンパク質を細胞膜に局在化するために必要であり、しばしば、ポリペプチド機能のために必要である。例えば、GTPアーゼシグナリングタンパク質のRasスーパーファミリーは、細胞内で機能するためにプレニル化されなければならない。例えば、Prendergast et al., Semin. Cancer Biol. 10: 443-452 (2000) および Khwaja et al., Lancet 355: 741-744 (2000) を参照。

【0063】

ガン細胞で変化する他の翻訳後修飾としては、限定されることなく、ポリペプチドのメチル化、アセチル化、アルギニル化またはアミノ酸残基のラセミ化が挙げられる。これらのケースでは、癌性細胞からのポリペプチドは、非癌性細胞からの対応するポリペプチドに比べ、増加した、または減少した量の翻訳後修飾を示し得る。

ガン細胞でのタンパク質のポリペプチド変化としては、タンパク質の異常なポリペプチドの開裂および異常なタンパク質-タンパク質相互作用が挙げられる。異常なポリペプチドの開裂は、正常細胞では通常生じないガン細胞におけるポリペプチドの開裂、または、癌性細胞での開裂の欠如であって正常細胞ではポリペプチドが開裂しているものであり得る。異常なタンパク質-タンパク質相互作用は、通常は互いに結合しないタンパク質同士の共有結合による架橋または非共有結合であり得る。あるいは、癌性細胞では、タンパク質が、非癌性細胞では結合している他のタンパク質に結合しないことがあり得る。開裂またはタンパク質-タンパク質相互作用における変化は、正常細胞に対する癌性細胞でのポリ

ペプチドの産生過剰または産生不足に起因し得、または、癌性細胞における1または2以上の他の翻訳後修飾における変化（上記参照）に起因し得る。例えば、Henschen-Edman, Ann. N. Y. Acad. Sci. 936: 580-593 (2001)を参照。

ポリペプチドの翻訳後修飾における変化、ならびにポリペプチドの開裂およびタンパク質-タンパク質相互作用における変化は、当該技術分野で知られる任意の方法により決定することができる。例えば、リン酸化における変化は、抗ホスホセリン、抗ホスホトレオニンまたは抗ホスホチロシン抗体を用いることによって、またはアミノ酸分析によって決定することができる。グリコシル化の変化は、異なる糖残基に特異的な抗体を用いて、炭水化物シーケンシングによって、または、例えばSDSポリアクリルアミドゲル電気泳動（PAGE）により決定することができる糖タンパク質のサイズの変化によって決定することができる。プレニル化、ラセミ化、メチル化、アセチル化およびアルギニル化などのタンパク質の翻訳後修飾の変化は、化学分析、タンパク質シーケンシング、アミノ酸分析によって、または特定の翻訳後修飾に特異的な抗体を用いることによって決定することができる。タンパク質-タンパク質相互作用およびポリペプチド開裂における変化は、限定されることなく、非変性PAGE（非共有結合性タンパク質-タンパク質相互作用用）、SDSPAGE（共有結合性タンパク質-タンパク質相互作用およびタンパク質開裂用）、化学開裂、タンパク質シーケンシングまたはイムノアッセイを含む、当該技術分野で知られた任意の方法によって分析することができる。

【0064】

他の態様では、本発明は翻訳後に修飾されたポリペプチドを提供する。1つの態様では、ポリペプチドは、翻訳後修飾の付加または除去によって、酵素的または化学的に修飾することができる。例えば、ポリペプチドは酵素的にグリコシル化または脱グリコシル化することができる。同様に、ポリペプチドは、MAPキナーゼ（例えばp38、ERKまたはJNK）またはチロシンキナーゼ（例えば、SrcまたはerbB2）などの精製キナーゼを用いてリン酸化することができる。ポリペプチドはまた、合成化学を介して修飾することができる。あるいは、所望の翻訳後修飾を伴うポリペプチドを発現している細胞または組織から、対象となるポリペプチドを単離することができる。他の態様では、対象となるポリペプチドをコードする核酸分子は、所望の様式でコードされたポリペプチドを翻訳後修飾することができる宿主細胞中に導入される。ポリペプチドが所望の翻訳後修飾のためのモチーフを含まない場合は、ポリペプチドが所望の翻訳後修飾のための部位を含むように、ポリペプチドをコードする核酸分子の核酸配列を変異させることによって、翻訳後修飾を変化させることができる。翻訳後修飾され得るアミノ酸配列は、当該技術分野で知られている。例えば、ウェブサイトwww.expasy.org上の上記プログラムを参照。核酸分子は次に、コードされたポリペプチドを翻訳後修飾することが可能な宿主細胞に導入される。同様に、コードされたポリペプチドが翻訳後修飾モチーフを含まないように核酸配列を変異させるか、またはコードされたポリペプチドを翻訳後修飾することができない宿主細胞にネイティブ核酸分子を導入することによって、翻訳後修飾された部位を除去することができる。

発現制御配列を選択する際には、多様な因子を考慮に入れるべきである。これらは、例えば、配列の相対的強度、その制御性、および、特に潜在的な二次構造に関する、この発明の核酸配列とのその適合性を含む。単細胞宿主は、選択したベクターとのその適合性、この発明の核酸配列によってコードされた産物の毒性、その分泌特性、ポリペプチドを正しく折りたたむその能力、その発酵または培養の要求性、およびこの発明の核酸配列によってコードされた産物のそれらからの精製の容易性を考慮して選択すべきである。

【0065】

組換え核酸分子、そして特に、この発明の発現ベクターは、この発明のポリペプチドを、異種宿主細胞で、組換えポリペプチドとして発現させるのに用いることができる。この発明のポリペプチドは全長であってもよく、または、この発明による核酸配列から組換え的に発現した全長より短いポリペプチド断片であってもよい。かかるポリペプチドは、類縁

体、誘導体およびミューテインを含み、生物活性を有しても有しなくてもよい。

本発明のベクターはまた、挿入された異種核酸からのRNAの*in vitro*での転写を可能にする要素をしばしば含む。かかるベクターは、典型的には、核酸インサートに隣接するファージプロモーター、例えば、T7、T3またはSP6からのものなどを含む。しばしば、2つの異なるかかるプロモーターが、挿入された核酸に隣接し、センス鎖およびアンチセンス鎖両方の*in vitro*での個別の産生を可能にする。

形質転換および宿主細胞に核酸を導入する他の方法（例えば、接合、プロトプラスト形質転換または融合、トランスフェクション、エレクトロポレーション、リポソーム送達、膜融合技術、高速DNAコートペレット（*high velocity DNA-coated pellets*）、ウイルス感染およびプロトプラスト融合）は、当該技術分野でよく知られた多様な方法によって達成することができる（例えば、上記Ausubelおよび上記Sambrook et alを参照）。細菌、酵母、植物または哺乳類細胞は、発現ベクター、例えばプラスミド、コスミドなどで形質転換またはトランスフェクトされ、該発現ベクターは対象となる核酸を含む。あるいは、細胞は、対象となる核酸を含むウイルス発現ベクターで感染させることができる。宿主細胞、ベクターおよび用いる形質転換方法に依存して、ポリペプチドの一過性または安定発現は、構成型または誘導型となるだろう。当業者は、ポリペプチドを一過性にまたは安定的に発現させるかどうか、および、タンパク質を構成的にまたは誘導的に発現させるかどうかを決めることができる。

【0066】

多種多様な単細胞性宿主細胞が、この発明のDNA配列を発現するのに有用である。これらの宿主は、よく知られた原核生物宿主および真核生物宿主、例えば、真菌、酵母の菌株、*Spodoptera frugiperda*（SF9）などの昆虫細胞、CHOなどの動物細胞、ならびに組織培養における植物細胞を含むことができる。適切な宿主細胞の代表的な例としては、限定されることなく、大腸菌、*Caulobacter crescentus*、ストレプトマイセス種、および*Salmonella typhimurium*などの細菌細胞、*Saccharomyces cerevisiae*、*Schizosaccharomyces pombe*、*Pichia pastoris*、*Pichia methanolic*などの酵母細胞、*Spodoptera frugiperda*からのもの、例えば、Sf9およびSf21細胞系、およびexpressSF（登録商標）細胞（Protein Sciences Corp., Meriden, CT, USA）、*Drosophila* S2細胞、および*Trichoplusia ni* High Five（登録商標）細胞（Invitrogen, Carlsbad, CA, USA）などの昆虫細胞系、および哺乳類細胞が挙げられる。典型的な哺乳類細胞としては、BHK細胞、BSC 1細胞、BSC 40細胞、BMT 10細胞、VERO細胞、COS 1細胞、COS 7細胞、チャイニーズハムスター卵巣（CHO）細胞、3T3細胞、NIH 3T3細胞、293細胞、HEPG2細胞、HeLa細胞、L細胞、MDCK細胞、HEK293細胞、WI38細胞、マウスES細胞系（例えば、129/SV、C57/BL6、DBA-1、129/SVJ系統からのもの）、K562細胞、Jurkat細胞、およびBW5147細胞が挙げられる。他の哺乳類細胞系はよく知られており、アメリカンタイプカルチャーコレクション（ATCC）（Manassas, VA, USA）およびCoriell Cell Repositories（Camden, NJ, USA）の国立一般医学研究所（NIGMS）Human Genetic Cell Repositoryから容易に入手できる。大腸に由来する細胞または細胞系が特に好ましい。なぜなら、それらはよりネイティブな翻訳後プロセッシングを提供し得るからである。特に好ましいのはヒト大腸細胞である。

組換えタンパク質のトランスフェクション、発現および精製の詳細は、十分に文書化されており、当業者に理解されている。細菌細胞発現系における外来遺伝子の組換え的産生に用いるそれぞれの工程の種々の技術的側面についてのさらなる詳細は、当該技術分野の多数のテキストおよび実験室マニュアルに見出すことができる。本明細書に参照によって組み込まれる、例えば、上記Ausubel（1992）、上記Ausubel（199

10

20

30

40

50

9)、上記 S a m b r o o k (1989)、および上記 S a m b r o o k (2001)を参照。

本発明のベクターおよび核酸を宿主細胞に導入するための方法は、当該技術分野でよく知られている。技術の選択は、第一に、導入する具体的なベクターおよび選択した宿主細胞に依存する。

核酸分子およびベクターは、多数の方法で、大腸菌などの原核生物に導入することができる。例えば、ラムダファージベクターは、典型的には、パッケージング抽出物(例えば、G i g a p a c k (登録商標)パッケージング抽出物、S t r a t a g e n e, L a J o l l a, C A, U S A)を用いてパッケージングされ、そしてパッケージングされたウイルスを用いて大腸菌を感染させる。

10

【0067】

プラスミドベクターは、典型的には、ケミカルコンピテントまたはエレクトロコンピテント細菌細胞に導入される。大腸菌細胞は、例えばC a C l₂、またはM g²⁺、M n²⁺、C a²⁺、R b⁺またはK⁺の溶液、ジメチルスルホキシド、ジチオトレイトールおよびヘキサミンコバルト(I I I)での処置でケミカルコンピテントにすることができ、H a n a h a n, J. M o l. B i o l. 166(4):557-80(1983)、そして、ベクターはヒートショックにより導入する。多種多様なケミカルコンピテント株は市販されてもいる(例えば、E p i c u r i a n C o l i (登録商標)X L 10-G o l d (登録商標)ウルトラコンピテント細胞(S t r a t a g e n e, L a J o l l a, C A, U S A)、D H 5 αコンピテント細胞(C l o n t e c h L a b o r a t o r i e s, P a l o A l t o, C A, U S A)、およびT O P 10ケミカルコンピテント大腸菌キット(I n v i t r o g e n, C a r l s b a d, C A, U S A))。細菌細胞は、エレクトロコンピテント、即ち、種々のプレパルス処置によって、エレクトロポレーションによる外生的DNAの取込みについてコンピテントにすることができる。ベクターは、エレクトロポレーションにより導入され、その後引き続き選択培地中で成長する。膨大な種類のプロトコルが、E l e c t r o p r o t o c o l s (B i o R a d, R i c h m o n d, C A, U S A) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf)でオンラインで提供されている。

20

ベクターは、酵母細胞に、スフェロプラスチ化(s p h e r o p l a s t i n g)、リチウム塩での処置、エレクトロポレーション、またはプロトプラスト融合によって導入することができる。スフェロプラストは、通常はグルスラーゼ(G l u s u l a s e)と呼ばれる、巻貝の腸の抽出物(s n a i l - g u t e x t r a c t)、またはA r t h r o b a c t e r l u t e u sからの酵素であるザイモリアーゼなどの加水分解酵素の、浸透圧安定剤、典型的には1Mのソルビトールの存在下における細胞壁の部分を除去する作用によって調製される。DNAをスフェロプラストに加え、混合物をポリエチレングリコール(P E G)とC a²⁺との溶液で共沈殿する。続いて、細胞をソルビトールの溶液に再懸濁し、溶融寒天(m o l t e n a g a r)と混合し、その後、ソルビトールを含む選択プレートの表面に層化する。

30

リチウム媒介性形質転換では、酵母細胞を、細胞壁を明らかに透過性にする酢酸リチウムで処理し、DNAを加え、細胞をP E Gで共沈殿する。細胞を短時間のヒートショックに供し、P E Gおよび酢酸リチウムがなくなるまで洗浄し、続いて、通常を選択培地を含むプレートに播種する。特別に調製した1本鎖キャリアDNAおよびある種の有機溶媒を用いることにより、増大した形質転換頻度を得ることができる。S c h i e s t l e t a l., C u r r. G e n e t. 16(5-6):339-46(1989)。

40

エレクトロポレーションでは、新たに成長した酵母の培養物を、典型的には洗浄し、ソルビトールなどの浸透圧保護剤中に懸濁し、DNAと混合し、そして細胞懸濁液をエレクトロポレーションデバイスでパルスする。続いて、細胞を選択培地を含むプレートの表面に播種する。B e c k e r e t a l., M e t h o d s E n z y m o l. 194:182-187(1991)。エレクトロポレーションによる形質転換の効率は、P E G

50

、1本鎖キャリアDNAおよび対数増殖期後期にある細胞を用いることにより、100倍以上高めることができる。YACなどの大型の構築物は、プロトプラスト融合によって導入することができる。

【0068】

哺乳類細胞および昆虫細胞は、パッケージングされたウイルスベクターによって直接感染させ、または、化学的もしくは電気的手段によってトランスフェクトすることができる。化学的トランスフェクションでは、DNAはCaPO₄で共沈殿するか、またはリポソーム系および非リポソーム系の脂質ベースの剤を用いて導入することができる。CaPO₄トランスフェクションについては市販のキットが入手可能であり(Cal Phos (登録商標) 哺乳類トランスフェクションキット、Clontech Laboratories, Palo Alto, CA, USA)、脂質媒介性トランスフェクションは、LIPOFECTAMINE (登録商標) 2000、LIPOFECTAMINE (登録商標) 試薬、CELLFECTIN (登録商標) 試薬およびLIPOFECTIN (登録商標) 試薬 (Invitrogen, Carlsbad, CA, USA)、DOTAPリポソーム系トランスフェクション試薬、FuGENE 6、X-tremeGENE Q2、DOSPER、(Roche Molecular Biochemicals, Indianapolis, IN USA)、Effectene (登録商標)、Polyfect (登録商標)、Superfect (登録商標) (Qiagen, Inc., Valencia, CA, USA)などの市販の試薬を用いて行うことができる。哺乳類細胞をエレクトロポレーションするためのプロトコルは、オンラインでエレクトロプロトコル (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf)、Norton et al. (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000)に見出すことができ、その全体は参照によって本明細書に組み込まれる。他のトランスフェクション技術としては、粒子衝撃によるトランスフェクションおよびマイクロインジェクションが挙げられる。例えば、Cheng et al., Proc. Natl. Acad. Sci. USA 90 (10): 4455-9 (1993)、Yang et al., Proc. Natl. Acad. Sci. USA 87 (24): 9568-72 (1990)を参照。組換え的に作製された本発明のタンパク質の作製の後に、任意に精製を行うことができる。

組換え的に発現されたタンパク質の精製は、現在当業者によく知られている。その開示の全体が参照によって本明細書に組み込まれ、従って本明細書では詳細に述べる必要のない、例えば、Thorner et al. (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000)、Harbin (ed.), Cloning, Gene Expression and Protein Purification: Experimental Procedures and Process Rationale, Oxford Univ. Press (2001)、Marshak et al., Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996)、およびRoe (ed.), Protein Purification Applications, Oxford University Press (2001)を参照。

しかしながら、簡潔に述べると、精製タグが、かかるタグの付加した発現ベクターの使用

を介して融合されている場合、精製は、少なくとも部分的には、該タグに適切な手段、例えば、ポリヒスチジンタグのための固定化金属アフィニティクロマトグラフィーの使用などによって行うことができる。当該技術分野に一般的な他の技術としては、硫酸アンモニウム分画、免疫沈降、高速タンパク質液体クロマトグラフィー (fast protein liquid chromatography) (FPLC)、高速液体クロマトグラフィー (HPLC)、および分取ゲル電気泳動が挙げられる。

【0069】

ポリペプチド

本発明の他の目的は、本発明の核酸分子によってコードされるポリペプチドを提供することである。好ましい態様では、ポリペプチドは、大腸特異的ポリペプチド (CSP) である。より一層好ましい態様では、ポリペプチドは、配列番号 148~250 のアミノ酸配列を含むポリペプチドに由来する。本明細書で定義されるポリペプチドは、上記で論じたように組換え的に作製してもよく、天然にタンパク質を発現している細胞から単離してもよく、または、本明細書の教示にしたがい、そして当業者によく知られた方法を用いて化学的に合成してもよい。

他の側面では、ポリペプチドは、ポリペプチドの断片を含んでもよく、ここで該断片は本明細書に記載されているとおりのものである。好ましい態様では、ポリペプチド断片は CSP の断片である。より好ましい態様では、断片は、配列番号 148~250 のアミノ酸配列を含むポリペプチドに由来する。CSP 全体の断片のみを含むポリペプチドはまた、CSP であるポリペプチドであってもなくてもよい。例えば、全長ポリペプチドは大腸特異的であってもよく、一方、その断片は大腸と同様に他の組織に見出されてもよい。CSP ではないポリペプチドは、それが断片、類縁体、ミューテイン、相同タンパク質または誘導体であるかどうかにかかわらず、特に動物を免疫し抗 CSP 抗体を調製するのに有用である。しかしながら、好ましい態様では、該部分または断片は CSP である。ポリペプチドが CSP かどうかを決定する方法は、下記で説明してある。

少なくとも 6 個の連続したアミノ酸の断片が、参照タンパク質の B 細胞および T 細胞エпитオープのマッピングに有用である。その開示の全体が参照によって本明細書に組み込まれる、例えば Geysen et al., Proc. Natl. Acad. Sci. USA 81: 3998-4002 (1984)、および米国特許第 4,708,871 号および第 5,595,915 号を参照。かかるエピトープマッピングに有用であるために、該断片は、それ自身、免疫原性であり、イムノドミナントエピトープの一部であり、またネイティブな抗体に認識される必要はないので、本発明の他の少なくとも 6 アミノ酸の断片は、かかる研究において有用性を有する。

【0070】

少なくとも 8 個の連続したアミノ酸、しばしば少なくとも 15 個の連続したアミノ酸の断片は、本発明のタンパク質を認識する抗体を惹起するためのイムノゲンとして有用である。その開示の全体が参照によって本明細書に組み込まれる、例えば Lerner, Nature 299: 592-596 (1982); Shinnick et al., Annu. Rev. Microbiol. 37: 425-46 (1983); Sutcliffe et al., Science 219: 660-6 (1983) を参照。上述の参考文献にさらに記載されているとおり、タンパク質などのキャリアに結合したすべての 8 マーは、事実上免疫原性を示す。これは、それらが、結合したペプチドについての抗体を惹起することができることを示す。従って、本発明のタンパク質の少なくとも 8 アミノ酸の断片のすべては、イムノゲンとしての有用性を有する。

少なくとも 8、9、10 または 12 個の連続したアミノ酸の断片はまた、抗体への (エピトープマッピングでのように)、および多量体複合体におけるサブユニットなどの天然の結合パートナーへのまたは対象タンパク質のレセプターもしくはリガンドへの、タンパク質全体またはその一部の結合の競合阻害剤として有用である。この競合阻害は、対象となるタンパク質に特異的に結合する分子の同定および分離を可能にする。その全体が参照によって本明細書に組み込まれる米国特許第 5,539,084 号および第 5,783,6

10

20

30

40

50

74号。

本発明のタンパク質またはタンパク質断片は、従って、少なくとも6アミノ酸長、典型的には、少なくとも8、9、10または12アミノ酸長、そして、しばしば少なくとも15アミノ酸長である。しばしば、本発明のタンパク質またはその断片は、少なくとも20アミノ酸長、さらには25アミノ酸、30アミノ酸、35アミノ酸または50アミノ酸長以上である。勿論、少なくとも75個のアミノ酸、100個のアミノ酸、またはさらには150個のアミノ酸を有する大型の断片もまた有用であり、ときには好ましい。当業者は、ポリペプチドをコードする核酸分子、例えばCSNAを切断し、ついでそれを組換え的に発現させることによってポリペプチドの断片を作製することができる。あるいは、全長ポリペプチドの一部を化学的に合成することによって断片を作製することができる。また、組換えポリペプチドまたは単離された天然に存在するポリペプチドを酵素的に開裂させることによって、断片を作製することもできる。ポリペプチド断片を作成する方法は、当該技術分野でよく知られている。例えば上記Sambrook (1989)、上記Sambrook (2001)、上記Ausubel (1992)、および上記Ausubel (1999)を参照。1つの態様では、本発明のポリペプチドの断片のみを含むポリペプチド、好ましくはCSPを、ポリペプチドの化学的または酵素的開裂によって作製することができる。好ましい態様では、ポリペプチド断片は、ポリペプチドの断片、好ましくはCSPをコードする核酸分子を、宿主細胞中で発現させることによって作製される。

10

【0071】

「ポリペプチド」はまた、本明細書で用いられる場合、具体的に例示されたポリペプチドのミュータント、融合タンパク質、相同タンパク質および対立遺伝子変異体を含むものとする。

20

変異タンパク質またはミューテインは、天然に存在するポリペプチドと同一のまたは異なる特性を有してもよく、ネイティブなタンパク質のアミノ酸配列と比べて、少なくとも1つのアミノ酸の挿入、重複、欠失、再編成または置換を含む。タンパク質の機能を変化させない微小な欠失および挿入はしばしばみられる。1つの態様では、ミューテインは大腸特異的であってもなくてもよい。好ましい態様では、ミューテインは大腸特異的である。好ましい態様では、ミューテインは、配列番号148～250のアミノ酸配列の1つと比べて、少なくとも1つのアミノ酸の挿入、重複、欠失、再編成または置換を含む。より好ましい態様では、ミューテインは、配列番号148～250のアミノ酸配列を含むCSPに対し、少なくとも50%の配列同一性、より好ましくは少なくとも60%の配列同一性、さらにより好ましくは少なくとも70%の配列同一性、より一層好ましくは少なくとも80%の配列同一性を示すものである。より一層好ましい態様では、ミューテインは、配列番号148～250のアミノ酸配列を含むCSPに対し、少なくとも85%、より好ましくは90%、さらにより好ましくは95%または96%、そしてより一層好ましくは少なくとも97%、98%、99%または99.5%の配列同一性を示す。

30

【0072】

ミューテインは、天然に存在する変異細胞、組織または生物から単離することによって作製することができる。ミューテインは、実験的に変異導入された細胞、組織または生物から単離することによって作製することができる。あるいは、ミューテインは、ポリペプチドの化学的操作によって、例えば、合成的または半合成的な化学的技法を用いてアミノ酸残基をタンパク質のアミノ酸残基に変化させることによって作製することができる。好ましい態様では、ミューテインは、天然に存在する核酸分子に比べて変化した核酸分子を含む宿主細胞から作製することができる。例えば、本発明の核酸配列に1または2以上の変異を導入し、次いでそれを組換え的に発現させることによってポリペプチドのミューテインを作製することができる。これらの変異は標的化することができ、この場合は特別にコードされたアミノ酸が変化しており、または非標的化することができ、この場合は、ポリペプチド内のランダムにコードされたアミノ酸が変化している。ランダムアミノ酸変化を伴うミューテインは、以下に説明するとおり、具体的な生物活性または特性について、特

40

50

に該ポリペプチドが大腸特異的かどうかをスクリーニングすることができる。複数のランダム変異を、当該技術分野でよく知られた方法によって、例えば、エラープローンPCR、シャッフリング、オリゴヌクレオチドを用いた変異導入、アセンブリPCR、セクシャルPCR変異導入、*in vivo*変異導入、カセット変異導入、エクスポネンシャルアンサンブル変異導入、リカーシブアンサンブル変異導入および部位特異的変異導入によって、遺伝子に導入することができる。標的化されたアミノ酸変化またはランダムアミノ酸変化を有するミューテインを作製する方法は、当該技術分野でよく知られている。それぞれが参照によって本明細書に組み込まれる、例えば上記Sambrook (1989)、上記Sambrook (2001)、上記Ausubel (1992) および上記Ausubel (1999)、米国特許第 5, 223, 408号、および上記で論じた参考文献を参照。 10

「ポリペプチド」は、本明細書中で用いる場合には、本明細書中で例示されたポリペプチドと相同なポリペプチドをも含むものとする。好ましい態様では、ポリペプチドはCSPと相同である。より一層好ましい態様では、ポリペプチドは、配列番号148～250のアミノ酸配列からなる群から選択されるCSPと相同である。好ましい態様では、相同ポリペプチドは、CSPと顕著な配列同一性を示すものである。より好ましい態様では、ポリペプチドは、配列番号148～250のアミノ酸配列を含むCSPと顕著な配列同一性を示すものである。より一層好ましい態様では、相同ポリペプチドは、配列番号148～250のアミノ酸配列を含むCSPと、少なくとも50%の配列同一性、より好ましくは少なくとも60%の配列同一性、さらにより好ましくは少なくとも70%の配列同一性、 20
より一層好ましくは少なくとも80%の配列同一性を示すものである。より一層好ましい態様では、相同ポリペプチドは、配列番号148～250のアミノ酸配列を含むCSPに対し、少なくとも85%、より好ましくは90%、さらにより好ましくは95%または96%、そしてより一層好ましくは少なくとも97%または98%の配列同一性を示すものである他の好ましい態様では、相同ポリペプチドは、配列番号148～250のアミノ酸配列を含むCSPに対し、少なくとも99%、より好ましくは99.5%、さらにより好ましくは99.6%、99.7%、99.8%または99.9%の配列同一性を示すものである。好ましい態様では、アミノ酸置換は、上記で論じたとおり、保存的なアミノ酸置換である。

【0073】

他の態様では、相同ポリペプチドは、CSNAに選択的にハイブリダイズする核酸分子によってコードされたものである。好ましい態様では、相同ポリペプチドは、本明細書で定義された低いストリンジェンシー、中程度のストリンジェンシーまたは高いストリンジェンシー条件下で、CSNAにハイブリダイズする核酸分子によってコードされる。より好ましい態様では、CSNAは、配列番号1～147からなる群から選択される。他の好ましい態様では、相同ポリペプチドは、本明細書で定義された低いストリンジェンシー、中程度のストリンジェンシーまたは高いストリンジェンシー条件下で、CSPをコードする核酸分子にハイブリダイズする核酸分子によってコードされる。より好ましい態様では、CSPは、配列番号148～250からなる群から選択される。 30

相同ポリペプチドは、他の種に由来する天然に存在するもの、特に、チンパンジー、アカゲザル、ヒヒまたはゴリラなどの他の霊長類に由来するものであって、相同性ポリペプチドが、配列番号148～250のものと顕著な配列同一性を示すアミノ酸配列を含むものであってもよい。相同ポリペプチドはまた、CSPがポリペプチドのファミリーの成員である場合には、ヒトの天然に存在するポリペプチドであってもよい。相同性ポリペプチドはまた、限定されずに、家畜種、例えば、イヌ、ネコ、マウス、ラット、ウサギ、モルモット、ハムスター、ウシ、ウマ、ヤギまたはブタを含む、霊長類ではない哺乳類種に由来する、天然に存在するポリペプチドであってもよい。相同ポリペプチドはまた、鳥類または爬虫類などの哺乳類ではない種に由来する、天然に存在するポリペプチドであってもよい。天然に存在する相同タンパク質は、ヒトまたは他の種から直接単離してもよい。あるいは、天然に存在する相同ポリペプチドをコードする核酸分子を単離し、相同ポリペプチ 40 50

ドを組換え的に発現させるために用いてもよい。他の態様では、相同ポリペプチドは、核酸分子のランダム突然変異とその後の核酸分子の発現によって、実験的に作製されたものであってもよい。他の態様では、相同ポリペプチドは、C S Pのコードされたアミノ酸を変化させるための1または2以上のコドンの特異的変異 (directed mutation) によって実験的に作製されたものであってもよい。さらに、相同タンパク質は、C S Pであるポリペプチドをコードしてもしなくてもよい。しかしながら、好ましい態様では、相同ポリペプチドはC S Pであるポリペプチドをコードする。

【0074】

タンパク質の関連性はまた、第2の機能的テストである、第1のタンパク質が、第2のタンパク質の抗体への結合を競合的に阻害する能力を用いて特徴付けすることができる。従って、本発明の他の側面は、本明細書中で詳細に説明したものと配列が同一である単離タンパク質を提供するだけでなく、本発明の種々の単離ポリペプチドの全部または一部への抗体の結合を競合的に阻害する単離タンパク質 (「交差反応性タンパク質」) を提供する。かかる競合阻害は、当該技術分野でよく知られているイムノアッセイを用いて容易に決定することができる。

上記で論じたとおり、1塩基多型 (SNPs) は真核生物のゲノムで頻繁に生じ、1つの種の1つの個体から決定された配列は、集団内に存在する他の対立遺伝子の形態と異なり得る。従って、「ポリペプチド」は、本明細書中で用いる場合は、C S Pをコードする核酸分子の対立遺伝子変異体によってコードされるポリペプチドをも含むものとする。好ましい態様では、ポリペプチドは、配列番号148~250からなる群から選択されるアミノ酸配列を有するポリペプチドをコードする遺伝子の対立遺伝子変異体によってコードされる。より一層好ましい態様では、ポリペプチドは、配列番号1~147からなる群から選択される核酸配列を有する遺伝子の対立遺伝子変異体によってコードされる。

他の態様では、本発明は、本発明の核酸分子によってコードされるポリペプチドの誘導体を含むポリペプチドを提供する。好ましい態様では、ポリペプチドはC S Pである。好ましい態様では、ポリペプチドは、配列番号148~250からなる群から選択されるアミノ酸配列を有し、またはそのミューテイン、対立遺伝子変異体、相同タンパク質または断片である。好ましい態様では、誘導体は、アセチル化、カルボキシル化、リン酸化、グリコシル化またはユビキチン化されている。他の好ましい態様では、誘導体は、例えば¹²⁵I、³²P、³⁵Sおよび³Hなどの放射性同位体によって標識されている。他の好ましい態様では、誘導体は、フルオロフォア、化学発光剤、酵素、および標識されたりガンドの特異的結合対の成員として機能する抗リガンドによって標識されている。

【0075】

ポリペプチド修飾は当業者によく知られており、科学的文献に詳細に記載されている。数種の特に一般的な修飾である、例えば、グリコシル化、脂質付加、硫化、グルタミン酸残基のガンマーカルボキシル化、ヒドロキシル化およびADP-リボシル化は、最も基本的なテキスト、例えば、Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993)などに記載されている。多数の詳細な総説がこの主題について利用可能であり、例えば、Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983)中、Worldによって提供されているもの、Seifter et al., Meth. Enzymol. 182: 626-646 (1990)およびRattan et al., Ann. N. Y. Acad. Sci. 663: 48-62 (1992)などがある。よく知られているように、そして上記の通り、ポリペプチドは常には全体が線状ではないと理解される。例えば、ポリペプチドは、ユビキチン化の結果として分枝状であることができ、そして、一般的に、天然のプロセッシングイベントおよび天然には生じないヒトの操作によってなされたイベントを含む翻訳後イベントの結果として、分子を有するまたは有しない環状であることができる。環状、分枝状および分枝環状ポリペプチドは、非翻訳的

な天然のプロセスによって、および同様に、完全に合成的な方法によって合成することができる。修飾は、ペプチド主鎖、アミノ酸側鎖およびアミノまたはカルボキシ末端を含む、ポリペプチドのどこでも生じ得る。実際、共有結合的な修飾による、ポリペプチド中のアミノ基もしくはカルボキシル基またはその両方のブロックは、天然に存在するポリペプチドおよび合成ポリペプチドにおいてよく見られ、そしてかかる修飾は本発明のポリペプチドにも同様に存在し得る。例えば、大腸菌で産生されたポリペプチドのアミノ末端残基は、タンパク溶解プロセッシングの前は、殆ど常にN-ホルミルメチオニンである。有用な合成後（および翻訳後）修飾は、フルオロフォアなどの検出可能な標識への結合を含む。多種多様なアミン反応性およびチオール反応性のフルオロフォアが合成され、これらは、非変性条件下で、一方では、リジン残基のN-末端アミノ基およびイプシロンアミノ基と、他方では、システイン残基のチオール基と反応する。

10

【0076】

タンパク質の、種々のアミン反応性またはチオール反応性フルオロフォアとの結合を可能にするキットが市販されている。例えばMolecular Probes, Inc.

(Eugene, OR, USA)は、アレクサフルオール350、アレクサフルオール430、フルオレセイン-EX、アレクサフルオール488、オレゴングリーン488、アレクサフルオール532、アレクサフルオール546、アレクサフルオール546、アレクサフルオール568、アレクサフルオール594およびテキサスレッド-Xへタンパク質を結合するためのキットを提供している。

多種多様なタンパク質のアミノ反応性およびチオール反応性フルオロフォアが市販されており、これは、アレクサフルオール（登録商標）350、アレクサフルオール（登録商標）488、アレクサフルオール（登録商標）532、アレクサフルオール（登録商標）546、アレクサフルオール（登録商標）568、アレクサフルオール（登録商標）594、アレクサフルオール（登録商標）647（モノクローナル抗体標識キットがから入手可能）、BODIPY色素、例えば、BODIPY 493/503、BODIPY FL、BODIPY R6G、BODIPY 530/550、BODIPY TMR、BODIPY 558/568、BODIPY 558/568、BODIPY 564/570、BODIPY 576/589、BODIPY 581/591、BODIPY TR、BODIPY 630/650、BODIPY 650/665、カスケードブルー、カスケードイエロー、ダンシル、リサミンローダミンB、マリナブルー、オレゴングリーン488、オレゴングリーン514、パシフィックブルー、ローダミン6G、ローダミングリーン、ローダミンレッド、テトラメチルローダミン、テキサスレッド（Molecular Probes, Inc., Eugene, OR, USAから入手可能）を含む。

20

30

本発明のポリペプチドはまた、二価性架橋試薬を用いて、フルオロフォア、他のタンパク質、および他の巨大分子に結合することができる。一般的なホモ二価性試薬としては、例えば、APG、AEDP、BASED、BMB、BMDB、BMH、BMOE、BM[P EO] 3、BM[P EO] 4、BS3、BSOCOES、DFDNB、DMA、DMP、DMS、DPDPB、DSG、DSP (Lomant's Reagent)、DSS、DST、DTBP、DTME、DTSSP、EGS、HBVS、Sulfo-BSOCOES、Sulfo-DST、Sulfo-EGS（すべてPierce, Rockford, IL, USAから入手可能）が挙げられ、一般的なヘテロ二価性架橋剤としては、ABH、AMAS、ANB-NOS、APDP、ASBA、BMPA、BMPH、BMPS、EDC、EMCA、EMCH、EMCS、KMUA、KMUH、GMBS、LC-SMCC、LC-SPDP、MBS、M2C2H、MPBH、MSA、NHS-ASA、PDPH、PMPI、SADP、SAED、SAND、SANPAH、SASD、SATP、SBAP、SFAD、SIA、SIAB、SMCC、SMPB、SMPH、SMPT、SPDP、Sulfo-EMCS、Sulfo-GMBS、Sulfo-HSAB、Sulfo-KMUS、Sulfo-LC-SPDP、Sulfo-MBS、Sulfo-NHS-LC-ASA、Sulfo-SADP、Sulfo-SANPAH、Sulfo-SIAB、Sulfo-SMCC、Sulfo-SMPB、Sulfo-LC-

40

50

S M P T、S V S B、T F C S (すべて P i e r c e, R o c k f o r d, I L, U S Aから入手可能)が挙げられる。

本発明のポリペプチド、断片および融合タンパク質は、かかる架橋剤を用いて、アミンまたはチオール反応性ではないフルオロフォアに結合させることができる。本発明のポリペプチド、断片および融合タンパク質に有効に結合させることができるタンパク質の標識としては、放射性標識、超音波検査用造影試薬およびMRI造影剤が挙げられる。

【0077】

本発明のポリペプチド、断片および融合タンパク質はまた、架橋剤を用いて、K L H、ウシチログロブリン、およびさらにはウシ血清アルブミン(B S A)などのキャリアタンパク質に、抗C S P抗体を惹起するための免疫原性を高めるために、有効に結合させることができる。 10

本発明のポリペプチド、断片および融合タンパク質はまた、ポリエチレングリコール(P E G)に有効に結合させることができる。P E G化は、補充療法のために静脈内投与されたタンパク質の血清半減期を増大させる。D e l g a d o e t a l., C r i t. R e v. T h e r. D r u g C a r r i e r S y s t. 9 (3-4): 249-304 (1992)、S c o t t e t a l., C u r r. P h a r m. D e s. 4 (6): 423-38 (1998)、D e S a n t i s e t a l., C u r r. O p i n. B i o t e c h n o l. 10 (4): 324-30 (1999)。これらの全体は、参照によって本明細書に組み込まれる。P E Gモノマーは、緩和な条件下で直接的な付着を可能にするトレシルクロライド(2, 2, 2-トリフルオロエタンスルホニルクロライド)で活性化されたP E Gモノマーを用いるP E G化によって、タンパク質に直接またはリンカーを介して付着させることができる。 20

さらに他の態様では、本発明は、本発明の核酸分子によってコードされるポリペプチドの類縁体を提供する。好ましい態様では、ポリペプチドはC S Pである。より好ましい態様では、類縁体は、配列番号148~250の1つのアミノ酸配列の一部または全部を有するポリペプチドに由来する。好ましい態様では、類縁体は、天然に存在するポリペプチドに比べ、1または2以上の、非天然アミノ酸の置換または非天然の残基間結合を含むものである。一般的に、非ペプチド類縁体は、C S Pと構造的に類似するが、1または2以上のペプチド結合が、 $-CH_2NH-$ 、 $-CH_2S-$ 、 $-CH_2-CH_2-$ 、 $-CH=CH-$ (シスおよびトランス)、 $-COCH_2-$ 、 $-CH(OH)CH_2-$ および $-CH_2S$ 30
O-からなる群から選択される結合に置換されている。他の態様では、非ペプチド類縁体は、より安定なペプチドを創出するために、C S Pの1または2以上のアミノ酸の、同型のD-アミノ酸とのまたは他の非天然アミノ酸との置換を含む。D-アミノ酸は、化学的なペプチド合成の最中に容易に組み込むことができる。D-アミノ酸で構成されたペプチドは、タンパク溶解性の攻撃により抵抗性である。D-アミノ酸の組み込みはまた、ペプチドに特定の3次元コンホメーションを付与するために用いることができる。化学合成の最中に一般的に添加される他のアミノ酸類縁体としては、オルニチン、ノルロイシン、リン酸化アミノ酸(典型的には、ホスホセリン、ホスホトレオニン、ホスホチロシン)、L-マロニルチロシン(L-malonyl tyrosine)、ホスホチロシンの非加水分解性類縁体(例えばK o l e e t a l., B i o c h e m. B i o p h y s. R e 40
s. C o m. 209: 817-821 (1995)を参照)、および種々のハロゲン化フェニルアラニン誘導体が挙げられる。

【0078】

非天然アミノ酸は、固相化学合成の最中に、または組換え技術によって組み込むことができるが、前者の方が典型的で、より一般的である。ペプチドの固相化学合成は、当該技術分野で十分に確立されている。手順は、とりわけ、C h a n e t a l. (e d s.) , F m o c S o l i d P h a s e P e p t i d e S y n t h e s i s : A P r a c t i c a l A p p r o a c h (P r a c t i c a l A p p r o a c h S e r i e s) , O x f o r d U n i v. P r e s s (M a r c h 2000)、J o n e s, A m i n o A c i d a n d P e p t i d e S y n t h e s i s (O x f o r d C h e 50

mistry Primers, No 7), Oxford Univ. Press (1992)、およびBodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993)に記載されており、これらの開示の全体は参照によって本明細書に組み込まれる。

また、検出可能な標識を有するアミノ酸類縁体を、誘導体および類縁体を提供するために、合成の最中に有効に組み込むこともできる。例えば、ビオチンは、ビオチノイルー(9-フルオレニルメトキシカルボニル)-L-リジン(FMOCビオシチン)(Molecular Probes, Eugene, OR, USA)を用いて付加することができる。ビオチンはまた、大腸菌BirA基質ペプチドの融合タンパク質への組み込みによって、酵素的に付加することもできる。ダブシル-L-リジンのFMOCおよびtBOC誘導体(Molecular Probes, Inc., Eugene, OR, USA)は、ダブシルクロモフォアを、合成の最中に、ペプチド配列の選択した部位に組み込むために用いることができる。蛍光共鳴エネルギー転送(FRET)システムにおいて、ダブシルクエンチャーとの対合のための最も一般的なフルオロフォアである、アミノナフタレン誘導体EDANSは、EDANS-FMOC-L-グルタミン酸または対応するtBOC誘導体(両方ともMolecular Probes, Inc., Eugene, OR, USAから)を用いることによって、ペプチドの自動合成の最中に導入することができる。テトラメチルローダミンフルオロフォアは、(FMOC)-TMR-L-リジン(Molecular Probes, Inc. Eugene, OR, U 20

化学合成の最中に組み込むことができる他の有用なアミノ酸類縁体としては、アリル側鎖保護を有するアスパラギン酸、グルタミン酸、リジン、およびチロシンの類縁体(Applied Biosystems, Inc., Foster City, CA, USA)が挙げられる。アリル側鎖により、環状、分枝鎖状、スルホン化、グリコシル化およびリン酸化ペプチドの合成が可能になる。

【0079】

化学合成の最中に組み込むことが可能な多数の他のFMOC保護非天然アミノ酸類縁体が市販されており、例えば、Fmoc-2-アミノピシクロ[2.2.1]ヘプタン-2-カルボン酸、Fmoc-3-エンド-アミノピシクロ[2.2.1]ヘプタン-2-エンド-カルボン酸、Fmoc-3-エキソ-アミノピシクロ[2.2.1]ヘプタン-2-エキソ-カルボン酸、Fmoc-3-エンド-アミノ-ピシクロ[2.2.1]ヘプト-5-エン-2-エンド-カルボン酸、Fmoc-3-エキソ-アミノ-ピシクロ[2.2.1]ヘプト-5-エン-2-エキソ-カルボン酸、Fmoc-シス-2-アミノ-1-シクロヘキサンカルボン酸、Fmoc-トランス-2-アミノ-1-シクロヘキサンカルボン酸、Fmoc-1-アミノ-1-シクロペンタンカルボン酸、Fmoc-シス-2-アミノ-1-シクロペンタンカルボン酸、Fmoc-1-アミノ-1-シクロプロパンカルボン酸、Fmoc-D-2-アミノ-4-(エチルチオ)酪酸、Fmoc-L-2-アミノ-4-(エチルチオ)酪酸、Fmoc-L-ブチオニン、Fmoc-S-メチル-L-システイン、Fmoc-2-アミノ安息香酸(アントラニル酸)、Fmoc-3-アミノ安息香酸、Fmoc-4-アミノ安息香酸、Fmoc-2-アミノベンゾフェノン-2'-カルボン酸、Fmoc-N-(4-アミノベンゾイル)-β-アラニン、Fmoc-2-アミノ-4,5-ジメトキシ安息香酸、Fmoc-4-アミノ馬尿酸、Fmoc-2-アミノ-3-ヒドロキシ安息香酸、Fmoc-2-アミノ-5-ヒドロキシ安息香酸、Fmoc-3-アミノ-4-ヒドロキシ安息香酸、Fmoc-4-アミノ-3-ヒドロキシ安息香酸、Fmoc-4-アミノ-2-ヒドロキシ安息香酸、Fmoc-5-アミノ-2-ヒドロキシ安息香酸、Fmoc-2-アミノ-3-メトキシ安息香酸、Fmoc-4-アミノ-3-メトキシ安息香酸、Fmoc-2-アミノ-3-メチル安息香酸、Fmoc-2-アミノ-5-メチル安息香酸、Fmoc-2-アミノ-6-メチル安息香酸、Fmoc-3-アミノ-2-メチル安息香酸、Fmoc-3-アミノ-4-メチル安息香酸 50

、Fmoc-4-アミノ-3-メチル安息香酸、Fmoc-3-アミノ-2-ナフトエ酸、Fmoc-D, L-3-アミノ-3-フェニルプロピオン酸、Fmoc-L-メチルドパ、Fmoc-2-アミノ-4, 6-ジメチル-3-ピリジンカルボン酸、Fmoc-D, L-アミノ-2-チオフエン酢酸、Fmoc-4-(カルボキシメチル)ピペラジン、Fmoc-4-カルボキシピペラジン、Fmoc-4-(カルボキシメチル)ホモピペラジン、Fmoc-4-フェニル-4-ピペリジンカルボン酸、Fmoc-L-1, 2, 3, 4-テトラヒドロノルハルマン-3-カルボン酸、Fmoc-L-チアゾリジン-4-カルボン酸が挙げられ、これらはすべてPeptide Laboratory (Richmond, CA, USA) から入手可能である。

非天然残基はまた、サプレッサーtRNA、典型的にはUAG終止コドンを認識するものを、所望の非天然アミノ酸の化学的アミノアシル化により操作することによって、生合成的に付加することができる。選択した終止コドンUAGをタンパク質遺伝子の対象となる部位に導入するために、慣用の部位特異的変異導入を用いる。アシル化サプレッサーtRNAと変異遺伝子とがin vitroの転写／翻訳系で混合された場合、非天然アミノ酸はUAGに応答して組込まれ、このアミノ酸を特定の位置に含むタンパク質を与える。Liu et al., Proc. Natl Acad. Sci. USA 96 (9): 4780-5 (1999)、Wang et al., Science 292 (5516): 498-500 (2001)。

10

【0080】

融合タンパク質

20

本発明はまた、本発明によるポリペプチドおよび断片それぞれの非相同性ポリペプチドへの融合体を提供する。好適態様において、このポリペプチドはCSPである。さらに好適な態様において、非相同性ポリペプチドに融合されるポリペプチドは、配列番号148～250のアミノ酸配列の一部または全部を含有するか、またはその突然変異タンパク質、相同性ポリペプチド、類縁体または誘導体である。さらにまた好適な態様において、融合タンパク質をコードする核酸分子は、配列番号1～147の核酸配列の一部または全部を包含するか、または選択的にハイブリダイズする核酸配列の一部または全部を包含するか、または配列番号1～147の核酸配列を包含する核酸分子に対して相同性である。

本発明による融合タンパク質は、本発明のタンパク質の少なくとも一個の断片を包含し、この断片は少なくとも6、代表的に少なくとも8、多くの場合に少なくとも15、また通常少なくとも16、17、18、19、または20のアミノ酸長を有する。融合体に包含させる本発明によるタンパク質の断片は、通常少なくとも25アミノ酸長、少なくとも50アミノ酸長であることができ、また少なくとも75、100アミノ酸長または150アミノ酸長であることさえできる。本発明のタンパク質全体を包含する融合体は、特に有用である。

30

本発明の融合タンパク質内に包含される非相同性ポリペプチドは、少なくとも6アミノ酸の長さ、多くの場合に少なくとも8アミノ酸の長さ、また通常少なくとも15、20、および25アミノ酸の長さを有する。さらに長いポリペプチド、例えばIgG Fc領域を、また完全タンパク質（例えば、GFPクロモホール含有タンパク質）さえもを包含する融合体は、特に有用である。

40

本発明のベクターおよび発現ベクターの説明について上記したように（この説明は、その全体を引用してここに組入れる）、本発明の融合タンパク質に包含させる非相同性ポリペプチドは通常、組換え発現されたタンパク質の精製および／または可視化が促進されるようにデザインされているポリペプチドを包含することができる。例えば、Ausubelによる上記文献、16章（1992年）。精製標識（tag）はまた、化学合成される融合体中に挿入することができるが、化学合成は代表的に、HPLCによる追加の精製で充分である充分な純度を提供する。しかしながら、上記可視化標識は、タンパク質が化学合成によって生成された場合でさえも、それらの有用性を保有し、またこのような包含は、本発明の融合タンパク質を本発明によるポリペプチドの存在の直接的検出可能マーカーとして有用にする。

50

【0081】

上記でまた説明したように、本発明の融合タンパク質に包含させる非相同性ポリペプチドは通常、分泌シグナルおよび／またはリーダー配列の配合によって、組換え発現タンパク質のー原核生物宿主の細胞周辺腔または細胞外環境中への、真核細胞の培養培地へのー分泌を促進する非相同性ポリペプチドを包含する。一例として、His⁶ タグ付けしたタンパク質は、Niアフィニティカラムで精製することができ、またGST融合タンパク質はグルタチオンアフィニティカラムで精製することができる。同様に、IgGのFcドメインを包含する融合タンパク質は、プロテインAまたはプロテインGカラムで精製することができ、またエピトープ標識、例えばmycを含有する融合タンパク質は、抗-c-myc抗体を含有する免疫親和性カラムを用いて精製することができる。このエピトープ標識は、精製後に開裂させることができる酵素開裂部位によって、必須遺伝子によりコードされるタンパク質から分離されていると好ましい。細胞表面上で発現させることができる融合タンパク質をコードする核酸分子にかかわる説明をまた、参照することができる。

別の本発明による有用なタンパク質融合体は、本発明のタンパク質を酵母ツーハイブリッド系におけるベイトとして使用することを可能にするタンパク質融合体を包含する。下記文献を参照することができる：Bartel et al. (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu et al., Yeast Hybrid Technologies, Eaton Publishing (2000); Fields et al., Trends Genet. 10 (8): 286-92 (1994); Mendelsohn et al., Curr. Opin. Biotechnol. 5 (5): 482-6 (1994); Luban et al., Curr. Opin. Biotechnol. 6 (1): 59-64 (1995); Allen et al., Trends Biochem. Sci. 20 (12): 511-6 (1995); Drees, Curr. Opin. Chem. Biol. 3 (1): 64-70 (1999); Topcu et al., Pharm. Res. 17 (9): 1049-55 (2000); Fashena et al., Gene 250 (1-2): 1-14 (2000); ; Colas et al., (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. Nature 380, 548-550; Norman, T. et al., (1999) Genetic selection of peptide inhibitors of biological pathways. Science 285, 591-595, Fabbrizio et al., (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. Oncogene 18, 4357-4363; Xu et al., (1997) Cells that register logical relationships among proteins. Proc Natl Acad Sci U S A 94, 12473-12478; Yang, et al., (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. Nuc. Acids Res. 23, 1152-1156; Kolonin et al., (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. Proc Natl Acad Sci U S A 95, 14266-14271; Cohen et al., (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. Proc Natl Acad Sci U S A 95, 14272-14277; Uetz, P.; Giot, L.; al, e. ;

10

20

30

40

50

Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, et al., (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574、これらの刊行物の全記載を、ここに引用して本明細書に組入れる。代表的に、このような融合は、大腸菌 (*E. coli*) LexAまたは酵母 GAL4 DNA結合性ドメインに対してなされる。核局限化シグナルに融合したベイトを発現する関連ベイトプラスミドを利用することができる。

10

【0082】

その他の有用な融合タンパク質は、上記されているように、ファージまたは細胞の表面にコードされたタンパク質の表示、内在性蛍光タンパク質（例えば、緑色蛍光タンパク質 (GFP)）への融合、および IgGFc 領域への融合を可能にする融合タンパク質を包含する（これらの説明を、その全体を引用しここに組入れる）。
本発明のポリペプチドおよび断片はまた、タンパク質毒素、例えばシュードモナス外毒素 A、ジフテリア毒素、シガ毒素 A、炭疽菌毒素致死因子、リシン (ricin) へ有用に融合することができ、これにより本発明のタンパク質を結合しまたは取り込む細胞の除去を行うことができる。

融合相手には、中でも、下記相手が包含される：myc、ヘマグルチニン (HA)、GST、免疫グロブリン、 β -ガラクトシダーゼ、ビオチン trpE、プロテイン A、 β -ラクタマーゼ、 α -アミラーゼ、 α -マルトース結合性タンパク質、アルコールデヒドロゲナーゼ、ポリヒスチジン（例えば、当該ポリペプチドのアミノおよび／またはカルボキシル末端の6個のヒスチジン）、lacZ、緑色蛍光タンパク質 (GFP)、酵母交配因子、GAL4 転写活性化または DNA 結合性ドメイン、ルシフェラーゼ、および血清タンパク質、例えばオボアルブミン、アルブミンおよび IgG の定常ドメイン。一例として、Ausubel による上記文献 (1992 年) および Ausubel による上記文献 (1999 年) を参照することができる。融合タンパク質はまた、特異的酵素開裂にかかわる部位、例えばファクター XIII、トリプシン、ペプシン、またはいずれかその他の当該技術分野で公知の酵素により認識される部位を含有することができる。上記したように、融合タンパク質は代表的に、当該技術分野で周知の技術（例えば、メリフィールド (Merrifield) 合成) を用いて化学的に合成されるか、または化学架橋により生成される組換え核酸法によって形成される。

20

30

融合タンパク質のもう一つの利点は、エピトープ標識を使用し、結合タンパク質または CSP に結合するその他の分子のスクリーニング用アフィニティ結合によって、融合タンパク質をプレートまたはカラムに結合させることができることにある。

以下でさらに説明するように、本発明による単離されたポリペプチド、突然変異タンパク質、融合タンパク質、相同性タンパク質または対立遺伝子変異体は、CSP、それらの対立遺伝子変異体および相同体を特異的に認識する抗体を作製するための特異的免疫原として容易に使用することができる。これらの抗体は次いで、中でも、本発明のポリペプチド、特に CSP にかかわる特異的試験（例えば血清などのタンパク質液体試料検出用 ELISA による、もしくは組織試料中のタンパク質検出用免疫組織化学またはレーザー走査サイトメトリーによる）、または細胞懸濁液中の細胞内タンパク質検出、または例えば免疫沈澱法によるなどの CSP の抗体媒介単離および／または精製、および CSP の特異的アゴニストまたはアンタゴニストとして使用するためのフローサイトメトリーによる分析に使用することができる。

40

【0083】

突然変異タンパク質、融合タンパク質、相同性タンパク質または対立遺伝子変異体を包含するポリペプチドが当該技術分野で公知の方法によって機能するかを測定することができる。一例として、機能を保有しながら、変化に耐性である残基は、当該技術分野で公知の

50

方法、例えばアラニンスキャニング突然変異誘発法 (Cunningham et al., Science 244 (4908): 1081-5 (1989)); トランスポゾンリンカースキャニング突然変異誘発法 (Chen et al., Gene 263 (1-2): 39-48 (2001)); ホモログーおよびアラニンスキャニング突然変異誘発法 (Jin et al., J. Mol. Biol. 226 (3): 851-65 (1992)); コンビナトリアルアラニンスキャニング法 (Weiss et al., Proc. Natl. Acad. Sci USA 97 (16): 8950-4 (2000)) を使用し、公知残基部位でタンパク質を変更し、引続き機能分析することにより、同定することができる。トランスポゾンリンカー走査キットは、市販されている (New England Biolabs, Beverly, MA, USA, カタログ番号 E7-102S; EZ::TN (登録商標) In-Frame Linker Insertion Kit, カタログ番号 EZ104KN, Epicentre Technologies Corporation, Madison, WI, USA)。

断片、相同性ポリペプチド、突然変異タンパク質、類縁体、誘導体および融合タンパク質を包含するポリペプチドの精製は周知であり、当業者の技術範囲内にある。例えば、Scopes, Protein Purification, 2d ed. (1987) を参照することができる。組換え発現ポリペプチドの精製は、上記で説明されている。化学合成されたペプチドの精製は、例えば HPLC によって容易に行うことができる。従って、本発明の単離タンパク質を、安定剤の存在または不存在下に、純粋な状態で、または実質的に純粋な状態で提供することは、本発明の1つの側面である。安定剤は、タンパク質様材料または非タンパク質様材料の両方を包含し、当該技術分野で周知である。アルブミンおよびポリエチレングリコール (PEG) などの安定剤は公知であり、また市販されている。

本発明による単離タンパク質を治療剤として、例えばワクチンに、また代用治療剤として使用する場合、高レベルの純度が好適であるが、本発明による単離タンパク質はまた、より低い純度でも有用である。一例として、本発明による部分的に精製されたタンパク質は、免疫原として使用することができ、これにより実験動物に抗体を産生させることができる。

好適態様において、本発明の精製されたタンパク質および実質的に精製されたタンパク質は、検出可能な両性電解質、アクリルアミドモノマー、ビス-アクリルアミドモノマーおよびポリアクリルアミドを欠く組成物の状態である。

【0084】

本発明によるポリペプチド、断片、類縁体、誘導体および融合体は、基体に有用に結合させることができる。この基体は、多孔質または非孔質の、平坦または非平坦形態であることができ、結合は共有または非共有状態であることができる。一例として、本発明によるポリペプチド、断片、類縁体、誘導体および融合体は、有孔基体、代表的にニトロセルロース、ポリビニリデンフルオライド (PVDF)、またはカチオン性に誘導体化されている親水性 PVDF を包含する慣用の膜に有用に結合させることができる; このように結合された本発明によるタンパク質、断片および融合体を使用し、例えば血清中の本発明の不動化タンパク質に対して特異的に結合する抗体を検出および定量することができる。

もう一つの例として、本発明によるポリペプチド、断片、類縁体、誘導体および融合体は、プラスチックなどの実質的に無孔基体に有用に結合させることができ、これにより本発明の不動化タンパク質に対して特異的に結合する抗体、例えば血清中のものを検出および定量することができる。このようなプラスチックには、ポメチルアクリル、ポリエチレン、ポリプロピレン、ポリアクリレート、ポリメチルメタアクリレート、ポリビニルクロライド、ポリテトラフルオロエチレン、ポリスチレン、ポリカーボネート、ポリアセタール、ポリスルホン、セルロースアセテート、セルロースナイトレート、ニトロセルロース、またはその混合物が包含される。この分析を標準的マイクロタイターディッシュで行

う場合、このプラスチックは代表的に、ポリスチレンである。

本発明によるポリペプチド、断片、類縁体、誘導体および融合体はまた、表面増強レーザー脱着イオン化源として使用するのに適する基体に結合させることもできる；このように結合された本発明によるタンパク質、断片または融合体は、表面結合タンパク質に対して十分な親和性または結合活性をもって結合する二次タンパク質の結合、引続く検出に有用であり、これによりこれらの間の生物学的相互作用が示される。本発明によるポリペプチド、断片および融合体はまた、表面プラズモン共鳴検出法で使用するのに適する基体に結合させることができる；このように結合された本発明によるタンパク質、断片または融合体は、表面結合タンパク質に対して十分な親和性または結合活性をもって結合する二次タンパク質の結合、引続く検出に有用であり、これによりこれらの間の生物学的相互作用が示される。 10

【0085】

抗体

他の側面において、本発明は、本発明による核酸分子によりコードされるポリペプチドに対し特異的に結合するその断片および誘導体を包含する抗体、ならびに当該ポリペプチドの断片、変異体、誘導体および類縁体に結合する抗体を提供する。好適態様において、これらの抗体は、CSPであるポリペプチド、またはその断片、変異体、誘導体、類縁体または融合タンパク質に対し特異的である。さらに好適な態様において、これらの抗体は、配列番号148～250を包含するポリペプチド、またはその断片、変異体、誘導体、類縁体または融合タンパク質に対し特異的である。 20

本発明による抗体は、そのネイティブな配座でタンパク質上に存在するか、またはいくつかの場合、例えばSDS中における可溶化によるなどして変性されたものとしてタンパク質上に存在する、線状エピトープ、不連続エピトープ、またはこのようなタンパク質またはタンパク質断片の立体配座エピトープに対し特異的であることができる。新規エピトープはまた、翻訳後修飾（PTM）についての疾患の組織対正常組織の差異による場合もある。一例として、CSPの特定の部位は、癌性細胞ではグリコシル化されている場合があるが、正常細胞ではグリコシル化されていない場合はなく、その逆もある。さらに、別のスプライス形態のCSPは癌を示唆することができる。CSPのC-またはN-末端の差別的分解はまた、抗癌治療用のマーカーまたは標的であることができる。一例として、CSPは、抗体が診断または治療用途で選択的に結合することができる新しいエピトープに暴露される癌細胞において分解されるN-末端であることができる。 30

当該技術分野で周知のように、抗体が混合物中の分子種中で識別することができる度合は、部分的に、混合物中の分子種の配座関連性に依存する；代表的に、本発明による抗体は、非CSPポリペプチドに対する優勢結合を越えて、少なくとも2倍、さらに代表的には少なくとも5倍、典型的に10倍、25倍、50倍、75倍以上、また多くの場合、100倍以上、また場合により、500倍または1000倍以上で識別するものとされる。本発明のタンパク質またはタンパク質断片の検出に使用する場合、この抗体をヒト大腸に由来する試料中の本発明によるタンパク質の存在の決定に使用することができるならば、本発明の抗体は十分に特異的である。

【0086】

代表的に、本発明によるタンパク質またはタンパク質断片に対する本発明の抗体（またはIgMペンタマーの場合などでは抗体マルチマー）の親和性または結合活性は、少なくとも約 1×10^{-6} モル（M）、代表的に少なくとも約 5×10^{-7} Mであり、少なくとも約 1×10^{-8} M、 5×10^{-9} M、 1×10^{-10} M、および 1×10^{-13} Mまでの親和性および結合活性は特別の有用性を提供する。 40

本発明の抗体は、いずれかの鳥類、爬虫類、または哺乳動物種からの天然産生形態、例えばIgG、IgM、IgD、IgE、IgY、およびIgAであることができる。

頻度は低いが、ヒト抗体はヒトドナーまたはヒト細胞から直接採取することができる。この場合、本発明のタンパク質に対する抗体は代表的に、本発明のタンパク質またはタンパク質断片による偶発的免疫感作、例えば自己免疫免疫感作から生成される。頻度は低いが 50

、このような抗体は代表的に、ポリクローナルである。さらに、各ポリクローナル抗体は、単離し、次いでクローン化し、モノクローナルを生成することができる。

ヒト抗体は、ヒト免疫グロブリン遺伝子を発現するトランスジェニック動物を用いてさらにしばしば得られ、このトランスジェニック動物は本発明のタンパク質免疫原により肯定的に (affirmatively) 免疫感作することができる。ヒト抗体を産生することができるヒト Ig トランスジェニックマウスおよびそこからのヒト抗体の産生方法は、中でも、下記刊行物に開示されており、これらの刊行物を、それらの全体の引用によってここに組入れる：米国特許第 6, 162, 963 号、同第 6, 150, 584 号、同第 6, 114, 598 号、同第 6, 075, 181 号、同第 5, 939, 598 号、同第 5, 877, 397 号、同第 5, 874, 299 号、同第 5, 814, 318 号、同第 5, 789, 650 号、同第 5, 770, 429 号、同第 5, 661, 016 号、同第 5, 633, 425 号、同第 5, 625, 126 号、同第 5, 569, 825 号、同第 5, 545, 807 号、同第 5, 545, 806 号および同第 5, 591, 669 号。このような抗体は代表的に、モノクローナルであり、代表的にマウス抗体の作製にかかわり開発された技術を用いて作製される。

投与された抗体に対するレシピエントの免疫応答はしばしば、別種、例えばマウスから誘導された抗体の投与による場合に比較して実質的に小さいことから、本発明による抗体を *in vivo* 診断または治療剤としてヒトに投与する場合、ヒト抗体が特に有用であり、またしばしば好適である。

【0087】

本発明による Ig G、Ig M、Ig D、Ig E、Ig Y、および Ig A 抗体はまた、齧歯類（代表的にマウス、およびラット、モルモットおよびハムスター）、ウサギ目（代表的にウサギ）、およびまたさらに大型の哺乳動物（例えば、ヒツジ、ヤギ、ウシ、およびウマ）、および別種の産卵する鳥類または爬虫類（例えば、ニワトリまたはワニ）を包含する別種から得ることもできる。一例として、鳥類抗体は 2000 年 5 月 25 日付けで公開された WO 00/29444 に記載の技術を用いて産生させることができる（これらの刊行物を、その全体の引用によって、ここに組入れる）。このような場合、トランスジェニックヒト抗体産生性非ヒト哺乳動物を用いると、偶発的免疫感作は不必要であり、またこの非ヒト哺乳動物は代表的に、本発明のタンパク質またはタンパク質断片を用い、標準的免疫感作方法に従い肯定的に免疫感作させることができる。

上記したように、キャリア、代表的にウシチログロブリン、キーホールリンペットヘモシアニン、またはウシ血清アルブミンなどのタンパク質に共有結合させる場合、上記いずれかの刊行物（これらの記載を、その全体の引用によってここに組入れる）に開示されているような二官能性リンカーを都合良く使用して、本発明のタンパク質の 8 個またはそれ以上の隣接アミノ酸を免疫原として効果的に使用することができる。

免疫原性はまた、本発明によるタンパク質および断片を別の分子に融合させることによって授与することができる。一例として、本発明のペプチドは、分枝したポリリジンコアマトリクスに対する固相合成により生成させることができる；これらの多重抗原性ペプチド (MAP) は、ワクチン作製において、高い純度、増大した活性、正確な化学的定義、および改良された安全性を備えている。Tam et al., Proc. Natl. Acad. Sci. USA 85: 5409-5413 (1988); Posnett et al., J. Biol. Chem. 263: 1719-1725 (1988)。

非ヒト哺乳動物または鳥類を免疫感作するための方法は、当該技術分野で十分に確立されている。下記刊行物を参照することができ、これらの記載を引用してここに組入れる：Harlow et al. (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan et al. (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies

: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck J. Dtsch. Tierarztl. Wochenschr. 103: 417-422 (1996)。免疫感作法はしばしば、フロイント完全アジュバントおよびフロイント不完全アジュバントなどのアジュバントを用いるか、用いない、多重免疫感作を包含し、また裸のDNA免疫感作を包含することができる(Moss, Semin. Immunol. 2: 317-327 (1990))。

【0088】

非ヒト哺乳動物および鳥類からの抗体は、ポリクローナルまたはモノクローナルであることができ、ポリクローナル抗体は本発明によるタンパク質の免疫組織化学的検出において或る利点を有し、他方、モノクローナル抗体は本発明によるタンパク質の特定のエピトープの同定および識別に利点を有する。鳥類からの抗体は、ヒト血清または組織における本発明によるタンパク質の検出に特別の利点を有することがある(Viking et al., Biosens. Bioelectron. 13: 1257-1262 (1998))。

免疫感作後、本発明の抗体は、いずれか当該技術分野で許容される技法を用いて産生させることができる。このような技術は当該技術分野で周知であり、下記刊行物を参照することができ、これらの記載を、それらの全体の引用によりここに組入れることから、ここでは詳細は不必要である: Coligan, 前出; Zola, supra; Howard et al. (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, 前出; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997)。

しかしながら、簡単に説明すると、このような技術は、中でも、ハイブリドーマによるモノクローナル抗体の生成および免疫グロブリン遺伝子またはその断片を発現するように操作された宿主細胞からの抗体もしくはその断片または誘導体の発現を包含する。これら2種の生成方法は、相互に排他的ではない: 本発明によるタンパク質またはタンパク質断片に特異的な抗体をコードする遺伝子は、ハイブリドーマからクローン化することができ、次いで別の宿主細胞で発現させることができる。これら2種の方法を一緒に行う必要はない。例えば、本発明によるタンパク質またはタンパク質断片に特異的な抗体をコードする遺伝子は、米国特許第5,627,052号(この記載を、その全体の引用によりここに組入れる)に詳細に記載されているように、所望のタンパク質に特異的であることが知られているB細胞から、または抗体-ディスプレイファージから、直接にクローン化することができる。

宿主細胞における組換え発現は、本発明による抗体の断片または誘導体が所望される場合、特に有用である。

完全抗体、抗体断片または抗体誘導体の組換え生成用の宿主細胞は、真核生物または原核生物であることができる。

原核生物宿主は、本発明のファージディスプレイ抗体の生成に特に有用である。

【0089】

ファージディスプレイ抗体技法は現時点で十分に確立されており、この方法では、抗体可変領域断片を、例えば遺伝子IIIタンパク質(pIII)または遺伝子VIIタンパク質(pVII)に融合させ、M13などの線状ファージの表面上に表示させる。例え

ば、下記刊行物を参照することができる：Sidhu, Curr. Opin. Biotechnol. 11 (6) : 610-6 (2000); Griffiths et al., Curr. Opin. Biotechnol. 9 (1) : 102-8 (1998); Hoogenboom et al., Immunotechnology, 4 (1) : 1-20 (1998); Rader et al., Current Opinion in Biotechnology 8 : 503-508 (1997); Aujame et al., Human Antibodies 8 : 155-168 (1997); Hoogenboom, Trends in Biotechnol. 15 : 62-70 (1997); de Kruif et al., 17 : 453-455 (1996); Barbas et al., Trends in Biotechnol. 14 : 230-234 (1996); Winter et al., Ann. Rev. Immunol. 433-455 (1994)。最近になって、このようなライブラリーからの抗体断片の産生、増殖、選別 (pan)、および使用に要する技術および方法は承認されている。一例として下記刊行物を参照することができ、これらの記載を、その全体の引用によりここに組入れる：Barbasによる上記文献 (2001年) ; Kayによる上記文献 ; Abelsonによる上記文献。

代表的に、ファージディスプレイ抗体断片は、scFv断片またはFab断片であり、所望により、ディスプレイファージから完全抗体への可変領域のクローン化および追加の真核生物または原核生物宿主細胞における十分な長さを有する抗体の発現により、十分な長さの抗体を生成させることができる。

真核細胞はまた、本発明による抗体、抗体断片および抗体誘導体の発現に有用であることができる。

一例として、本発明の抗体断片は、ピチア パストリス (*Pichia pastoris*) およびサッカロマイセス セレビシアエ (*Saccharomyces cerevisiae*) で生成させることができる。例えば、下記刊行物を参照することができ、これらの記載を、それら全体の引用によりここに組入れる：Takahashi et al., Biosci. Biotechnol. Biochem. 64 (10) : 2138-44 (2000); Freyre et al., J. Biotechnol. 76 (2-3) : 157-63 (2000); Fischer et al., Biotechnol. Appl. Biochem. 30 (Pt 2) : 117-20 (1999); Pennell et al., Res. Immunol. 149 (6) : 599-603 (1998); Eldin et al., J. Immunol. Methods. 201 (1) : 67-75 (1997); Frenken et al., Res. Immunol. 149 (6) : 589-99 (1998); Shusta et al., Nature Biotechnol. 16 (8) : 773-7 (1998)。

本発明の抗体断片および誘導体を包含する抗体は、昆虫細胞で生成させることもできる。例えば、下記刊行物を参照することができ、これらの記載を、それら全体の引用によりここに組入れる：Li et al., Protein Expr. Purif. 21 (1) : 121-8 (2001); Ailor et al., Biotechnol. Bioeng. 58 (2-3) : 196-203 (1998); Hsu et al., Biotechnol. Prog. 13 (1) : 96-104 (1997); Edelman et al., Immunology 91 (1) : 13-9 (1997); および Nesbit et al., J. Immunol. Methods 151 (1-2) : 201-8 (1992)。

【0090】

本発明による抗体、ならびにその断片および誘導体はまた、植物細胞、特にトウモロコシまたはタバコにおいて産生させることもでき、下記刊行物を参照することができ、これらの記載を、それら全体の引用によりここに組入れる：Giddings et al., Nature Biotechnol. 18 (11) : 1151-5 (2000);

Gavilondo et al., *Biotechniques* 29 (1): 128-38 (2000); Fischer et al., *J. Biol. Regul. Homeost. Agents* 14 (2): 83-92 (2000); Fischer et al., *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer et al., *Biol. Chem.* 380 (7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); および Ma et al., *Plant Physiol.* 109 (2): 341-6 (1995)。

本発明による抗体断片および誘導体を包含する抗体は、トランスジェニック非ヒト哺乳動物ミルクで産生させることもできる。例えば、下記刊行物を参照することができ、これらの記載を、それらの全体の引用によりここに組入れる: Pollock et al., *J. Immunol. Methods* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); Limonta et al., *Immunotechnology* 1: 107-13 (1995)。

本発明による抗体、抗体断片および抗体誘導体の組換え発現に有用な哺乳動物細胞は、CHO細胞、COS細胞、293細胞および骨髓腫細胞を包含する。

Verma et al., *J. Immunol. Methods* 216 (1-2): 165-81 (1998) (これらの記載を引用してここに組入れる) には、抗体の発現にかかわる細菌、酵母、昆虫および哺乳動物発現系を発表し、比較している。

本発明による抗体はまた、Merk et al., *J. Biochem. (Tokyo)* 125 (2): 328-33 (1999) および Ryabova et al., *Nature Biotechnol.* 15 (1): 79-84 (1997) にさらに説明されているように、無細胞翻訳によって生成させることもでき、また Pollock et al., *J. Immunol. Methods* 231 (1-2): 147-57 (1999) にさらに説明されているように、トランスジェニック動物のミルクで生成させることもできる (これらの記載を、それらの全体の引用によりここに組入れる)。

本発明はさらに、1種または2種以上の本発明によるタンパク質およびタンパク質断片に特異的に結合する抗体断片、本発明の単離核酸によりコードされた1種または2種以上のタンパク質およびタンパク質断片に特異的に結合する抗体断片、もしくはその結合を1種または2種以上の本発明のタンパク質およびタンパク質断片により、または本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片により競合的に阻害することができる抗体断片を提供する。

このような有用な断片の中には、Fab、Fab'、Fv、F(ab)'₂ および単鎖Fv(scFv)断片がある。その他有用な断片は、Hudson, *Curr. Opin. Biotechnol.* 9 (4): 395-402 (1998) に記載されている。

【0091】

もう一つの側面において、本発明は、本発明によるタンパク質およびタンパク質断片に特異的に結合する抗体誘導体、本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片に特異的に結合する抗体誘導体、もしくはその結合を1種または2種以上の本発明のタンパク質およびタンパク質断片により、または本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片により競合的に阻害することができる抗体誘導体を提供する。

このような有用な誘導体の中には、キメラ、霊長類化、およびヒト化抗体があり、このような誘導体はヒトにおいては免疫原性に乏しく、従って非ヒト哺乳動物種からの未修飾抗体に比較して、*in vivo*投与に適している。

キメラ抗体は代表的に、1種の動物、代表的にヒトの定常領域に融合された別の1種の動

物、代表的にマウスの免疫グロブリンの重鎖および／または軽鎖可変領域（CDRおよびフレームワーク残部（residues）を包含する）を含有する。例えば、下記刊行物を参照することができ、これらの記載を、それら全体の引用によりここに組入れる：米国特許第5,807,715号；Morrison et al., Proc. Natl.

Acad. Sci USA. 81(21): 6851-5 (1984)；Sharone et al., Nature 309(5966): 364-7 (1984)；Takeda et al., Nature 314(6010): 452-4 (1985)。霊長類化およびヒト化抗体は代表的に、非霊長類またはヒト抗体のV領域フレームワーク中にグラフトされたマウス抗体からの重鎖および／または軽鎖CDRを包含し、通常ヒト定常領域をさらに包含する。下記刊行物を参照することができ、これらの記載を、それら全体の引用によりここに組入れる：Riechmann et al., Nature 332(6162): 323-7 (1988)；Co et al., Nature 351(6326): 501-2 (1991)；米国特許第6,054,297号；同第5,821,337号；同第5,770,196号；同第5,766,886号；同第5,821,123号；同第5,869,619号；同第6,180,377号；同第6,013,256号；同第5,693,761号；および同第6,180,370号。

本発明によるもう1種の有用な抗体誘導体は、ヘテロマー状抗体複合体および抗体融合体、例えばダイアボディ（二重特異性抗体）、単鎖ダイアボディ、およびイントラボディ（intrabodies）を包含する。

【0092】

本発明の抗体をコードする核酸は、本発明の抗体のクローン化または発現用の組換えベクターを形成する別の核酸に作動可能に結合させることができるものと考えられる。本発明は、真核生物の形質導入用であるか、トランスフェクション用であるかまたは遺伝子治療用であるかに関係なく、コード配列またはその一部を含有する全部の組換えベクターを包含する。このようなベクターは、当業者に公知である慣用の分子生物学的技術を用いて調製することができ、またフレームワークおよびCDRまたはその一部を包含する免疫グロブリンV領域用のDNAコード配列、および細胞内移送用のシグナル配列を伴うか、または伴わない適当なプロモーターを含有することができる。このようなベクターは、当業者に公知である慣用の技術によって真核細胞中に形質導入することができ、またはトランスフェクションすることができる（Marasco et al., Proc. Natl. Acad. Sci. (USA) 90: 7889-7893 (1993)；Duan et al., Proc. Natl. Acad. Sci. (USA) 91: 5075-5079 (1994)）。

その断片および誘導体を包含する本発明による抗体は、有用に標識することができる。従って、もう一つの態様において、本発明は1種または2種以上の本発明によるタンパク質およびタンパク質断片に特異的に結合する標識した抗体、本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片に特異的に結合する標識した抗体、もしくはその結合を1種または2種以上の本発明のタンパク質およびタンパク質断片により、または本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片により競合的に阻害することができる標識した抗体を提供する。この標識の選択は、部分的に、所望の用途に依存する。

一例として、本発明の抗体を組織試料の免疫組織化学的染色に用いる場合、この標識は好ましくは、検出可能な生成物の生成および局所的沈着を触媒する酵素である。

抗体に共有結合し、それらの免疫組織化学的可視化を可能にする代表的酵素は周知であり、アルカリホスファターゼ、 β -ガラクトシダーゼ、グルコースオキシダーゼ、西洋ワサビペルオキシダーゼ（HRP）およびウレアーゼを包含する。目視で検出できる生成物の生成および沈着用の代表的基質は、 α -ニトロフェニル- β -D-ガラクトピラノシド（ONPG）； α -フェニレンジアミンジヒドロクロライド（OPD）； p -ニトロフェニルホスフェート（PNPP）； p -ニトロフェニル- β -D-ガラクトピラノシド（PN

10

20

30

40

50

PG) ; 3', 3'-ジアミノベンジジン (DAB) ; 3-アミノ-9-エチルカルバゾール (AEC) ; 4-クロロ-1-ナフトール (CN) ; 5-ブロモ-4-クロロ-3-インドリル-ホスフェート (BCIP) ; ABTS (登録商標) ; Blu o Gal ; ヨードニトロテトラゾリウム (INT) ; ニトロブルーテトラゾリウムクロライド (NBT) ; フェナジンメソサルフェート (PMS) ; フェノールフタレインモノホスフェート (PMP) ; テトラメチルベンジジン (TMB) ; テトラニトロブルーテトラゾリウム (TNBT) ; X-Gal ; X-Gluc ; および X-グルコシドを包含する。

【0093】

別種の基質を用いて、発光性である局所沈着生成物を生成させることもできる。一例として、過酸化水素 (H_2O_2)、西洋ワサビペルオキシダーゼ (HRP) を存在させ、環状ジアシルヒドラジド化合物、例えばルミノールの酸化を触媒させることができる。この酸化の直後、ルミノールは励起状態 (中間反応生成物) であり、光を放射することによって基底状態に減衰する。この光放射の強力な増強は、増強剤、例えばフェノール系化合物によってもたらされる。利点には、放射線を用いることなく、また少量のみの抗体を要して、高感度、高解像力、および迅速な検出が得られることが包含される。一例として、下記刊行物を参照することができ、これらの記載を、それらの全体の引用によりここに組入れる: Thorpe et al., Methods Enzymol. 133: 331-53 (1986) ; Kricka et al., J. Immunoassay 17 (1): 67-83 (1996) ; および Lundqvist et al., J. Biolumin. Chemilumin. 10 (6): 353-9 (1995)。このような強化化学発光検出 (ECL) 用キットは市販されている。

抗体はまた、コロイド状金を用いて標識することができる。

もう一つの例として、本発明の抗体を、例えばフローサイトメトリーによる検出、走査性レーザーサイトメトリーによる検出、または蛍光免疫分析に使用する場合、これらはフルオロフォアにより標識すると有用であることができる。

本発明の抗体に有用に結合させることができるフルオロフォア標識には、広く種々のフルオロフォアが存在する。

フローサイトメトリー用途の場合、細胞外検出用および細胞内検出用の両方において、共通して有用なフルオロフォアは、フルオレセインイソチオシアネート (FITC) ; アロフィコシアニン (APC) ; R-フィコエリトリン (PE) ; ペリジニクロロフィルタンパク質 (PerCP) ; テキサスレッド (Texas Red) ; Cy3、Cy5、蛍光共鳴エネルギータンデムフルオロフォア、例えば、PerCP-Cy5. 5、PE-Cy5、PE-Cy5. 5、PE-Cy7、PE-テキサスレッド、および APC-Cy7 であることができる。

別種のフルオロフォアは、中でも、アレクサフルオール (登録商標) 350、アレクサフルオール (登録商標) 488、アレクサフルオール (登録商標) 532、アレクサフルオール (登録商標) 546、アレクサフルオール (登録商標) 568、アレクサフルオール (登録商標) 594、アレクサフルオール (登録商標) 647 (Molecular Probes, Inc. Eugene, OR, USA から入手することができるモノクロール抗体標識キット) ; BODIPY 染料、例えば BODIPY 493/503、BODIPY FL、BODIPY R6G、BODIPY 530/550、BODIPY TMR、BODIPY 558/568、BODIPY 558/568、BODIPY 564/570、BODIPY 576/589、BODIPY 581/591、BODIPY TR、BODIPY 630/650、BODIPY 650/665、カスケードブルー、カスケードイエロー、ダンシル、リッサミンローダミン B (Lissamine rhodamine B)、マリナブルー、オレゴングリーン 488、オレゴングリーン 514、パシフィックブルー、ローダミン 6G、ローダミングリーン、ローダミンレッド、テトラメチルローダミン、テキサスレッド (Molecular Probes, Inc. Eugene, OR, USA から入手することができる)、および Cy2、Cy3、Cy3. 5、Cy5、Cy5. 5、Cy7 を包含し、これら全部が本発明の抗体の蛍

光標識に有用である。

【0094】

標識したアビジン、ストレプトアビジン、カプトアビジンまたはニュートラアビジンを用いる二次検出の場合、本発明の抗体はビオチンにより有用に標識することができる。

本発明による抗体を、例えばウエスタンブロット用途に用いる場合、これらは放射性同位元素、例えば ^{33}P 、 ^{32}P 、 ^{35}S 、 ^3H 、および ^{125}I により有用に標識することができる。

別の例として、本発明による抗体を放射性免疫治療に用いる場合、この標識は有利には、 ^{228}Th 、 ^{227}Ac 、 ^{225}Ac 、 ^{223}Ra 、 ^{213}Bi 、 ^{212}Pb 、 ^{212}Bi 、 ^{211}At 、 ^{203}Pb 、 ^{194}Os 、 ^{188}Re 、 ^{186}Re 、 ^{153}Sm 、 ^{149}Tb 、 ^{131}I 、 ^{125}I 、 ^{111}In 、 ^{105}Rh 、 $^{99\text{m}}\text{Tc}$ 、 ^{97}Ru 、 ^{90}Y 、 ^{90}Sr 、 ^{88}Y 、 ^{72}Se 、 ^{67}Cu または ^{47}Sc であることができる。

もう一つの例として、本発明による抗体をin vivo診断用途に用いる場合、これらはMRI造影剤、例えばガドリニウムジエチレントリアミンペンタ酢酸(DTPA)に共有結合させることによって(Lauffer et al., Radiology 207 (2): 529-38 (1998))、または放射性同位元素標識することによって検出可能にすることができる。

理解されるように、上記標識の使用は、その用途が記載されている用途に制限されない。その断片および誘導体を包含する本発明による抗体はまた、毒素に結合させることができ、これにより本発明によるタンパク質を表示および/または発現する細胞に対して毒素の除去作用を標的化することができる。通常、このようなイムノトキシンにおける抗体は、シュードモナス外毒素A、ジフテリア毒素、シガ毒素A、炭疽菌毒素致死因子、またはリシンに結合させる。下記刊行物を参照することができ、これらの記載を、それらの全体の引用によりここに組入れる: Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); および Frankel et al. (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998)。

本発明による抗体は、基体に有利に結合させることができ、従ってもう一つの側面において、本発明は、基体に結合されている1種または2種以上の本発明によるタンパク質およびタンパク質断片に特異的に結合する抗体、本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片に特異的に結合する抗体、もしくはその結合を1種または2種以上の本発明のタンパク質およびタンパク質断片により、または本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片により競合的に阻害することができる抗体を提供する。

【0095】

基体は、多孔質または非多孔質で平坦または非平坦であることができる。

一例として、本発明による抗体は、濾過媒体、例えば免疫親和性クロマトグラフィー目的用のNH₂-活性セファロースまたはCNBr-活性化セファロースに有用に結合させることができる。

一例として、本発明による抗体は、代表的にビオチン-ストレプトアビジン相互反応によって、常磁性微小球体に結合させることができ、この微小球体は次いで、本発明によるタンパク質を発現または表示する細胞の単離に使用することができる。もう一つの例において、本発明による抗体は、ELISA用の微量滴定板の表面上に有用に結合させることができる。

上記したように、本発明による抗体は真核生物および原核生物細胞で産生することができる。従って、もう一つの側面において、本発明は本発明による抗体を発現する細胞を提供し、これら細胞にはハイブリドーマ細胞、B細胞、形質細胞、および本発明による抗体を発現させるために組換えにより修飾された宿主細胞が包含される。

さらにもう一つの側面において、本発明は、1種または2種以上の本発明によるタンパク

10

20

30

40

50

質およびタンパク質断片に対する特異的結合、本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片に対する特異的結合、もしくはその結合を1種または2種以上の本発明のタンパク質およびタンパク質断片により、または本発明の単離核酸によりコードされる1種または2種以上のタンパク質およびタンパク質断片により競合的に阻害することができる特異的結合を惹起するアプタマーを提供する。

要約すると、本発明の教示が提供されることによって、当業者は本発明による抗体の生物学的性質の変更に使用することができる、所定の抗体分子の安定性または寿命、免疫原性、毒性、親和性または収率を増加または減少させる方法を含む、種々の方法、または特定の用途にさらに適するようにすることができるいずれか別の方法によってそれを変更する方法を入手することができる。

10

【0096】

トランスジェニック動物および細胞

他の側面では、本発明は、本発明の核酸分子を含むトランスジェニック細胞および非ヒト生物を提供する。好ましい態様では、トランスジェニック細胞および非ヒト生物は、CSPをコードする核酸分子を含む。好ましい態様では、CSPは配列番号148～250から選択されるアミノ酸配列、またはその断片、ミューテイン、相同タンパク質または対立遺伝子変異体を含む。他の好ましい態様では、トランスジェニック細胞および非ヒト生物は、本発明のCSNA、好ましくは配列番号1～147からなる群から選択される核酸配列、またはその部分、実質的に類似の核酸分子、対立遺伝子変異体もしくはハイブリダイズする核酸分子を含む、CSNAを含む。

20

他の態様では、トランスジェニック細胞および非ヒト生物は、ヒトCSGの内在性オーソログの標的化された破壊または置換を有する。トランスジェニック細胞は、胚性幹細胞または体細胞であることができる。トランスジェニック非ヒト生物は、キメラ、非キメラヘテロ接合体、および非キメラホモ接合体であることができる。トランスジェニック動物を作製する方法は、当該技術分野でよく知られている。例えば、Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999)、Jackson et al., Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000) および Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999) を参照。

30

本発明の核酸分子を動物に導入し、トランスジェニック動物の創始系統を作製するために、当該技術分野で知られている任意の技術を用いることができる。かかる技法としては、前核 (pronuclear) マイクロインジェクション (例えば Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994)、Carver et al., Biotechnology (NY) 11: 1263-1270 (1993)、Wright et al., Biotechnology (NY) 9: 830-834 (1991) および米国特許第4, 873, 191号 (1989))、生殖系列、胚盤胞または胚へのレトロウイルス媒介性遺伝子移入 (例えば、Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985) を参照)、胚性幹細胞における遺伝子ターゲティング (例えば、Thompson et al., Cell 56: 313-321 (1989) を参照)、細胞または胚のエレクトロポレーション (例えば、Lo, 1983, Mol Cell Biol. 3: 1803-1814 (1983) を参照)、遺伝子銃を用いた導入 (例えば、Ulmer et al., Science 259: 1745 (1993) を参照)、胚性多能性幹細胞への核酸構築物の導入および胚盤胞への幹細胞の移し戻し、および精子媒介性遺伝子移入 (例えば、Lavitrano et al., Cell 57: 717-723 (1989) を参照) が挙げられるが、これらに限定されない。

40

50

【0097】

他の技術としては、例えば、静止状態にした培養胚、胎児または成人細胞からの核の、除核卵母細胞への核移入が挙げられる（例えば、Campbell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810-813 (1997)を参照）。本発明は、そのすべての細胞に導入遺伝子（即ち、本発明の核酸分子）を持つトランスジェニック動物、ならびに、その細胞のすべてではないが、いくつかに移入遺伝子を持つ動物、即ちモザイク動物またはキメラ動物を提供する。

導入遺伝子は、単一の導入遺伝子として、または、コンカテマー状、例えば頭－頭タンデムまたは頭－尾タンデム状に、複数のコピーとして組込むことができる。導入遺伝子はまた、たとえばLasko et al., et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)の教示にしたがい、特定の細胞タイプに選択的に導入し、活性化させることができる。かかる細胞タイプ特異的な活性化に必要な調節配列は、対象となる特定の細胞タイプに依存し、それは当業者には明らかであろう。

トランスジェニック動物が一旦創出されれば、組換え遺伝子の発現は、標準的な技法を利用してアッセイされ得る。初期スクリーニングは、導入遺伝子の組み込みが起ったことを確認すべく動物組織を分析するために、サザンブロット分析またはPCR技法により達成され得る。トランスジェニック動物の組織における導入遺伝子のmRNA発現レベルはまた、動物から得られた組織試料のノーザンブロット分析、in situハイブリダイゼーション分析、および逆転写酵素PCR（RT-PCR）が含まれるが、これらに限定されない技法を用いて評価してもよい。トランスジェニック遺伝子発現組織の試料はまた、導入遺伝子産物に特異的な抗体を用いて、免疫細胞化学的に、または免疫組織化学的に評価してもよい。

創始動物が一旦作製されれば、それらを、繁殖、同系交配、異系交配、または交雑させて、特定の動物コロニーを作製してもよい。かかる繁殖戦略の例としては、別個の系統を樹立するために、1つよりも多い組み込み部位を有する創始動物を異系交配すること、各導入遺伝子の付加的発現の効果が理由で、高いレベルで導入遺伝子を発現する複合遺伝子導入をもたらすために、別個の系統を同系交配させること、発現を増大させ、かつDNA分析による動物のスクリーニングの必要性をなくすために、対象となる組み込み部位に関してホモ接合性の動物を生産するために、ヘテロ接合性のトランスジェニック動物を交雑させること、複合ヘテロ接合性またはホモ接合性系統を生産するために、別個のホモ接合性系統を交雑させること、および対象となる実験モデルに適切な、明瞭なバックグラウンドに導入遺伝子を配置させるために繁殖させることが挙げられるが、これらに限定されない。

【0098】

本発明のトランスジェニック動物は、本発明のポリペプチドの生物機能を説明し、異常発現に関連した症状および／または疾患を研究し、かかる症状および／または疾患を改善するのに効果的な化合物をスクリーニングするのに有用な動物モデル系を含むが、これに限定されない用途を有する。

トランスジェニック動物を標的遺伝子の破壊によって創出する方法も、当該技術分野でよく知られている。一般的に、内在性標的遺伝子に相同ないくつかのヌクレオチド配列を含むようにベクターを設計する。ベクターを細胞に導入し、それが、染色体の配列との相同組換えを介して、内在性遺伝子に組込まれ、それによって内在性遺伝子の機能を破壊できるようにする。導入遺伝子はまた、特定の細胞タイプに選択的に導入することができ、こうしてこの細胞タイプでのみ、内在性遺伝子が不活性化される。かかる細胞タイプ特異的な不活性化に必要な調節配列は、対象となる特定の細胞タイプに依存しており、当業者には明らかだろう。例えば、Smithies et al., Nature 317: 230-234 (1985)、Thomas et al., Cell 51: 503-512 (1987)、Thompson et al., Cell 5: 313-321

(1989)を参照。

1つの態様では、選択マーカーおよび／またはネガティブ選択マーカーを用いて、あるいは用いずに、内在性核酸配列（遺伝子のコード領域または調節領域）に相同的なDNAに隣接した突然変異体である本発明の非機能性核酸分子（または完全に非関連のDNA配列）を使用して、*in vivo*で本発明のポリペプチドを発現する細胞をトランスフェクトすることができる。別の態様では、当該技術分野にて既知の技法を用いて、対象となる遺伝子を含有するが、発現はしない細胞におけるノックアウトを創出する。標的相同組換えを介したDNA構築物の挿入は、結果として標的遺伝子の不活性化を招く。かかるアプローチは、研究および農業の分野に特に適しており、そこで胚性幹細胞の改変によって、不活性な標的遺伝子を有する動物の子孫を創出することができる。例えば、上記Thomasおよび上記Thompsonを参照。しかしながら、組換えDNA構築物が、当業者には明らかであろう適切なウイルスベクターを用いて、必要とされる部位に*in vivo*で直接的に投与されるか、または標的化されるという条件で、このアプローチは、ヒトにおける使用にルーチンで適合され得る。

【0099】

本発明のさらなる態様では、本発明のポリペプチドを発現するように遺伝子操作された細胞、あるいは本発明のポリペプチドを発現しないように遺伝子操作された（例えば、ノックアウト）細胞は、患者に*in vivo*で投与される。かかる細胞は、動物、または患者、またはMHC適合ドナーから得てもよく、線維芽細胞、骨髄細胞、血液細胞（例えば、リンパ球）、脂肪細胞、筋細胞、内皮細胞などを含むことができるが、これらに限定されない。細胞は、組換えDNA技法を用いて、*in vitro*で遺伝子操作され、細胞に本発明のポリペプチドのコード配列を導入するか、あるいは例えば形質導入（ウイルスベクター、および好ましくは細胞ゲノムに導入遺伝子を組み込むベクターを用いて）、またはプラスミド、コスミド、YAC、裸のDNA、エレクトロポレーション、リポソーム等の使用を含むが、これらに限定されないトランスフェクション手法により、本発明のポリペプチドに関するコード配列および／または内在性調節配列を崩壊させる。

本発明のポリペプチドのコード配列は、強力な構成型もしくは誘導型プロモーターまたはプロモーター／エンハンサーの制御下に置くことにより、本発明のポリペプチドの発現、好ましくは分泌を達成することができる。本発明のポリペプチドを発現、好ましくは分泌する操作細胞を全身的に、例えば、循環にて、または腹腔内にて患者に導入することができる。

あるいは、細胞をマトリックス中に組み込み、体内へ移植することができ、例えば、遺伝子操作した線維芽細胞を、皮膚移植片の一部として移植することができ、または、遺伝子操作した内皮細胞を、リンパ管または血管移植片の一部として移植することができる。例えば、各々その全体が参照により本明細書に組込まれる米国特許第5,399,349号および第5,460,959号を参照。

投与されるべき細胞が非自系または非MHC適合細胞である場合、それらは、導入細胞に対する宿主免疫応答の発生を防ぐ既知の技法を用いて、投与され得る。例えば、細胞は、直に接している細胞外環境との構成成分の交換が可能であるものの、導入細胞が宿主免疫系により認識されることが不可能なカプセル化形態で導入されてもよい。

本発明のトランスジェニックおよび「ノックアウト」動物は、本発明のポリペプチドの生物機能を説明し、異常な発現に関連した症状および／または疾患を研究し、かかる症状および／または疾患を改善するのに効果的な化合物をスクリーニングするのに有用な動物モデル系を含むが、これらに限定されない用途を有する。

【0100】

コンピュータ可読な手段

本発明のもう一つの側面は、本発明の核酸およびアミノ酸配列を保存するためのコンピュータ可読な手段に関する。好適態様において、本発明は、本明細書に記載されている配列番号1～147および配列番号148～250の一组の完全配列またはいずれかの組合せ配列を保存するためのコンピュータ可読な手段を提供する。このコンピュータ可読な手段

の記録は、読取りおよび表示のため、ならびに、或る基準に適合するデータ、配列の比較、一組の基準に適合する配列の整列または順序などにかかわる質問に対するデータのロケーションを可能にするプログラムを適用するためのコンピュータに接続させるためにアクセスすることができる。

本発明の核酸およびアミノ酸配列は、検索分析に有用であるデータベースにおける、また配列分析アルゴリズムにおける構成要素として特に有用である。本明細書で用いられている「本発明の核酸配列」および「本発明のアミノ酸配列」の用語は、コンピュータ可読な形態に縮小されるか、または縮小することができるか、もしくは保存することができる、本発明によるポリヌクレオチドまたはポリペプチドのいずれかの検出可能な化学的または物理的特徴を意味する。これらに制限されないものとして、これらにはクロマトグラフィ

10

ー走査データまたはピークデータ、写真データまたはそこからの走査データ、および質量分析のデータが包含される。

本発明は、本発明による配列がそこに保存されているコンピュータ可読な媒体を提供する。コンピュータ可読な媒体は、1種または2種以上の下記のことを包含することができる：本発明の核酸配列の配列を包含する核酸配列；本発明のアミノ酸配列の配列を包含するアミノ酸配列；少なくとも一つの上記配列が本発明の核酸配列の配列を包含する一組の核酸配列；少なくとも一つの上記配列が本発明のアミノ酸配列の配列を包含する一組のアミノ酸配列；1種または2種以上の本発明による核酸配列の配列を包含する核酸配列を表わす一組のデータ；本発明のアミノ酸配列の配列を包含するアミノ酸配列をコードする核酸配列を表わす一組のデータ；少なくとも一つの上記配列が本発明の核酸配列の配列を包含

20

【0101】

本発明によりまた、文字配列、特に遺伝子配列の分析方法が提供される。好適な配列分析方法は、例えば配列相同性分析（例えば、同一性および類似性の分析）、RNA構造分析、配列のアセンブリ、分岐論的分析、配列モチーフ分析、オープンリーディングフレーム

30

決定、核酸塩基の呼び出し、およびシーケンシングクロマトグラムピーク分析の方法を包含する。

コンピューターベースの方法は、核酸配列の同一性または類似性の特定を行う方法を提供する。この方法は、コンピュータ可読な媒体において本発明の核酸の配列を包含する核酸配列を用意する工程；およびこの核酸配列を少なくとも1種の核酸またはアミノ酸配列と比較し、配列同一性または類似性を確定する工程；を包含する。

コンピューターベースの方法はまた、アミノ酸相同性の同定を行う方法を提供し、この方法は、コンピュータ可読な媒体において本発明のアミノ酸の配列を包含するアミノ酸配列を用意する工程；およびこのアミノ酸配列を少なくとも1種の核酸またはアミノ酸配列と比較し、相同性を確定する工程；を包含する。

40

コンピューターベースの方法はさらにまた、単一核酸配列中に核酸配列を重複させるアセンブリ方法を提供する。この方法は、コンピュータ可読な媒体において本発明の核酸の配列を包含する第一の核酸配列を用意する工程；およびこの第一の核酸配列と第二の核酸配列との間の少なくとも一つの重複領域についてスクリーニングする工程；を包含する。

【0102】

大腸癌の診断方法

本発明はまた、癌の定量的および定性的診断分析方法、ならびに、大腸癌を有するか、または有する恐れがあるヒト患者における、もしくは大腸癌を発病する危険性を有する患者におけるCSNAまたはCSPの発現と、正常ヒト対照におけるCSNAまたはCSPの発現とを比較することによって、癌を検出、診断、モニタリング、病期分類および予見す

50

る方法を提供する。本発明の目的に関して、「C S N Aの発現」または「P S N P発現」の用語は、当該技術分野で公知のいずれかの方法により測定することができるC S G m R N Aの量または患者の細胞、組織または患者自身において当該技術分野で公知のいずれかの方法により測定することができる転写のレベルを意味する。同様に、「C S Pの発現」または「C S P発現」の用語は、当該技術分野で公知のいずれかの方法により測定することができるC S Pの量または当該技術分野で公知のいずれかの方法により測定することができる転写のレベルを意味する。

本発明は、患者における大腸癌、特に扁平上皮癌の診断方法を提供し、この方法は、細胞、組織、臓器または体液中のC S N AまたはC S Pレベルの変化について、好ましくは同一タイプの正常ヒト対照の細胞、組織、臓器または体液中のC S N AまたはC S Pのレベルと比較することによって分析する方法であり、この方法において、或る場合の患者におけるC S N AまたはC S Pのレベルに対する正常ヒト対照に対する増加または減少は、大腸癌の存在に関連するか、または疾患の好発性に関連する。もう一つの好適態様において、本発明は、正常対照からのm R N Aに比較したC S Gのm R N Aの構造における変化を分析することによる患者における大腸癌の診断方法を提供する。これらの変化には、これらに制限されないものとして、異常スプライシング、ポリアデニル化における変化および／または5'ヌクレオチドキャッピングにおける変化が包含される。さらにもう一つの好適態様において、本発明はC S Pの変化を正常対照からのC S Pに比較して分析することによる患者の大腸癌の診断方法を提供する。これらの変化には、例えばC S Pのグリコシル化および／またはリン酸化における変化、もしくは細胞内のC S P局在化が包含される。

【0103】

好適態様において、C S N Aの発現は、配列番号148～250から選択されるアミノ酸配列、その相同体、対立遺伝子変異体または断片をコードするm R N Aの量を測定することにより測定される。さらに好適な態様において、測定されるC S N A発現は、配列番号1～147、もしくはそのハイブリダイズする核酸、相同性核酸または対立遺伝子変異体、またはこれらの核酸のいずれかの一部から選択されるC S N A m R N Aの発現レベルである。C S N A発現は、当該技術分野で公知のいずれかの方法、例えば上述の方法によって測定することができ、このような方法にはノーザンブロットによるm R N A発現の測定、定量的または定性的逆転写酵素P C R (R T-P C R)の測定、マイクロアレイ法、ドットブロットまたはスロットブロット法、または*i n s i t u*ハイブリダイゼーション法が包含される。

一例として、A u s u b e lによる上記文献(1992年)；A u s u b e lによる上記文献(1999年)；S a m b r o o kによる上記文献(1989年)；およびS a m b r o o kによる上記文献(2001年)を参照することができる。C S N A転写は、対象となるC S Gのプロモーターまでフックされたレセプター遺伝子を用いるか、または核ランナーオフアッセイを行う方法を包含する当該技術分野で公知のいずれかの方法により測定することができる。m R N A構造の変化、例えば異常スプライシング変異体は、当業者に公知のいずれかの方法により測定することができ、このような方法にはR T-P C R、引続くシーケンシング分析または制限分析が包含される。必要に応じて、C S N A発現は、公知対照、例えば正常大腸核酸と比較し、発現の変化を検出することができる。

もう一つの好適態様において、C S Pの発現は、配列番号148～250、もしくはその相同体、対立遺伝子変異体または断片からなる群から選択されるアミノ酸配列を有するC S Pのレベルを測定することによって測定される。このようなレベルは好ましくは、正常および異常レベルの測定を包含し、少なくとも1種の細胞、組織、臓器および／または体液で測定する。すなわち、一例として、正常対照の体液、細胞または組織試料に比較したC S N AまたはC S Pの過剰または過少発現を診断する本発明による診断方法は、大腸癌の存在の診断に使用することができる。C S Pの発現レベルは、当該技術分野で公知のいずれかの方法により測定することができ、このような方法には上記記載の方法が包含される。好適態様において、C S P発現レベルは、放射性免疫検定法、競合結合検定法、E

L I S A、ウェスタンブロッティング、F A C S、免疫組織化学法、免疫沈降法、プロテオミクス手法：二次元ゲル電気泳動（2 D 電気泳動）および無ゲルベースの手法、例えば質量分析またはタンパク質相互作用プロファイリングにより測定することができる。一例として、H a r l o wによる上記文献（1999年）；A u s u b e lによる上記文献（1992年）；およびA u s u b e lによる上記文献（1999年）を参照することができる。C S P構造における変化は、当該技術分野で公知のいずれかの方法により測定することができ、このような方法は、例えばホスホセリン、ホスホトレオニンまたはホスホチロシン残基を特異的に認識する抗体を使用する方法、二次元ポリアクリルアミドゲル電気泳動（2 D P A G E）を使用する方法および／またはタンパク質のアミノ酸残基を化学的に分析する方法を包含する。同上。

10

【0104】

好適態様において、ラジオイムノアッセイ（R I A）またはE L I S Aが用いられる。C S Pに特異的な抗体は、これがすでに利用できる状態にない場合には調製する。好適態様において、この抗体はモノクローナル抗体である。抗-C S P抗体を固体支持体に結合させ、次いでこの固体支持体上のいずれかの遊離タンパク質結合部位をタンパク質、例えばウシ血清アルブミンによりブロックする。C S Pが抗-C S P抗体に結合する条件下に、この固体支持体上で抗体とともに対象となる試料をインキュベートする。試料を取り除き、この固体支持体を洗浄し、未結合物質を除去し、次いでこの固体支持体に検出可能な試薬（R I Aの場合、放射性物質、またE L I S Aの場合、酵素）に結合されている抗-C S P抗体を添加し、次いで標識した抗体に対するC S Pの結合が生じる条件下にインキュベートする。E L I S Aの場合、1種または2種以上の基質を添加し、試料中のC S Pの量に基づく着色した反応生成物を生成させる。R I Aの場合、この固体支持体を当該技術分野で公知のいずれかの方法により放射能減衰シグナルについて計量する。R I AおよびE L I S Aの両方にかかわる定量的結果は代表的に、標準曲線を引用して得られる。C S Pレベルの別の測定方法は当該技術分野で公知である。一例として、競合アッセイを用いることができ、この方法では、抗-C S P抗体を固体支持体に結合させ、次いでこの固体支持体とともに割当てられた量の標識されたC S Pおよび対象となる試料をインキュベートする。固体支持体に結合された標識したC S Pの検出量は、試料中のC S Pの量と相関させることができる。

20

プロテオミクス手法の場合、2 D P A G Eは周知の技術である。血清などの試料からの各タンパク質の単離は、等電点および分子量によるタンパク質の順次的分離を用いて達成される。代表的に、ポリペプチドが等電点により先ず分離され（第一次元）、次いで電流を用いてサイズにより分離される（第二次元）。一般に、第二次元分離は第一次元分離に対して直角である。相違する配列を有する2種のタンパク質は、サイズおよび電荷の両方に基づき同一であることから、2 D P A G Eの結果は、各タンパク質が特別のスポットを占有する略短形のゲルである。化学的または抗体プローブによるスポットの分析、または引続くタンパク質マイクロシーケンシングは、所定のタンパク質の相対的存在量および試料中のタンパク質の同定を明らかにすることができる。

30

【0105】

C S N Aの発現レベルは、当該技術分野で公知のいずれかの方法により測定することができる。かかる方法には、P C Rおよびその他の核酸法、例えば診断および種々の悪性腫瘍のモニタリングのための悪性細胞の検出に使用することができるリガーゼ連鎖反応（L C R）および核酸配列ベースの増幅（N A S B A）が包含される。一例として、逆転写酵素P C R（R T-P C R）は、数千の別種のm R N A種の複合混合物中における特定のm R N A集団の存在の検出に使用することができる強力な技術である。R T-P C Rにおいて、m R N A種を先ず、酵素、逆転写酵素を用いて相補性D N A（c D N A）に逆転写する；このc D N Aを次いで、標準P C R反応でのように増幅させる。固体支持体上に整列された特定のD N A分子（例えば、オリゴヌクレオチド）は、対象となる1種または2種以上のC S N Aの発現の検出および発現レベルの定量の両方に使用することができる。この方法において、1種または2種以上のC S N Aの全

40

50

部または一部を基体に固定する。RNAを含有し得る対象となる試料を、ハイブリダイゼーションが固体支持体上のDNAと対象となる試料中の核酸分子との間で生じる条件下に、固体支持体とともにインキュベートする。基体に結合したDNAと試料中の核酸分子との間のハイブリダイゼーションは、数種の手段により検出し、また定量することができ、このような手段は、これらに制限されないものとして、核酸分子またはハイブリットが検出されるようにデザインされた二次分子に対する放射能標識または蛍光標識を包含する。

【0106】

上記試験は、種々の細胞、体液および／または組織抽出物、例えば患者から得られたホモジネートまたは可溶化組織に由来する試料に対し行うことができる。組織抽出物は、組織生検および解剖材料から日常的に得られる。本発明において有用な体液は、血液、尿、唾液またはいずれか他の身体分泌液またはその派生物を包含する。血液の用語は、全血、血漿、血清またはいずれかの血液の派生物を包含することを意味する。好適態様において、CSNAまたはCSPの発現にかかわり試験される試料は、これらに制限されないものとして、大腸組織、気管支胞洗浄(BAL)により得られる液体、痰、細胞培養で増殖した大腸細胞、血液、血清、リンパ節組織およびリンパ液を包含する。もう一つの好適態様において、特に原発性大腸癌の転移が知られているか、または予測される場合、この試料には、これらに制限されないものとして、脳、骨、骨髄、肝臓、副腎および大腸からの組織が包含される。一般に、これらの組織は、生検により試料採取することができ、この生検には、これらに制限されないものとして、針生検、例えば経胸腔針吸引、頸部縦隔鏡検査、内視鏡リンパ節生検、ビデオ補助胸腔鏡、試験開胸、骨髄生検、および骨髄吸引が包含される。Scottによる上記文献およびKaneによる上記文献中のFranklinによる529-570頁を参照することができる。早期の安価な検出の場合、喀痰試料中の細胞のCSNAまたはCSPにおける変化にかかわる分析は特に有用である。喀痰試料の採取方法および分析方法は、Franklinによる上記文献に記載されている。本発明による全方法は、場合により、CSNAまたはCSPの発現レベルの測定に加え、1種または2種以上の別種癌マーカーの発現レベルの測定を包含することができる。かなりの場合、さらに1種の癌マーカーの使用は誤った陽性または誤った陰性の可能性を減少させる。一態様において、1種または2種以上の別種の癌マーカーは、本明細書に記載されているような別種のCSNAまたはCSPを包含する。本発明で有用な別種の癌マーカーは、検査される癌に依存し、当業者にとって公知である。好適態様において、特定のCSNAまたはCSPに加え、少なくとも1種の別種の癌マーカーを測定する。さらに好適な態様において、少なくとも2種の別種の癌マーカーを使用する。さらにまた好適態様において、少なくとも3種、さらに好ましくは少なくとも5種、さらにまた好ましいものとして少なくとも10種の癌マーカーを使用する。

【0107】

診断

一態様において、本発明は大腸癌を有するものと予測される患者からの試料中の1種または2種以上のCSNAまたはCSPの発現レベルおよび／または構造変化を測定する方法を提供する。一般に、この方法は、患者から試料を得る工程、CSNAおよび／またはCSPの発現レベルまたは構造変化を測定する工程、および次いでCSNAまたはCSPの発現レベルから患者が大腸癌を有するかを推定する工程を包含する。一般に、CSNAまたはCSPの発現が対照に比較して高い場合、大腸癌が示唆され、またCSNAまたはCSPの発現レベルが、好ましくは正常ヒト対照の同一細胞、組織または体液に比較し、少なくとも2倍またはそれ以上である場合、さらに好ましくは少なくとも5倍高い場合、さらにまた好ましくは少なくとも10倍高い場合、診断は陽性であると考えられる。これに対して、対照に比較して低いCSNAまたはCSP発現が大腸癌を示唆する場合、CSNAまたはCSPの発現レベルが、好ましくは正常ヒト対照の同一細胞、組織または体液に比較し、少なくとも2倍低い、さらに好ましくは少なくとも5倍低い、さらにまた好ましくは少なくとも10倍低いならば、この診断分析は陽性であると考えられる。正常ヒト対照は、相違する患者から、または同一患者の冒されていない組織からのものであることが

できる。

本発明はまた、大腸癌が患者において転移された癌であるかを診断する方法を提供する。その大腸癌が転移された癌であるかは、種々の組織の1種または2種以上のC S N AまたはC S Pの発現レベルおよび／または構造変化を測定することによって特定することができる。或る組織に対応する非癌組織（別の個人からの同一組織）よりも高レベルでC S N AまたはC S Pが存在する場合であって、高レベルのC S N AまたはC S P発現が大腸癌に関連する場合、転移が示唆される。同様に、或る組織に対応する非癌組織よりも低レベルでC S N AまたはC S P発現が存在する場合であって、低レベルのC S N AまたはC S P発現が大腸癌に関連する場合、転移が示唆される。さらにまた、大腸癌に関連するC S N AまたはC S Pの構造変化の存在はまた、転移を示唆する。

10

一般に、対照と比較してC S N AまたはC S Pの高発現が転移を示唆する場合であって、正常ヒト対照の好ましくは同一細胞、組織または体液と比較し、C S N AまたはC S Pの発現レベルが少なくとも2倍高い、さらに好ましくは少なくとも5倍高い、さらに一層好ましくは少なくとも10倍高いならば、転移にかかわる検査は陽性であると考えられる。これに対して、対照に対して低いC S N AまたはC S Pの発現が転移を示唆する場合であって、正常ヒト対照の好ましくは同一細胞、組織または体液と比較し、C S N AまたはC S Pの発現レベルが少なくとも2倍低い、さらに好ましくは少なくとも5倍低い、さらに一層好ましくは少なくとも10倍低いならば、転移にかかわる検査は陽性であると考えられる。

本発明のC S N AまたはC S Pは、大腸癌または別の大腸関連障害が関連する発現パターンを認識するアレイまたは多検体試験における要素として使用することができる。さらに、核酸またはタンパク質の配列は、大腸癌のパターン認識用のコンピュータプログラムにおける要素として使用することができる。

20

【0108】

病期分類

本発明はまた、ヒト患者における大腸癌の病期分類方法を提供する。この方法は、大腸癌を有するヒト患者を特定し、次いでこのようなヒト患者からの細胞、組織または体液を1種または2種以上のC S N AまたはC S Pの発現レベルおよび／または構造変化について分析することを包含する。まず、種々の患者からの1種または2種以上の腫瘍を、当該技術分野で周知の方法に従い病期分類し、次いで1種または2種以上のC S N AまたはC S Pの発現レベルを各病期について測定し、C S N AおよびC S Pのそれぞれにかかわる標準発現レベルを得る。次いで、その病期が不明である患者からの生物学的試料において、そのC S N AまたはC S P発現レベルを測定する。この患者からのC S N AまたはC S P発現レベルを次いで、標準発現レベルと比較する。患者からのC S N AおよびC S Pの発現レベルと標準発現レベルとを比較することによって、当該腫瘍の病期を決定することができる。同一方法を、C S N AまたはC S Pの構造変化を用いて行い、大腸癌の病期を決定することができる。

30

【0109】

モニタリング

本発明はさらに、ヒト患者における大腸癌のモニタリング方法を提供する。転移が存在するかを決定するために、または転移が存在したとして、転移が生じ始めた時点を決めるために、ヒト患者をモニタリングすることができる。また、前癌病変が癌になるか否かを決定するために、ヒト患者をモニタリングすることができる。また、治療、例えば化学療法、放射線治療または手術が大腸癌を減少または消滅させたか否かを決定するために、ヒト患者をモニタリングすることができる。この方法は、大腸癌についてモニタリングすることを望むヒト患者を特定し、このヒト患者からの細胞、組織または体液を1種または2種以上のC S N AまたはC S Pの発現レベルについて定期的に評価し、次いで予め得たC S N AまたはC S P発現レベルに対し、経過時間にわたってそのC S N AまたはC S Pレベルを比較することを包含する。患者はまた、大腸癌に関連するC S N AまたはC S Pの1種または2種以上の構造変化を測定することによってモニタリングすることもできる。

40

50

C S N A または C S P の増加した発現が、転移、処置の失敗、または前癌病変の癌病変への転化に関連する場合、C S N A または C S P の発現レベルの増加の検出は、それぞれこの腫瘍が転移癌であること、処置は失敗であること、または病変は癌であることを示唆する。これが真実である場合、減少した発現レベルは、転移がないこと、治療が効果的であること、または新生物病変への進行が失敗に終わっていることを示唆するものと、当業者は認識するものと見なされる。C S N A または C S P の発現減少が、転移、治療の失敗、前癌病変の癌病変への転化に関連する場合、C S N A または C S P 発現レベルの減少の検出は、それぞれこの腫瘍が転移であること、処置は失敗であること、または病変は癌であることを示唆する。好適態様において、C S N A または C S P のレベルは、前の患者と同一の細胞種、組織または体液から測定される。大腸癌転移の開始にかかわる患者のモニタリングは、定期的に行い、好ましくは一年四回行うが、この頻度は多くても、または少なくともよい。

本明細書に記載されている方法はまた、C S N A または C S P の発現レベルの増加または減少を伴う疾患または障害の発生または発生の危険性について対象を特定するための予後検査として利用することができる。本発明は、被験試料をヒト患者から採取し、次いで1種または2種以上のC S N A および／またはC S P を検出する方法を提供する。正常ヒト対照に比較した高レベル（または低レベル）のC S N A および／またはC S P の存在によって、当該ヒト患者における癌、特に大腸癌の発病にかかわる危険性が診断される。本発明の1種または2種以上のC S N A および／またはC S P の減少した（または増加した）発現または活性に対する治療剤の効果はまた、臨床試験におけるヒト患者のC S N A および／またはC S P の発現レベルの分析によって、またはヒト細胞におけるなどの *i n v i t r o* スクリーニングアッセイによって、モニタリングすることができる。この方法で、遺伝子発現パターンは、このような場合があるとして、試験される治療剤に対するヒト患者または細胞の生理学的応答を示すマーカーとしての役目を果たすことができる。

【0110】

遺伝子欠陥または変異の検出

本発明の方法はまた、C S G における遺伝子欠陥または変異の検出に使用し、これにより遺伝子欠陥を有するヒトが大腸癌を発病しやすいかを測定することができ、または遺伝子欠陥がヒトに存在する大腸癌に対し応答性であるか、または部分的に応答性であるかを測定することができる。遺伝子欠陥は、一例として、本発明のC S G からの1種または2種以上のヌクレオチドの欠失、挿入および／または置換の存在、C S G の染色体再構成、C S G の異常修飾（例えば、ゲノムDNAのメチル化パターン）、またはC S G の対立遺伝子の喪失を確認することによって検出することができる。本発明のC S G における、このような欠陥の検出方法は、本明細書の教示に従い当業者に知られている。

【0111】

非癌性大腸疾病の検出方法

本発明はまた、非癌性大腸疾病を有することが予測されるか、または有することが知られている患者からの試料中の1種または2種以上のC S N A および／またはC S P の発現レベルおよび／または構造変化の測定方法を提供する。一般に、この方法は、患者から試料を採取する工程、C S N A および／またはC S P の発現レベルおよび／または構造変化を測定する工程、このC S N A および／またはC S P の発現レベルおよび／または構造変化を、正常大腸対照と比較する工程、および次いで、当該患者が非癌性大腸疾病を有するかどうかを確認する工程を包含する。一般に、対照に比較して高レベルのC S N A またはC S P 発現が特定の非癌性大腸疾病を示唆する場合であって、C S N A またはC S P の発現レベルが、正常ヒト対照の、好ましくは同一細胞、組織または体液中よりも少なくとも2倍高い、好ましくは5倍高い、さらに好ましくは少なくとも10倍高い場合、この診断試験は陽性であると考えられる。これに対して、C S N A またはC S P の低い発現レベルが非癌性大腸疾病を示唆する場合であって、C S N A またはC S P の発現レベルが、正常ヒト対照の、好ましくは同一細胞、組織または体液中よりも少なくとも2倍低い、好ましくは5倍低い、さらに好ましくは少なくとも10倍低い場合、この診断試験は陽性であると考

えられる。

当業者は、C S N Aおよび／またはC S Pを特定の非癌性大腸疾病と関係付けられるかは、対象となる非癌性大腸疾病を有する患者からの大腸組織を採取し、次いでこの組織において、C S N Aおよび／またはC S Pが正常大腸組織に比較して高レベルまたは低レベルで発現していることを測定することによって決定することができる。もう一つの例において、C S N AまたはC S Pが特定の非癌性大腸疾病において構造変化を示すかは、対象となる非癌性大腸疾病を有する患者からの大腸組織を採取し、正常大腸組織に比較して1種または2種以上のC S N Aおよび／またはC S Pの構造変化を測定することによって決定することができる。

【0112】

大腸組織の同定方法

もう一つの側面において、本発明は大腸組織の同定方法を提供する。これらの方法は、例えば法医学、大腸細胞の分化、および発育において、また組織工学において特に有用である。

一態様において、本発明は、試料が大腸組織であるか、または大腸組織様特徴を有するかを測定する方法を提供する。この方法は、大腸組織または大腸組織様特徴を有する組織であると予想される試料を採取する工程、この試料が1種または2種以上のC S N Aおよび／またはC S Pを発現するかを測定する工程、および試料が1種または2種以上のC S N Aおよび／またはC S Pを発現した場合、試料は大腸組織を含有するものと結論する工程を包含する。好適態様において、C S N Aは配列番号148～250、もしくはその相同体、対立遺伝子変異体または断片から選択されるアミノ酸配列を有するポリペプチドをコードする。

さらに好適な態様において、C S N Aは配列番号1～147、もしくはそのハイブリダイズする核酸、対立遺伝子変異体または一部から選択されるヌクレオチド配列を有する。試料がC S N Aを発現するかにかかわる測定は、当該技術分野で公知のいずれかの方法によって達成することができる。好適方法は、マイクロアレイに対するハイブリダイゼーション、ノーザンブロットハイブリダイゼーション、および定量的または定性的R T - P C Rを包含する。もう一つの好適態様において、この方法は、C S Pが発現されるかについて測定することによって実施することができる。試料がC S Pを発現するかにかかわる測定は、当該技術分野で公知のいずれかの方法によって実施することができる。好適方法は、ウエスタンブロット、E L I S A、R I Aおよび2 D P A G Eを包含する。一態様において、C S Pは配列番号148～250、もしくはその相同体、対立遺伝子変異体または断片から選択されるアミノ酸配列を有する。もう一つの好適態様において、少なくとも2種のC S N Aおよび／またはC S Pの発現を測定する。さらに好適な態様において、少なくとも3種、さらに好ましくは4種、さらにまた好ましくは5種のC S N Aおよび／またはC S Pの発現を測定する。

一態様において、この方法は、未知試料が大腸組織であるかについての決定に使用することができる。これは法医学において特に有用であり、この場合、顕微鏡またはその他の手段によって同定することができない組織の小さい損傷した組織を犯罪または事故現場から採取する。もう一つの態様において、この方法は、組織が大腸組織に分化または成長するかについての測定に使用することができる。これは、種々の作用物質の細胞または組織培養物への添加効果のモニタリングに、例えば組織操作による新しい大腸組織の生成に、重要である。これらの作用物質には、例えば成長および分化因子、細胞外マトリックスタンパク質および培地が包含される。組織発生および分化に対する効果を測定することができる別種の因子には、細胞または組織中への遺伝子転移、p Hの変更、水：空気界面および種々の別の培養条件が包含される。

【0113】

大腸組織の生成および修飾方法

もう一つの側面において、本発明は操作された(e n g i n e e r e d)大腸組織または細胞の生成方法を提供する。一態様において、この方法は、細胞を用意する工程、この細

10

20

30

40

50

胞中にC S N AまたはC S Gを導入する工程、およびこれらの細胞が大腸組織細胞の1種または2種以上の特徴を示す条件下に、これらの細胞を成長させる工程を包含する。好適態様において、これらの細胞は多分化能を有する。当該技術分野で周知のように、正常大腸組織は、多数の相違する細胞種を包含する。従って、一態様において、この操作された大腸組織または細胞は、これらの細胞種の1種を包含する。もう一つの態様において、操作された大腸組織または細胞は、1種以上の大腸細胞を含有する。さらに、これらの細胞または組織の培養条件は、大腸細胞組織の完全な分化および成長を達成するための操作を必要とすることがある。培養条件の取り扱い方法は、当該技術分野で周知である。

1種または2種以上のC S Pをコードする核酸分子は、細胞、好ましくは多能性細胞中に導入される。好適態様において、この核酸分子は、配列番号148～250、もしくはその相同タンパク質、類縁体、対立遺伝子変異体または断片から選択されるアミノ酸配列を有するC S Pをコードする。さらに好適な態様において、この核酸分子は、配列番号1～147、もしくはそのハイブリダイズする核酸、対立遺伝子変異体または一部から選択されるヌクレオチド配列を有する。もう一つのさらに好適な態様において、C S Gが細胞中に導入される。細胞中への核酸分子の導入方法および発現ベクターは、当該技術分野で周知であり、上記文献に詳細に記載されている。

人工大腸組織は、彼等の大腸機能のいくらか、または全部を失った患者の処置に使用することができる。

【0114】

医薬組成物

他の側面において、本発明は、本発明による核酸分子、ポリペプチド、抗体、抗体誘導体、抗体断片、アゴニスト、アンタゴニストおよびインヒビターを含有する医薬組成物を提供する。好適態様において、この医薬組成物は、C S N Aまたはその一部を含有する。さらに好適な態様において、C S N Aは配列番号1～147、そこにハイブリダイズする核酸、その対立遺伝子変異体、またはこれと実質的に同一の配列を有する核酸からなる群から選択されるヌクレオチド配列を有する。もう一つの好適態様において、医薬組成物は、C S Pまたはその断片を含有する。さらに好適な態様において、C S Pは、配列番号148～250、これと相同性であるポリペプチド、このポリペプチドの全部または一部を包含する融合タンパク質、もしくはその類縁体または誘導体からなる群から選択されるアミノ酸配列を有する。もう一つの好適態様において、医薬組成物は、抗C S P抗体、好ましくは配列番号148～250からなる群から選択されるアミノ酸を有するC S P、これと相同性であるポリペプチドに結合する抗体、このポリペプチドの全部または一部を含有する融合タンパク質、もしくはその類縁体または誘導体を含有する。

このような組成物は代表的に、薬学的に許容される担体または賦形剤中に、および／または薬学的に許容される担体または賦形剤を用いて、本発明の治療剤を約0.1～90重量%の量で含有する。

医薬組成物は十分に確立された技術であり、下記刊行物に開示されており、これらの記載を、それらの全体の引用によりここに組入れる：Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000)；Ansel et al., Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999)；およびKibbe (ed.), Handbook of Pharmaceutical Excipients, American Pharmaceutical Association, 3rd ed. (2000)。

【0115】

簡単には、本発明の医薬組成物の組成は、投与のために選択される経路に依存する。本発明で利用される医薬組成物は、経腸および非経口の両方を包含する種々の経路によって投与することができ、このような経路には、経口、静脈内、筋肉内、皮下、吸入、局所、舌

10

20

30

40

50

下、直腸、動脈内、脊髄内、くも膜下、心室内、経粘膜、経皮、鼻内、腹腔内、肺内、および子宮内が含まれる。

経口剤型は、患者によって消化される錠剤、丸剤、糖衣錠、カプセル剤、液体、ゲル、シロップ、スラリー、懸濁剤などとして調剤することができる。

経口投与用組成物の固形製剤は、適当な担体または賦形剤、例えば炭水化物またはタンパク質充填剤、例えば、乳糖、ショ糖、マンニトールまたはソルビトールを含む糖；トウモロコシ、麦、米、ジャガイモ、またはその他の植物からのデンプン；メチルセルロース、ヒドロキシプロピルメチルセルロース、ナトリウムカルボキシメチルセルロース、または微結晶セルロースなどのセルロース類；アラビアゴムおよびトラガカントガムを含むガム類；ゼラチンおよびコラーゲンなどのタンパク質類；カオリン、炭酸カルシウム、リン酸二カルシウム、塩化ナトリウムなどの無機化合物；およびアカシアおよびアルギン酸などの助剤を含有することができる。

10

分解および／または溶解を促進する助剤を添加することができ、例えば架橋したポリビニルピロリドン、寒天、アルギン酸、またはその塩、例えばアルギン酸ナトリウム塩、微結晶セルロース、トウモロコシデンプン、ナトリウムデンプングリコレート、およびアルギン酸を添加することができる。

使用することができる錠剤結合剤は、アカシア、メチルセルロース、ナトリウムカルボキシメチルセルロース、ポリビニルピロリドン（ポビドン（Povidone）（登録商標））、ヒドロキシプロピルメチルセルロース、ショ糖、デンプンおよびエチルセルロースを含む。

20

使用することができる滑剤は、ステアリン酸マグネシウム、ステアリン酸、シリコーン液、タルク、ワックス類、油類、およびコロイド状シリカを含む。

前記材料を含む充填剤、分解および／または溶解を促進する助剤、錠剤結合剤および滑剤は、単独でまたは組み合わせて使用することができる。

固形経口剤型は、全体的に均一である必要はない。一例として、糖衣錠芯部を適当なコーティング、例えば濃縮糖溶液コーティング（これはまた、アラビアゴム、タルク、ポリビニルピロリドン、カルボポールゲル、ポリエチレングリコール、および／または二酸化チタンを含有することができる）、ラッカー溶液、および適当な有機溶媒または溶媒混合物と組合せて使用することができる。

【0116】

30

本発明の経口剤型は、ゼラチンから形成されているプッシュフィット型（push-fit）カプセル剤、ならびにゼラチンおよびコーティング、例えばグリセロールまたはソルビトールから形成されている軟質のシールされたカプセル剤を含む。プッシュフィット型カプセル剤は、充填剤または結合剤（例えば、乳糖またはデンプン）、滑剤（例えば、タルクまたはステアリン酸マグネシウム）、および必要に応じて、安定剤と混合された活性成分を含有することができる。軟質カプセル剤において、活性成分は安定剤を含有するか、または含有していない適当な液体、例えば脂肪油、液体、液状ポリエチレングリコールに溶解または懸濁することができる。

さらに、錠剤または糖衣錠コーティングには染料または顔料を添加することができ、これにより製品を確認することができ、または活性化合物の量、すなわち用量を特定することができる。

40

経口（経腸）投与用の医薬組成物の液体製剤は、水または水性媒質中で調製され、種々の懸濁剤、例えばメチルセルロース、アルギネート、トラガカントガム、ペクチン、ケルジン（Kelgin）、カラゲニン、アカシア、ポリビニルピロリドンおよびポリビニルアルコールを含有することができる。液体製剤はまた、溶液、エマルジョン、シロップ、およびエリキシルを含むし、活性成分（1種または2種以上）とともに湿潤剤（1種または2種以上）、甘味剤および着色剤、ならびに風味付与剤を含有することができる。

本発明の医薬組成物はまた、非経口投与用に調剤することができる。非経口投与用製剤は、水性または非水性の等張性無菌注射溶液または懸濁液の形態であることができる。

静脈注射の場合、水溶性形態の本発明による化合物を生理学的に許容される液状媒質、例

50

例えば5%デキストロース（「D5」）、生理学的に緩衝されている塩類溶液、0.9%塩類溶液、ハンクス溶液（Hanks' solution）、またはリンゲル溶液中で調剤され、または凍結乾燥物である場合、これらの液体と混合される。静脈用製剤は、担体、賦形剤または安定剤を含有することができ、これらには、これらに制限されないものとして、カルシウム、ヒト血清アルブミン、クエン酸塩、酢酸塩、塩化カルシウム、炭酸塩またはその他の塩が包含される。

筋肉内製剤、例えば適当な可溶性形態の本発明による化合物の無菌製剤は、医薬用賦形剤、例えば注射用水、0.9%塩類溶液、または5%グルコース溶液に溶解し、次いで投与することができる。別法として、適当な不溶性形態の化合物を調製し、水性基剤または医薬として許容される油性基剤、例えば長鎖状脂肪酸のエステル（例えば、オレイン酸エチル）、脂肪油（例えば、ゴマ油）、トリグリセライド類またはリポソーム中の懸濁液として投与することができる。

10

【0117】

組成物の非経口製剤は、種々の担体、例えば植物油、ジメチルアセトアミド、ジメチルホルムアミド、乳酸エチル、炭酸エチル、ミリスチン酸イソプロピル、エタノール、ポリオール類（グリセロール、プロピレングリコール、液状ポリエチレングリコールなど）を含有することができる。

水性注射用懸濁液はまた、懸濁液の粘度を高める物質、例えばナトリウムカルボキシメチルセルロース、ソルビトールまたはデキストランを含有することができる。非脂質ポリカチオンアミノポリマー類はまた、送達用に使用することができる。必要に応じて、この懸濁液はまた、適当な安定剤または僅かに濃厚な溶液の生成を可能にするために化合物の溶解度を高める助剤を含有することができる。

20

本発明の医薬組成物はまた、注射可能な長期間貯蔵製剤の形態に調剤することができる。この注射可能なデポ製剤は、生体分解性ポリマー、例えばポリ乳酸-ポリグリコライド中で化合物の微小カプセル封入マトリックスを形成することによって製造することができる。ポリマーに対する医薬の割合および使用される特定のポリマーの性質に応じて、医薬放出速度を制御することができる。別の生体分解性ポリマーの例は、ポリ（オルトエステル）およびポリ（無水物）を包含する。デポ型注射用製剤はまた、身体組織に適合することができる微小エマルジョン中に医薬をトラップさせることによって製造される。

本発明の医薬組成物は、局所投与することができる。

30

局所使用の場合、本発明の化合物はまた、皮膚または鼻および喉の粘膜に施用するのに適する形態に調製することができ、またローション、クリーム、軟膏、液体スプレーまたは吸入液、点眼剤、チンキ、トローチ、または喉塗布剤の形態をとることができる。このような局所用製剤はまた、活性成分の表面浸透を容易にするために、ジメチルスルホキシド（DMSO）のような化学化合物を含有することができる。別の経皮製剤、代表的に貼布型放出製剤において、医薬活性化合物は1種または2種以上の皮膚浸透剤、例えば2-N-メチルーピロリドン（NMP）またはアゾン（Azon）を用いて調剤することができる。局所用半固体軟膏製剤は代表的に、約1%~20%、例えば5~10%の濃度の活性成分を担体、例えば医薬用クリーム基剤中に含有する。

眼または耳に施用する場合、本発明の化合物は、軟膏、クリーム、ローション、ペイントまたは粉末として、疎水性または親水性基剤中で調剤される液状または半液状形態に調製することができる。

40

【0118】

直腸投与の場合、本発明の化合物は慣用の担体、例えばカカオ脂、ワックスまたはその他のグリセリドと混合されている座薬の形態で投与することができる。

吸入製剤はまた、容易に調剤することができる。吸入剤の場合、種々の粉末および液体製剤を調製することができる。エアロゾル製剤の場合、化合物または塩形態の化合物の無菌製剤を吸入器、例えば計量剤型吸入器およびネブライザーで 사용할 ことができる。エアロゾル化形態は、呼吸器障害の処置に特に有用であることができる。

あるいは、本発明による化合物は、供給時点で適当な薬学的に許容される担体中で再構成

50

するのに適する粉末形態であることができる。

本発明の医薬組成物中の薬学的に活性な化合物は、種々の酸の塩として提供することができる、これらに制限されないものとして、塩酸塩、硫酸塩、酢酸塩、乳酸塩、酒石酸塩、リンゴ酸塩およびコハク酸塩を包含する。これらの塩は、対応する遊離塩基形態に比較し、水性または別のプロトン性溶媒中にさらに溶解性である傾向を有する。

医薬組成物を調製した後、これらは適当な容器内に包装することができ、適応症状の処置用にラベルを付けることができる。

活性化合物は、意図する目的の達成に有効な量で存在させる。有効用量の決定は充分に、当業者の能力の範囲内にある。

「治療有効用量」の用語は、活性成分、例えばC S Pポリペプチド、融合タンパク質またはその断片、C S Pに特異的な抗体、C S Pのアゴニスト、アンタゴニストまたはインヒビターの、疾患の兆候または症状を軽減するか、またはその進行を防止する量である。医療分野で理解されているように、治療は、望ましいが必要ではない。

本発明の医薬剤の治療有効量は、最初に*in vitro*試験、例えば細胞培養試験により推定することができ、次いでモデル動物、通常マウス、ラット、イヌまたはブタにおいて試験する。動物モデルはまた、初期好適濃度範囲および投与経路の決定に使用することができる。

【0119】

一例として、ED50（全体の50%で治療的に有効な用量）およびLD50（全体の50%を致死させる用量）は、動物モデル系の1種または2種以上の細胞培養により決定することができる。治療効果に対する毒性の用量比は、治療指数であり、LD50/ED50として表わすことができる。大きい治療指数を示す医薬組成物は好適である。

細胞培養試験および動物実験から得られるデータは、ヒト用途用の初期用量範囲の組成に使用され、好ましくは僅かな毒性または無毒性のED50を包含する循環濃度範囲を提供する。投与後、または順次投与の間、この活性剤の循環濃度は、当業者に周知の医薬動態因子、例えば使用される剤型、患者の感受性、および投与経路に依存して、この範囲内で変化する。

正確な用量は、処置を要する患者に特異的な因子の観点から、実施者によって決定される。実施者が考慮することができる因子は、疾患状態の重篤度、対象の一般健康状態、年齢、体重、対象の性別、食事、投与の時期および頻度、医薬の組合せ、反応への感受性、および治療に対する耐容/応答を包含する。長期作用性医薬組成物は、特定の製剤の半減期およびクリアランスに応じて、3～4日間置きに、一週間置きに、または2週間置きに一回、投与することができる。

通常の用量は、投与経路に依存して、約1gの総量まで0.1から100,000マイクログラムまで変えることができ、治療剤が本発明によるタンパク質または抗体である場合、治療性タンパク質または抗体製剤は代表的に、0.01mg～30mg/患者の体重kg（例えば、1mg/kg～5mg/kg）の一日薬用量で投与される。医薬製剤は、所望により、一日多回の用量で投与することができ、これにより所望の総合一日薬用量を得ることができる。

特定の剤型および供給方法にかかわる指針は、刊行物に提供されており、また一般に当該技術分野の実施者に利用することができる。当業者はタンパク質またはそれらのインヒビターについてよりも、ヌクレオチドについて相違する製剤を使用することができるものと見なされる。同様に、ポリヌクレオチドまたはポリペプチドの放出は、特定の細胞、状態、場所などに特異的なものと見なされる。

医療技術の当業者にとって公知である慣用の方法を使用し、本発明の医薬製剤（1種または2種以上）を患者に投与することができる。本発明の医薬組成物は単独で、または別種の治療剤と組合せて、または別種の治療剤を介在させて投与することができる。

【0120】

治療方法

本発明はまた、本発明の遺伝子における欠陥、例えば発現、活性、分布、場所および/ま

たは溶解性にかかわる欠陥を有する対象の処置方法を提供する。これらの欠陥は大腸機能障害として現れる。本明細書で使用するものとして、「処置」の用語は、疾患の一時的軽減および予防（防止）を包含する医学的に許容される全てのタイプの治療的介入を包含する。「処置」の用語は、僅かな改善を包含する疾患のいずれかの改善を包含する。これらの方法は、以下で説明する。

【0121】

遺伝子治療およびワクチン

本発明の単離核酸はまた、本発明によるポリペプチドの *in vivo* 発現の誘導に使用することができる。 *in vivo* 発現は、遺伝子治療目的用のベクター、代表的にウイルスベクター、多くの場合、複製不全レトロウイルス、アデノウイルスまたはアデノ随伴ウイルス（AAV）に基づくベクターから誘導することができる。 *in vivo* 発現はまた、「裸の」核酸ワクチン注射目的で、核酸に対して内因性のシグナルから誘導することができ、またはベクター、多くの場合、プラスミドベクター、例えば pVAXI（*In vitro* gen, Carlsbad, CA, USA）から誘導することができ、さらに下記刊行物に開示されており、これらの記載を、それら全体の引用によりここに組入れる：米国特許第5,589,466号；同第5,679,647号；同第5,804,566号；同第5,830,877号；同第5,843,913号；同第5,880,104号；同第5,958,891号；同第5,985,847号；同第6,017,897号；同第6,110,898号；および同第6,204,250号。癌治療の場合、ベクターはまた、腫瘍選択性であると好ましい。一例として、Doronin et al., 20 J. Virol. 75: 3314-24 (2001) を参照することができる。

本発明の治療方法のもう一つの態様において、本発明の核酸を含有する治療有効量の医薬組成物を投与する。この核酸は、CSP、融合タンパク質またはその断片の発現を誘導するベクターにおいて、またはこのようなベクターを用いることなく、産生させることができる。CSPの発現を誘導することができる核酸組成物を投与し、例えば天然CSPにおける欠陥を補足することができ、またはDNAワクチンとして投与することができる。ウイルス、複製欠失レトロウイルス、アデノウイルス、アデノ随伴ウイルス（AAV）、ヘルペスウイルス、またはワクシニアウイルスから誘導される発現ベクターは、canプラスミドとして使用することができる。例えば、Cid-Arreguiによる上記文献を参照することができる。好適態様において、この核酸分子は、配列番号148～250の30 アミノ酸配列を有するCSP、もしくはその断片、融合タンパク質、対立遺伝子変異体または相同体をコードする。

さらにもう一つの本発明による治療方法において、CSP、その融合体またはその断片を発現する宿主細胞を含有する医薬組成物を投与することができる。このような場合、細胞は代表的に、異種間または同種間拒絶が回避されるようにオーソロガスであり、またCSP産生または活性にかかわる欠陥を補足するために投与される。好適態様において、この細胞中の核酸分子は、配列番号148～250のアミノ酸配列を有するCSP、もしくはその断片、融合タンパク質、対立遺伝子変異体または相同体をコードする。

【0122】

アンチセンス投与

アンチセンス核酸組成物またはCSGアンチセンス核酸の発現を誘導するベクターは、疾患の病理学的基礎として過剰産生または異常タンパク質が存在する状況において、CSGの転写および／または翻訳を下方調節するために投与される。

治療に有用なアンチセンス組成物は、CSGのコード領域または非コード領域に対し相補性である配列を有することができる。一例として、転写開始部位から誘導されるオリゴヌクレオチド、例えば出発部位から-10～+10のオリゴヌクレオチドは好適である。

触媒的アンチセンス組成物、例えばCSG転写体に配列特異的にハイブリダイズすることができるリボザイム類はまた、治療に有用である。一例として、下記刊行物を参照することができ、これらの記載を、それらの全体の引用によりここに組入れる：Phylactou, Adv. Drug Deliv. Rev. 44 (2-3): 97-108 (50

2000); Phylactou et al., Hum. Mol. Genet. 7 (10): 1649-53 (1998); Rossi, Ciba Found. Symp. 209: 195-204 (1997); および Sigurdsson et al., Trends Biotechnol. 13 (8): 286-9 (1995)

。本発明による治療方法に有用な別種の核酸は、CSGゲノム座中またはその付近で三重らせんを形成することができる核酸である。このような三重らせん状オリゴヌクレオチドは、転写を抑制することができる。例えば、下記文献を参照することができ、これらの記載を、その全体の引用によりここに組入れる: Intody et al., Nucleic Acids Res. 28 (21): 4283-90 (2000); McGuffie et al., Cancer Res. 60 (14): 3790-9 (2000) 10。このような3量体形成オリゴ(TFO)を含有する医薬組成物は、疾患の病態生理学的基礎として過剰発現または異常タンパク質の生成が存在する状況で投与される。好適態様において、アンチセンス分子はCSP、好ましくは配列番号148~250のアミノ酸配列、もしくははその断片、対立遺伝子変異体または相同体を包含するCSPをコードする核酸分子から誘導される。さらに好適な態様において、アンチセンス分子は、配列番号1~147のヌクレオチド配列、もしくははその一部、対立遺伝子変異体、実質的に類似のまたはハイブリダイズする核酸を有する核酸分子から誘導される。

【0123】

ポリペプチド投与

本発明による治療方法の一態様において、CSP、融合タンパク質、もしくはその類縁体または誘導体を含有する治療有効量の医薬組成物を、臨床的に重大なCSP欠陥を有する対象に投与する。

タンパク質組成物は、例えば天然CSPの欠陥を補足するために投与される。別の態様において、タンパク質組成物はワクチンとして投与され、これによりCSPに対する体液および/または細胞における免疫応答が惹起される。この免疫応答はCSP活性の調節に使用することができ、または免疫原に応じて、異常形態または異常に発現された形態、例えば突然変異体または不適当に発現したイソ型に対し免疫感作するために使用することができ。さらに別の態様において、毒素分子を含有するタンパク質融合体を投与し、CSPが異常蓄積する細胞を除去することができる。

好適態様において、ポリペプチドは配列番号148~250のアミノ酸配列、もしくははその融合タンパク質、対立遺伝子変異体、相同体、類縁体または誘導体を包含するCSPである。さらに好適な態様において、このポリペプチドは、配列番号1~147のヌクレオチド配列を有する核酸分子、もしくははその一部、対立遺伝子変異体、実質的に類似の核酸、またはハイブリダイズする核酸によってコードされる。

【0124】

抗体、アゴニストおよびアンタゴニスト投与

本発明による治療方法の別の態様において、治療有効量の本発明の抗体（その断片または誘導体を包含する）を含有する本発明による医薬組成物が投与される。周知のように、抗体組成物は、例えばCSP活性に拮抗させるために、または治療剤をCSPが存在および/または蓄積している部位に標的化するために、投与される。好適態様において、この抗体は、配列番号148~250のアミノ酸配列を有するCSP、もしくははその融合タンパク質、対立遺伝子変異体、相同体、類縁体または誘導体に特異的に結合する。さらに好適な態様において、この抗体は、配列番号1~147のヌクレオチド配列を有する核酸分子によってコードされるCSP、もしくははその一部、対立遺伝子変異体、実質的に類似の核酸、またはハイブリダイズする核酸に特異的に結合する。

本発明はまた、CSPに結合するか、またはCSPの発現または活性に対し調節効果を有するモジュレーターの特定方法を提供する。CSPの発現または活性を減少させるモジュレーター（アンタゴニスト）は、大腸癌の処置に有用であると信じられている。このような選別試験は、当業者にとって公知であり、これらに制限されないものとして、細胞に基 50

づく試験法および無細胞試験法を包含する。C S Pの領域に対し特異的に結合することがコンピュータ画像化により予測される小型分子をまた、大腸癌の画像化および処置に使用するために、デザインし、合成し、また試験することができる。さらに、分子ライブラリーを、本発明により同定されたC S Pに結合する分子の能力の評価によって抗癌剤について選別することができる。このライブラリーにおいて、C S Pに対し結合可能性であるものと同定された分子は、大腸癌処置における使用をさらに評価するための鍵となる候補である。好適態様において、これらの分子は細胞におけるC S Pの発現および／または活性を下方調節するものと見なされる。

本発明による治療方法の別の態様において、C S Pの非抗体アンタゴニストを含有する医薬組成物が投与される。C S Pのアンタゴニストは、当該技術分野で一般的に知られている方法を用いて生成させることができる。特に、精製C S Pを使用し、医薬剤のライブラリー、多くの場合に小型分子のライブラリーを選別することができ、これによりC S Pに特異的に結合し、C S Pの少なくとも1種の活性に拮抗することができるものを特定することができる。

別の態様において、C S Pのアゴニストを含有する医薬組成物が投与される。アゴニストは、アンタゴニストの同定に用いられる方法と類似の方法を用いて同定することができる。

好適態様において、これらのアンタゴニストまたはアゴニストは、配列番号148～250のアミノ酸配列を有するC S P、もしくはその融合タンパク質、対立遺伝子変異体、相同体、類縁体または誘導体に特異的に結合し、それぞれ拮抗または作用する。さらに好適な態様において、これらのアンタゴニストまたはアゴニストは、配列番号1～147のヌクレオチド配列を有する核酸分子によってコードされるC S P、もしくはその一部、対立遺伝子変異体、実質的に類似の核酸、またはハイブリダイズする核酸に特異的に結合し、それぞれ拮抗または作用する。

【0125】

大腸組織のターゲティング

本発明はまた、本発明によるポリペプチドまたはそれに対する抗体を治療剤に結合させ、治療剤を大腸または大腸の特定の細胞に送達できるようにする方法を提供する。好適態様において、抗C S P抗体を治療剤と結合させ、次いでこのような治療剤を必要とする患者に投与する。大腸組織が選択的に破壊されることが求められる場合、これらの治療剤は毒素であることができる。この方法は、大腸癌細胞をターゲティングし、これらを殺滅するのに有用である。もう一つの態様において、この治療剤は成長因子または分化因子であることができる。これらは大腸細胞の機能促進に有用である。

もう一つの態様において、抗C S P抗体は、例えば磁気共鳴画像化、C TまたはP E Tを用いて検出することができる画像化剤に結合させることができる。この方法は、大腸機能の測定およびモニタリングに、大腸腫瘍の確定に、および非癌性大腸疾病の確定に有用である。

【0126】

例

例1：遺伝子発現分析

C S Gは、データマイニングソフトウェアパッケージC L A S P（登録商標）（候補リード自動検索プログラム（C a n d i d a t e L e a d A u t o m a t i c S e a r c h P r o g r a m））を用いた、I n c y t e G e n o m i c s I n c（P a l o A l t o, C A）から入手可能なL I F E S E Q（登録商標）G o l dデータベースにおける遺伝子発現の系統的分析によって同定した。C L A S P（登録商標）は、I n c y t eのデータベースを照会し、特定の組織に特異的で、かつ癌を有する患者からの組織に差別的に発現されている遺伝子を同定する1組のアルゴリズムである。L I F E S E Q（登録商標）G o l dは、どの遺伝子が、体内の種々の組織で発現されているかについての情報、ならびに正常な状態および病的な状態の両方における発現の力学についての情報を含んでいる。C L A S P（登録商標）は、最初にL I F E S E Q（登録商標）G o l dデ

10

20

30

40

50

ータベースを乳房、卵巣および前立腺などの所定の組織種に並べ替える。CLASP（登録商標）は、各組織試料を疾患の状態によって類別する。疾患の状態は、「健康」、「癌」、「癌に関連」、「他の疾患」および「その他」を含む。疾患の状態を類別することで、組織および癌特異的分子標的を同定する能力が改善される。CLASP（登録商標）は次に、（１）所定の組織種に、他の組織種に比べて選択的に発現され、かつ（２）同一または異なる組織に関して、「癌」の疾患状態において、他の疾患状態に比べて差別的に発現されている遺伝子について同時並行検索を行う。この並べ替えは、発現レベルにおける最小の変化、および差別的発現パターンが、遺伝子に関する組織試料全域で、統計的に有意とみなされるように観察される最小の頻度を特定する数学的および統計学的フィルターを用いて行われる。CLASP（登録商標）アルゴリズムは、各組織種および各疾患状態における特定の遺伝子の相対的な存在量を定量する。 10

本発明のCSGを見出すために、次の具体的なCLASP（登録商標）プロファイル、組織特異的発現（CLASP1）、癌組織でのみ検出可能な発現（CLASP2）、所定の癌について高度に差別的な発現（CLASP4）、癌組織において差別的な発現（CLASP5）、を用い、そしてcDNAライブラリーは60のユニークな組織種（LifeSeq（登録商標）の以前のバージョンは、48の組織種を有した）に分けられた。遺伝子またはESTは「遺伝子ビン（gene bin）」にグループ化され、各ビンは、共通のコンティグを共有する配列のクラスターと一緒にグループ化したものである。各遺伝子の発現レベルは、各組織種について計算した。差別的発現の有意性は、サンプルサイズ、ならびに異なるライブラリーおよび各ライブラリー内における相対的な遺伝子の存在量の変動を考慮に入れた、厳密な統計学的有意性検定によって計算した（統計学的に有意な発現を決定するのに用いる等式については、987頁の等式1および988頁の等式2を含む、Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7 (10): 986-995 (1997)を参照。その内容は参照によって組込まれる）。差別的に発現された組織特異的遺伝子は、標的組織の存在量レベルの、他のすべての組織に対するパーセンテージに基づいて選択した（組織特異性）。癌患者からの正常組織もしくは非腫瘍組織のライブラリーにおける各遺伝子の発現レベルは、腫瘍もしくは疾患に関連する組織ライブラリーでの発現レベルと比較した（癌特異性）。結果は、統計学的有意性について分析した。 20 30

【0127】

厳密なCLASP（登録商標）プロファイルの基準に合致する標的遺伝子の選択は以下の通りであった。

（a）CLASP1：組織特異的発現：CLASP1候補になるには、遺伝子は、対象となる組織において、他のすべての組織に比べ、統計学的に有意な発現を示さなければならない。遺伝子は、90%の信頼度でかかる差別的発現を示す場合のみ、CLASP1候補として選択される。

（b）CLASP2：癌組織でのみ検出可能な発現：CLASP2候補となるためには、遺伝子は、腫瘍組織で検出可能な発現を、そして正常個体からのライブラリーおよび罹病患者から得た正常組織からのライブラリーでの検出不能な発現を示さなければならない。加えて、かかる遺伝子はまた、さらに、対象となる腫瘍組織に対して特異性を示さなければならない。 40

（c）CLASP5：癌組織における差別的発現：CLASP5候補となるためには、遺伝子は、対象となる組織の腫瘍ライブラリーにおいて、すべての組織についての正常ライブラリーに比べ、差別的に発現されていなければならない。遺伝子は、90%の信頼度でかかる差別的発現を示す場合のみ、CLASP5候補として選択される。

【0128】

配列番号1～147についてのCLASPスコアを以下に列挙する。

【外1】

配列番号1	DEX0253_1	CLASP2
配列番号 2	DEX0253_2	CLASP2
配列番号 3	DEX0253_3	CLASP2
配列番号 4	DEX0253_4	CLASP2 CLASP1
配列番号 5	DEX0253_5	CLASP2 CLASP1
配列番号 6	DEX0253_6	CLASP5 CLASP1
配列番号 7	DEX0253_7	CLASP5 CLASP1
配列番号 8	DEX0253_8	CLASP2
配列番号 9	DEX0253_9	CLASP2
配列番号 10	DEX0253_10	CLASP2
配列番号 11	DEX0253_11	CLASP2 CLASP1
配列番号 12	DEX0253_12	CLASP2
配列番号 13	DEX0253_13	CLASP2
配列番号 14	DEX0253_14	CLASP5 CLASP1
配列番号 15	DEX0253_15	CLASP2
配列番号 16	DEX0253_16	CLASP2
配列番号 17	DEX0253_17	CLASP2
配列番号 18	DEX0253_18	CLASP2
配列番号 19	DEX0253_19	CLASP2
配列番号 20	DEX0253_20	CLASP2

10

20

30

【 0 1 2 9 】

【 外 2 】

配列番号 21	DEX0253_21	CLASP2
配列番号 22	DEX0253_22	CLASP2
配列番号 23	DEX0253_23	CLASP2
配列番号 24	DEX0253_24	CLASP2 CLASP1
配列番号 25	DEX0253_25	CLASP2 CLASP1
配列番号 26	DEX0253_26	CLASP2 CLASP1
配列番号 27	DEX0253_27	CLASP2 CLASP1
配列番号 28	DEX0253_28	CLASP2
配列番号 29	DEX0253_29	CLASP2
配列番号 30	DEX0253_30	CLASP2
配列番号 31	DEX0253_31	CLASP2
配列番号 32	DEX0253_32	CLASP2
配列番号 33	DEX0253_33	CLASP2
配列番号 34	DEX0253_34	CLASP2
配列番号 35	DEX0253_35	CLASP2
配列番号 36	DEX0253_36	CLASP2
配列番号 37	DEX0253_37	CLASP2
配列番号 38	DEX0253_38	CLASP2
配列番号 39	DEX0253_39	CLASP2
配列番号 40	DEX0253_40	CLASP2

10

20

30

【 0 1 3 0 】

【 外 3 】

配列番号 41	DEX0253_41	CLASP2
配列番号 42	DEX0253_42	CLASP2
配列番号 43	DEX0253_43	CLASP2
配列番号 44	DEX0253_44	CLASP2
配列番号 45	DEX0253_45	CLASP2
配列番号 46	DEX0253_46	CLASP2
配列番号 47	DEX0253_47	CLASP2
配列番号 48	DEX0253_48	CLASP2 CLASP1
配列番号 50	DEX0253_50	CLASP2 CLASP1
配列番号 51	DEX0253_51	CLASP2 CLASP1
配列番号 52	DEX0253_52	CLASP2 CLASP1
配列番号 53	DEX0253_53	CLASP2
配列番号 54	DEX0253_54	CLASP2 CLASP1
配列番号 55	DEX0253_55	CLASP2 CLASP1
配列番号 56	DEX0253_56	CLASP2
配列番号 57	DEX0253_57	CLASP2
配列番号 58	DEX0253_58	CLASP2
配列番号 59	DEX0253_59	CLASP2
配列番号 60	DEX0253_60	CLASP2

10

20

30

【 0 1 3 1 】

【 外 4 】

配列番号61 DEX0253_61 CLASP2
配列番号62 DEX0253_62 CLASP5 CLASP1
配列番号63 DEX0253_63 CLASP5 CLASP1
配列番号64 DEX0253_64 CLASP5 CLASP1
配列番号65 DEX0253_65 CLASP5 CLASP1
配列番号66 DEX0253_66 CLASP2
配列番号68 DEX0253_68 CLASP2
配列番号69 DEX0253_69 CLASP2 CLASP1
配列番号70 DEX0253_70 CLASP5 CLASP1
配列番号71 DEX0253_71 CLASP5 CLASP1
配列番号72 DEX0253_72 CLASP2
配列番号73 DEX0253_73 CLASP2
配列番号74 DEX0253_74 CLASP2
配列番号75 DEX0253_75 CLASP2
配列番号76 DEX0253_76 CLASP2 CLASP1CLASP4
配列番号77 DEX0253_77 CLASP2
配列番号78 DEX0253_78 CLASP5 CLASP1
配列番号79 DEX0253_79 CLASP5 CLASP1
配列番号80 DEX0253_80 CLASP5 CLASP1

10

20

30

【 0 1 3 2 】

【 外 5 】

配列番号 81	DEX0253_81	CLASP5 CLASP1	
配列番号 82	DEX0253_82	CLASP2	
配列番号 83	DEX0253_83	CLASP2	
配列番号 84	DEX0253_84	CLASP1	
配列番号 85	DEX0253_85	CLASP1	
配列番号 86	DEX0253_86	CLASP2	10
配列番号 87	DEX0253_87	CLASP2	
配列番号 88	DEX0253_88	CLASP2	
配列番号 89	DEX0253_89	CLASP2	
配列番号 90	DEX0253_90	CLASP2	
配列番号 91	DEX0253_91	CLASP2	
配列番号 92	DEX0253_92	CLASP2	20
配列番号 93	DEX0253_93	CLASP2	
配列番号 94	DEX0253_94	CLASP2	
配列番号 95	DEX0253_95	CLASP2 CLASP1	
配列番号 96	DEX0253_96	CLASP2 CLASP1	
配列番号 97	DEX0253_97	CLASP2	
配列番号 98	DEX0253_98	CLASP2	
配列番号 99	DEX0253_99	CLASP2	30
配列番号 100	DEX0253_100	CLASP2	

【 0 1 3 3 】

【 外 6 】

配列番号 101	DEX0253_101	CLASP2
配列番号 102	DEX0253_102	CLASP2
配列番号 103	DEX0253_103	CLASP2
配列番号 104	DEX0253_104	CLASP5 CLASP1
配列番号 105	DEX0253_105	CLASP2
配列番号 106	DEX0253_106	CLASP2
配列番号 107	DEX0253_107	CLASP2
配列番号 108	DEX0253_108	CLASP2
配列番号 109	DEX0253_109	CLASP2 CLASP1
配列番号 110	DEX0253_110	CLASP2
配列番号 111	DEX0253_111	CLASP2
配列番号 112	DEX0253_112	CLASP2
配列番号 113	DEX0253_113	CLASP2
配列番号 114	DEX0253_114	CLASP2
配列番号 115	DEX0253_115	CLASP5 CLASP1
配列番号 116	DEX0253_116	CLASP2
配列番号 117	DEX0253_117	CLASP2
配列番号 118	DEX0253_118	CLASP2
配列番号 119	DEX0253_119	CLASP2
配列番号 120	DEX0253_120	CLASP2

10

20

30

【 0 1 3 4 】

【 外 7 】

配列番号 121	DEX0253_121	CLASP2
配列番号 123	DEX0253_123	CLASP2
配列番号 124	DEX0253_124	CLASP2
配列番号 125	DEX0253_125	CLASP5 CLASP1
配列番号 126	DEX0253_126	CLASP5 CLASP1
配列番号 127	DEX0253_127	CLASP2
配列番号 128	DEX0253_128	CLASP2
配列番号 129	DEX0253_129	CLASP2
配列番号 130	DEX0253_130	CLASP2
配列番号 131	DEX0253_131	CLASP2
配列番号 132	DEX0253_132	CLASP2
配列番号 133	DEX0253_133	CLASP5 CLASP1
配列番号 134	DEX0253_134	CLASP2
配列番号 135	DEX0253_135	CLASP2
配列番号 136	DEX0253_136	CLASP2
配列番号 139	DEX0253_139	CLASP2
配列番号 140	DEX0253_140	CLASP5
配列番号 141	DEX0253_141	CLASP2
配列番号 142	DEX0253_142	CLASP2
配列番号 143	DEX0253_143	CLASP2
配列番号 144	DEX0253_144	CLASP2
配列番号 145	DEX0253_145	CLASP2
配列番号 146	DEX0253_146	CLASP2
配列番号 147	DEX0253_147	CLASP2

10

20

30

40

50

【0135】

例2：遺伝子発現の相対的定量

蛍光性タックマン (Taqman) プロブを用いたリアルタイム定量的PCRは、Taq DNAポリメラーゼの5' - 3'ヌクレアーゼ活性を利用する定量検出システムである。この方法は、5'をレポーター色素で、下流の3'をクエンチャー色素で標識した内部蛍光オリゴヌクレオチドプロブ (タックマン) を用いる。PCRの間にTaq DNAポリメラーゼの5' - 3'ヌクレアーゼ活性がレポーターを放出し、次にその蛍光がモデル7700シーケンス検出システム (PE Applied Biosystems, Foster City, CA, USA) のレーザー検出器により検出することができる。内因性対照の増幅を使用し、反応に添加された試料RNAの量が標準化され、逆転写酵素 (RT) の効率が正規化される。シクロフィリン、グリセルアルデヒド-3-リン酸

デヒドロゲナーゼ (GAPDH)、ATPase、または18SリボソームRNA (rRNA) のいずれかが内部対照として用いられる。研究される全試料間の相対定量化を計算するために、1つのサンプル試料のための標的RNAレベルを結果比較のためのベース (キャリブレーター) として用いた。「キャリブレーター」に対する定量化は、標準曲線法または比較法 (User Bulletin #2: ABI PRISM 7700 シークエンス検出システム) を用いて得ることができる。

標的遺伝子の組織分布およびレベルを、正常組織および癌組織の全ての試料について評価した。全RNAは、正常組織、癌組織、および癌とそれに対応する対応隣接組織から抽出された。続いて逆転写酵素を用いて第1鎖cDNAが調製され、各標的遺伝子に特異的なプライマーとタックマンプローブとを用いてポリメラーゼ連鎖反応が行われた。結果は、ABI PRISM 7700 シークエンス検出装置を用いて分析された。無名数は、キャリブレーター組織と比較した場合の特定組織中にある標的遺伝子の相対発現レベルである。

当業者は、適切なプライマーを設計することができる。こうして、正常組織および他の癌組織に対するCSNAの相対発現レベルを決定することができる。すべての数値は、正常胸腺 (キャリブレータ) と比較してある。これらのRNA試料は市販のプールであり、異なる個体からの特定の組織の試料をプールすることによって作出されている。

対応試料のペアおよび1つの癌および1つの正常/組織の正常隣接部分におけるCSNAの相対発現レベルもまた決定できる。すべての数値は、正常胸腺 (キャリブレータ) と比較してある。対応ペアは、特定の組織についての癌試料からのmRNA、およびその同一の個体からの同一の組織についての正常隣接試料からのmRNAによって形成される。

対応試料の分析において、CSNAは、対象となる組織に対して高度の組織特異性を示す。これらの結果は、正常プール試料で得られた組織特異性の結果を確認するものである。さらに、癌試料、および同一の個体からの同系の正常隣接組織におけるmRNA発現のレベルが比較される。この比較は、癌の病期についての具体的な指標を与える (例えば、癌試料における、正常隣接試料に比べて高いレベルのmRNA発現)。

結局、高レベルの組織特異性に加えて、テストした対応試料におけるmRNAの過剰発現は、配列番号1~147が、癌についての診断マーカーであることを示すものである。

【0136】

【外8】

DEX0095_71

cln150 -sqcln033

配列

配列 ID #

Dex0095_71

DEX0253_104

(cln150-sqcln033)

(配列番号 104)

半定量的PCRは、以下のプライマーを用いて行なった：

【表1】

10

20

30

40

プライマー	DexSeqID	から	まで	プライマー長
Sqcln033F	DEX0253_104	36	56	21
Sqcln033R	DEX0253_104	256	236	21

12の異なる組織由来の12の正常な試料におけるsqcln0033の相対発現レベルを決定した。これらのRNA試料は市販のプールであり、異なる個体からの特定の組織の試料をプールすることによって作出されている。ポリメラーゼ連鎖反応（PCR）技術を用い、デュプリケートした4つの10倍連続希釈cDNAから、発現レベルを分析した。0、1、10、100および1000の相対発現レベルを用いて、遺伝子発現を評価した。最も希釈された試料における陽性反応は、最も高い相対発現値を示す。

10

【0137】

【表2】

組織	正常
乳房	0
大腸	1000
子宮内膜	0
腎臓	0
肝臓	0
肺	0
卵巣	0
前立腺	0
小腸	1000
胃	0
精巣	10
子宮	0

20

30

40

【0138】

上記表の発現の相対レベルは、大腸および小腸を除いて、ここで試験した正常組織のいずれもが、sqcln033遺伝子発現の感知されるレベルを示さない。

12の異なる組織由来の12の癌試料におけるsqcln033の発現の相対レベルを決定した。ポリメラーゼ連鎖反応（PCR）技術を用い、デュプリケートした4つの10倍連続希釈cDNAから、発現レベルを分析した。0、1、10、100および1000の相対発現レベルを用いて、遺伝子発現を評価した。最も希釈された試料における陽性反応は、最も高い相対発現値を示す。

【0139】

【表3】

50

組織	癌
膀胱	1
乳房	0
大腸	1000
腎臓	1
肝臓	0
肺	0
卵巣	0
脾臓	0
前立腺	1
胃	0
精巣	1
子宮	0

10

20

30

40

50

【0140】

表2において、発現の相対レベルは、sqcln033が膀胱、腎臓、前立腺および精巣の癌で極めて低いレベルで発現していることを示している。高レベルの発現が大腸癌について観察された。

6つの対応ペアの大腸試料において、sqcln033の発現の相対レベルを決定した。各対応ペアは、特定の組織に対する癌試料由来のmRNAおよび同じ個体由来の同じ組織に対する正常隣接試料由来のmRNAによって形成されている。

ポリメラーゼ連鎖反応（PCR）技術を用い、デュプリケートした4つの10倍連続希釈cDNAから、発現レベルを分析した。0、1、10、100および1000の相対発現レベルを用いて、遺伝子発現を評価した。最も希釈された試料における陽性反応は、最も高い相対発現値を示す。

【0141】

【表4】

試料 ID	組織	癌	NAT
9609B019	大腸	10	100
9709C074RA	大腸	1	10
9705F002D	大腸	10	100
9608B012	大腸	0	100
4004709A1	大腸	10	10
9707C004GB	大腸	100	10

【0142】

上記表において発現の相対レベルは、sqcln033が、6つの大腸癌試料の4つにお

いて低いレベルで、そして1つにおいて中程度のレベルで発現されていることを示している。発現の中程度のレベルは、対応正常隣接組織（NAT）において観察された。

とくに上記表の結果に示された半定量的PCR実験から、大腸組織に対するsqCln033の高い程度の特異性が観察される。

実験は、定量的PCR実験のためのプライマーおよびプローブを設計および試験するために行なわれている。

【0143】

例3：タンパク質発現

C S N Aはポリメラーゼ連鎖反応（PCR）により増幅され、C S N Aをコードする増幅DNA断片が大腸菌での発現のためのpET-21d内にサブクローニングされた。C S N Aのコード配列に加え、C S N Aのコード配列のNH²末端に隣接する2個のアミノ酸、Met-AlaおよびC S N Aのコード配列のCOOH末端に隣接する6個のヒスチジンが、それぞれ開始Met／制限部位および精製タグとして用いるために組み込まれた。適切な分子量の過剰発現タンパク質バンドが、クマシーブルー染色されたポリアクリルアミドゲル上で容易に観察された。このタンパク質バンドは、6×ヒスチジンタグに対するモノクローナル抗体を用いたウェスタンブロット分析により確認された。

C S Pの大規模な精製は、6リットルの細菌培養物から得た細胞ペーストを用いて達成され、固定化金属アフィニティークロマトグラフィー（IMAC）を使用して精製された。全細胞溶解産物より分離された可溶性分画はニッケルキレート樹脂とともにインキュベートされた。カラムを充填し、カラム容量の5倍量の洗浄緩衝液で洗浄した。C S Pは、各種濃度のイミダゾール緩衝液により段階的に溶出された。

【0144】

例4：タンパク質融合

簡単に述べると、IgG分子のヒトFc部分は、下記配列の5'および3'末端に及ぶプライマーを用い、PCR増幅することができる。これらプライマーはまた、発現ベクター、好ましくは哺乳動物発現ベクターへのクローニングを容易にするのに好都合な制限酵素部位を持つべきである。例えば、pC4（受託番号209646）を用いる場合、ヒトFc部分はBamHIクローニング部位に連結できる。3' BamHI部位が破壊されるべきであることに留意されたい。次に、ヒトFc部分を含むベクターはBamHIにより再度制限処理され、ベクターは直線化され、実施例1に記載のPCR法により単離された本発明のポリヌクレオチドがこのBamHI部位に連結される。このポリヌクレオチドは終止コドンなしにクローニングされ、さもなければ融合タンパク質は産生されないことに留意されたい。天然に存在するシグナル配列を用いて分泌タンパク質を産生する場合には、pC4は第2シグナルペプチドを必要としない。あるいは、天然に存在するシグナル配列を使用しない場合には、ベクターは異種シグナル配列を含むように改変することができる。例えば、国際公開第96/34891号を参照。

【0145】

例5：ポリペプチド由来の抗体の産生

一般に、かかる手法は、ポリペプチド、またはより好ましくは分泌ペプチド発現細胞で、動物（好ましくは、マウス）を免疫化することを包含する。かかる細胞は、任意の適切な組織培養培地にて培養されてもよいが、10%のウシ胎児血清（約56℃にて不活性化される）で、ならびに非必須アミノ酸約10g/l、ペニシリン約1,000U/ml、およびストレプトマイシン約100μg/mlを補充したアール変法イーグル培地にて細胞を培養することが好ましい。

かかるマウスの脾細胞を摘出し、適切な骨髓腫細胞系と融合させる。本発明によれば、いかなる適切な骨髓腫細胞系を用いてもよいが、ATCCから入手可能な親骨髓腫細胞系（SP20）を用いることが好ましい。融合後、得られたハイブリドーマ細胞は、HAT培地にて選択的に保持され、続いてWands et al., Gastroenterology 80: 225-232 (1981)に記載されるように、限界希釈によりクローニングされる。

次に、かかる選択により得られたハイブリドーマ細胞をアッセイして、ポリペプチドを結合することが可能な抗体を分泌するクローンを同定する。あるいは、ポリペプチドに結合することが可能なさらなる抗体を、抗イディオタイプ抗体を用いた二段階手法にて産生することができる。かかる方法は、抗体自身が抗原であるという事実を利用するものであり、したがって、二次抗体に結合する抗体を得ることが可能である。この方法に従って、タンパク質特異的抗体を用いて、動物、好ましくはマウスを免疫化する。次に、かかる動物の脾細胞を用いて、ハイブリドーマ細胞を生産し、ハイブリドーマ細胞をスクリーニングして、タンパク質特異的抗体に結合する能力が、ポリペプチドにより阻止され得る抗体を産生するクローンを同定する。かかる抗体は、タンパク質特異的抗体に対する抗イディオタイプ抗体を含み、それを用いて、動物を免疫化し、さらなるタンパク質特異的抗体の形成を誘発することができる。Jameson-Wolf法を用い、以下のエピトープを予測した。(Jameson and Wolf, CABIOS, 4(1), 181-186, 1988。その内容は参照によって組込まれる)。

【0146】

【外9】

DEX0253_151 抗原性指数 (Jameson-Wolf)

位置	抗原性指数	平均長さ
68-108	1.02	41

DEX0253_154 抗原性指数 (Jameson-Wolf)

位置	抗原性指数	平均長さ
13-31	1.00	19

DEX0253_155 抗原性指数 (Jameson-Wolf)

位置	抗原性指数	平均長さ
8-22	1.25	15
67-98	1.06	32

DEX0253_156 抗原性指数 (Jameson-Wolf)

位置	抗原性指数	平均長さ
14-33	1.13	20

DEX0253_161 抗原性指数 (Jameson-Wolf)

位置	抗原性指数	平均長さ
32-46	0.98	15

DEX0253_170 抗原性指数 (Jameson-Wolf)

位置	抗原性指数	平均長さ
6-48	1.02	43

【0147】

【外10】

DEX0253_172	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
19-28	1.00	10
DEX0253_174	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
13-29	1.01	17
DEX0253_178	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
2-17	0.99	16
DEX0253_179	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
7-20	0.96	14
DEX0253_184	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
18-33	1.00	16
DEX0253_187	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
42-54	1.09	13

10

20

30

【 0 1 4 8 】

【 外 1 1 】

DEX0253_188	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
14-41	1.03	28
81-97	0.96	17
DEX0253_192	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
25-75	0.97	51
DEX0253_197	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
31-58	1.06	28
DEX0253_198	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
73-82	1.22	10
154-165	1.21	12
213-224	1.06	12
38-58	1.05	21
314-323	1.03	10
250-267	0.99	18
275-295	0.98	21
137-149	0.96	13

10

20

30

【 0 1 4 9 】

【 外 1 2 】

DEX0253_203	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
3-28	0.99	26
DEX0253_204	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
29-39	1.10	11
52-65	0.97	14
DEX0253_208	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
3-14	1.04	12
DEX0253_211	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
33-54	1.22	22
59-76	1.08	18
DEX0253_212	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
22-31	1.00	10
DEX0253_214	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
7-22	1.24	16

10

20

30

【 0 1 5 0 】

【 外 1 3 】

DEX0253_217	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
34-46	1.06	13
DEX0253_218	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
23-32	1.16	10
DEX0253_219	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
19-30	1.24	12
DEX0253_221	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
58-69	1.08	12
42-51	1.04	10
DEX0253_222	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
58-69	1.08	12
42-51	1.04	10
DEX0253_226	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
8-18	1.01	11

10

20

30

【 0 1 5 1 】

【 外 1 4 】

DEX0253_229	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
8-36	1.02	29
DEX0253_237	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
14-31	0.91	18
DEX0253_245	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
3-27	1.12	25
DEX0253_246	抗原性指数 (Jameson-Wolf)	
位置	抗原性指数	平均長さ
23-32	0.93	10

10

20

30

【0152】

本発明のBSPの翻訳後修飾（PTM）の例を以下に列挙する。さらに、かかる翻訳後修飾に特異的に結合する抗体は、診断薬としてまたは治療薬として有用でありえる。ProSiteデータベース（Baïrochet al., Nucleic Acids Res. 25 (1): 217-221 (1997)）。その内容は参照によって組込まれる）を用いて、次のPTMが、本発明のLSPについて予測された（http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html、最後にアクセスしたのは2001年10月23日）。PTMの十分な定義のためには、<http://www.expasy.org/cgi-bin/prosite-list.pl>参照、最後にアクセスしたのは2001年10月23日。

【0153】

【外15】

- DEX0253_148 Tyr リン酸化部位 3-9;
DEX0253_150 Ck2 リン酸化部位 19-22;23-26;
DEX0253_151 Ck2 リン酸化部位 83-86;128-131;
ミリスチル 28-33;49-54;
Pkc リン酸化部位 77-79;83-85;
105-107;139-141;
DEX0253_152 ロイシンジッパー6-27;
Pkc リン酸化部位 16-18;
DEX0253_153 Pkc リン酸化部位 16-18;24-26;
DEX0253_154 Ck2 リン酸化部位 2-5;23-26;
DEX0253_155 アミド化 12-15;
ミリスチル 12-17;31-36;34-39;
Pkc リン酸化部位 75-77;
DEX0253_156 Ck2 リン酸化部位 8-11;
Pkc リン酸化部位 16-18;
DEX0253_158 Pkc リン酸化部位 2-4;
DEX0253_159 ミリスチル 26-31;
Pkc リン酸化部位 9-11;14-16;

10

20

30

【 0 1 5 4 】

【 外 1 6 】

- DEX0253_160 ミリスチル 13-18;
 Pkc リン酸化部位 31-33;
- DEX0253_161 Pkc リン酸化部位 8-10;17-19;
- DEX0253_162 Ck2 リン酸化部位 28-31;
 ミリスチル 48-53;
 Pkc リン酸化部位 28-30;
- DEX0253_164 Asn グリコシル化 21-24;
 ミリスチル 17-22;
- DEX0253_165 Pkc リン酸化部位 44-46;
- DEX0253_166 Asn グリコシル化 7-10;
 Camp リン酸化部位 3-6;
 Ck2 リン酸化部位 19-22;
 ミリスチル 32-37;
 Pkc リン酸化部位 6-8;45-47;
 Tyr リン酸化部位 28-34;
- DEX0253_167 Asn グリコシル化 14-17;
 ミリスチル 23-28;
 Pkc リン酸化部位 9-11;

10

20

30

【 0 1 5 5 】

【 外 1 7 】

- DEX0253_168 Asn グリコシル化 38-41;
Ck2 リン酸化部位 40-43;
ミリスチル 49-54;
Pkc リン酸化部位 16-18;35-37;
- DEX0253_170 Camp リン酸化部位 43-46;
ミリスチル 36-41;
- DEX0253_171 Asn グリコシル化 37-40;
Ck2 リン酸化部位 79-82;86-89;
ミリスチル 22-27;34-39;47-52;
Pkc リン酸化部位 7-9;15-17;
- DEX0253_172 Asn グリコシル化 7-10;11-14;
21-24;
Pkc リン酸化部位 52-54;55-57;
- DEX0253_173 Ck2 リン酸化部位 27-30;
- DEX0253_174 ミリスチル 56-61;
- DEX0253_175 ミリスチル 24-29;
- DEX0253_176 Pkc リン酸化部位 8-10;
- DEX0253_177 Pkc リン酸化部位 9-11;

10

20

【 0 1 5 6 】

【 外 1 8 】

30

- DEX0253_178 Ck2 リン酸化部位 52-55;
ミリスチル 24-29;
Pkc リン酸化部位 45-47;
- DEX0253_179 Asn グリコシル化 63-66;
Ck2 リン酸化部位 25-28;
Pkc リン酸化部位 27-29;
- DEX0253_180 Ck2 リン酸化部位 26-29;
- DEX0253_181 アミド化 51-54;
Pkc リン酸化部位 6-8;51-53;
- DEX0253_183 Ck2 リン酸化部位 17-20;
Pkc リン酸化部位 36-38;39-41;
- DEX0253_184 Pkc リン酸化部位 31-33;
- DEX0253_186 Pkc リン酸化部位 18-20;
- DEX0253_187 Ck2 リン酸化部位 42-45;77-80;
- DEX0253_188 Ck2 リン酸化部位 21-24;86-89;
ミリスチル 25-30;
Pkc リン酸化部位 73-75;
- DEX0253_189 アミド化 25-28;
Camp リン酸化部位 10-13;
Ck2 リン酸化部位 37-40;

10

20

30

【 0 1 5 7 】

【 外 1 9 】

- DEX0253_191 Ck2 リン酸化部位 11-14;
ミリスチル 16-21;
- DEX0253_192 Ck2 リン酸化部位 48-51;61-64;
ミリスチル 57-62;
Pkc リン酸化部位 22-24;58-60;
- DEX0253_193 Ck2 リン酸化部位 43-46;
Pkc リン酸化部位 18-20;34-36;
- DEX0253_196 Asn グリコシル化 3-6;
Ck2 リン酸化部位 8-11;
- DEX0253_197 ミリスチル 4-9;
- DEX0253_198 Asn グリコシル化 46-49;54-57;
99-102;123-126;135-138;182-185;
195-198;223-226;229-232;
272-275;305-308;
Camp リン酸化部位 259-262;
Ck2 リン酸化部位 75-78;82-85;
114-117;197-200;208-211;247-250;
308-311;382-385;395-398;
ミリスチル 173-178;316-321;
320-325;334-339;339-344;
Pkc リン酸化部位 63-65;82-84;
101-103;114-116;157-159;187-189;
197-199;212-214;225-227;293-295;
395-397;
レセプターサイトカイン 1 126-138;
レセプターサイトカイン 2 304-310;

10

20

30

40

【 0 1 5 8 】

【 外 2 0 】

- DEX0253_199 Camp リン酸化部位 6-9;
DEX0253_202 アミド化 38-41;
Asn グリコシル化 2-5;45-48;
DEX0253_203 Camp リン酸化部位 24-27;
Pkc リン酸化部位 20-22;
DEX0253_204 Asn グリコシル化 65-68;
Pkc リン酸化部位 54-56;
DEX0253_205 Ck2 リン酸化部位 6-9;
ミリスチル 46-51;47-52;59-64;
63-68;78-83;
Pkc リン酸化部位 6-8;19-21;
DEX0253_207 Ck2 リン酸化部位 57-60;61-64;
ミリスチル 69-74;
Pkc リン酸化部位 34-36;
DEX0253_208 Camp リン酸化部位 5-8;
Ck2 リン酸化部位 37-40;
ミリスチル 26-31;

10

20

【 0 1 5 9 】

【 外 2 1 】

30

- DEX0253_209 Pkc リン酸化部位 52-54;
DEX0253_210 Camp リン酸化部位 34-37;
Ck2 リン酸化部位 9-12;
Pkc リン酸化部位 9-11;33-35;
DEX0253_211 アミド化 67-70;
Camp リン酸化部位 69-72;
Ck2 リン酸化部位 5-8;10-13;23-26;
ミリスチル 35-40;54-59;
Pkc リン酸化部位 67-69;
DEX0253_212 Prokar リポタンパク質 6-16;
DEX0253_213 Asn グリコシル化 44-47;
Camp リン酸化部位 37-40;
Ck2 リン酸化部位 2-5;
Pkc リン酸化部位 8-10;
DEX0253_215 Asn グリコシル化 3-6;
Ck2 リン酸化部位 28-31;
DEX0253_216 Ck2 リン酸化部位 9-12;
Pkc リン酸化部位 32-34;

10

20

30

【 0 1 6 0 】

【 外 2 2 】

- DEX0253_217 Camp リン酸化部位 28-31;
ミリスチル 12-17;
Rieske_2 35-40;
- DEX0253_218 Ck2 リン酸化部位 21-24;
ミリスチル 29-34;
- DEX0253_221 Camp リン酸化部位 51-54;
Ck2 リン酸化部位 29-32;98-101;
ミリスチル 21-26;70-75;75-80;
Pkc リン酸化部位 50-52;57-59;
104-106;
Prokar リポタンパク質
66-76;68-78;
- DEX0253_222 Camp リン酸化部位 51-54;
Ck2 リン酸化部位 29-32;98-101;
ミリスチル 21-26;70-75;75-80;
Pkc リン酸化部位 50-52;57-59;
104-106;
Prokar リポタンパク質
66-76;68-78;
- DEX0253_223 Ck2 リン酸化部位 75-78;
- DEX0253_224 Ck2 リン酸化部位 96-99;
ミリスチル 72-77;76-81;105-110;
Pkc リン酸化部位 63-65;87-89;
- DEX0253_225 Asn グリコシル化 3-6;8-11;9-12;
Pkc リン酸化部位 12-14;

10

20

30

40

【 0 1 6 1 】

【 外 2 3 】

- DEX0253_226 Camp リン酸化部位 14-17;
DEX0253_228 Ck2 リン酸化部位 14-17;
DEX0253_229 Asn グリコシル化 41-44;
Ck2 リン酸化部位 20-23;39-42;
ミリスチル 25-30;30-35;37-42;45-50;
Pkc リン酸化部位 9-11;20-22;
DEX0253_230 Ck2 リン酸化部位 15-18;
DEX0253_231 Camp リン酸化部位 25-28;
DEX0253_232 Ck2 リン酸化部位 7-10; Rgd 8-10;
DEX0253_235 Ck2 リン酸化部位 21-24;
ミリスチル 4-9;
Pkc リン酸化部位 21-23;
DEX0253_236 Ck2 リン酸化部位 12-15;26-29;
ミリスチル 10-15;24-29;30-35;
DEX0253_237 Ck2 リン酸化部位 41-44;
DEX0253_238 Ck2 リン酸化部位 2-5;
Pkc リン酸化部位 46-48;
DEX0253_239 Pkc リン酸化部位 9-11;17-19;
DEX0253_240 Camp リン酸化部位 2-5;
Ck2 リン酸化部位 5-8;

10

20

30

【 0 1 6 2 】

【 外 2 4 】

- DEX0253_241 Pkc リン酸化部位 20-22;
 DEX0253_242 Ck2 リン酸化部位 30-33;
 ミリスチル 4-9;27-32;36-41;
 DEX0253_243 Asn グリコシル化 40-43;
 Pkc リン酸化部位 12-14;48-50;
 DEX0253_245 Asn グリコシル化 7-10;
 グリコサミノグリカン 9-12;
 ミリスチル 10-15;19-24;
 Pkc リン酸化部位 23-25;
 DEX0253_246 Ck2 リン酸化部位 3-6;35-38;
 DEX0253_247 Ck2 リン酸化部位 12-15;27-30;
 DEX0253_248 グリコサミノグリカン 30-33;
 ミリスチル 18-23;
 Pkc リン酸化部位 22-24;
 DEX0253_250 Asn グリコシル化 21-24;
 Ck2 リン酸化部位 14-17;28-31;
 ミリスチル 26-31;

10

20

【0163】

例6：ポリヌクレオチドに対応する遺伝子における変化の決定方法

30

RNAは、対象となる表現型を有する個々の患者または個体の家族からのRNAの単離によって決定される。次に、当該技術分野において既知のプロトコルを用いて、これらのRNA試料からcDNAが生成される。これらの技法を詳述した多くの実験室マニュアルの代表例である、例えば、上記Sambrook（2001）を参照。次いで、cDNAを、配列番号1～147において対象となる領域を取り囲むプライマーを用いるPCR用の鋳型として使用する。PCR条件は、Sidransky et al., Science 252（5006）：706-9（1991）に記載される緩衝溶液を用いて、95℃での30秒間、52～58℃での60～120秒、および70℃での60～120秒の35サイクルから構成される。Sidransky et al., Science 278（5340）：1054-9（1997）も参照。

40

次に、SequiThermポリメラーゼ（Epicentre Technologies）を用いて、T4ポリヌクレオチドキナーゼで5'末端において標識したプライマーを使用して、PCR産物を配列決定する。選択エクソンのイントロン-エクソン境界もまた決定され、ゲノムPCR産物を分析して、結果を確認する。次に、突然変異が疑われるPCR産物をクローニングして、配列決定して、直接シーケンスの結果を確認する。Holton et al., Nucleic Acids Res., 19：1156（1991）に記載されるように、PCR産物をT尾状（T-tailed）ベクターにクローニングし、T7ポリメラーゼ（United States Biochemical）を用いて配列決定する。病気に冒されていない個体に存在しない突然変異により、病気に冒された個体を同定する。

50

ゲノム再配列もまた決定することができる。ゲノムクローンを、ジゴキシゲニンデオキシ
 ウリジン 5' - 三リン酸塩 (Boehringer Mannheim) でニック
 トランスレーションし、Johnson, C, et al., Methods Cell
 Biol. 35: 73-99 (1991) に記載されるように、FISHを実施す
 る。標識プローブを用いたハイブリダイゼーションは、対応するゲノム座への特異的ハイ
 ブリダイゼーションのための、大過剰のヒト c o t - 1 DNA を用いて行なわれる。4,
 6 - ジアミジノ - 2 - フェニルインドールおよびヨウ化プロピジウムで染色体を対比染色
 し、C および R バンドの組合せを生じさせる。正確なマッピングのための整列画像 (a l
 i g n e d i m a g e) は、冷却電荷結合素子カメラ (Photometrics, T
 u c s o n, A Z) および可変励起波長フィルターを併合したトリプルバンドフィルタ
 ーセット (Chroma Technology, Brattleboro, VT) を
 用いて得られる。同書。画像取込み、分析および染色体分画長測定は、I See Gra
 p h i c a l プログラムシステム (Inovision Corporation, Du
 r h a m, N C) を用いて行なう。プローブによりハイブリダイズされるゲノム領域の
 染色体変化は、挿入、欠失および転座として同定される。これらの変化は、関連疾患用診
 断マーカーとして用いられる。

10

【0164】

例7：生物学的試料においてポリペプチドの異常レベルを検出する方法

抗体 - サンドイッチ E L I S A を用いて、試料中、好ましくは生物試料中のポリペプチド
 を検出する。マイクロタイタープレートのウェルを、最終濃度 0 . 2 ~ 1 0 μ g / m l で
 、特異抗体により被覆する。抗体は、モノクローナルまたはポリクローナルであり、上記
 の方法により産生される。ウェルへのポリペプチドの非特異的結合が低減するように、ウ
 ェルをブロッキングする。次に、被覆したウェルを、ポリペプチドを含有する試料ととも
 に、室温にて2時間より長い時間インキュベートする。好ましくは、試料の連続希釈物を用
 いて、結果を確認すべきである。次に、プレートを脱イオン水または蒸留水で3回洗浄
 して、未結合のポリペプチドを除去する。次に、25 ~ 400 n g の濃度の特異抗体 - ア
 ルカリホスファターゼ複合体 50 μ l を添加して、室温にて2時間インキュベートする。
 再びプレートを脱イオン水または蒸留水で3回洗浄し、未結合の複合体を除去する。次に
 、各ウェルに、75 μ l の 4 - メチルウンベリフェリルホスフェート (M U P) または p
 - ニトロフェニルホスフェート (N P P) 基質溶液を添加し、室温にて1時間インキュベ
 ートする。

20

30

マイクロタイタープレートリーダーにより、反応を測定する。対照試料の連続希釈物を用
 いて、標準曲線を作成し、X 軸上にポリペプチド濃度を (対数目盛)、Y 軸上に蛍光また
 は吸光度 (均等目盛) をプロットする。標準曲線を用いて、試料中のポリペプチド濃度を
 計算する。

【0165】

例8：ポリペプチドの製剤化

分泌ポリペプチド組成物は、個々の患者の臨床状態 (特に、分泌ポリペプチド単独での処
 置の副作用)、送達部位、投与方法、投与のスケジューリング、および開業医に既知の他
 の要素を考慮して、良質の医療のための原則 (g o o d m e d i c a l p r a c t i c
 e) に沿った様式で、処方されて、投薬される。したがって、本明細書の目的に関する「
 有効量」は、かかる考慮により決定される。

40

一般的な提案として、1回用量あたり非経口的に投与される分泌ポリペプチドの薬学的総
 有効量は、約 1 μ g / k g (患者の体重) / 日 ~ 10 m g / k g (患者の体重) / 日の範
 囲であろうが、上述するように、これは治療上の判断に左右されるであろう。より好まし
 くは、この用量は、少なくとも 0 . 01 m g / k g / 日であり、ヒトに関して最も好まし
 くは、ホルモンに関して約 0 . 01 ~ 1 m g / k g / 日である。連続的に与えられる場合
 、分泌ポリペプチドは、典型的には、1日あたり1 ~ 4回の注射により、または連続皮下
 注入により (例えば、ミニポンプを用いて)、約 1 μ g / k g / 時 ~ 約 50 m g / k g /
 時の用量速度にて投与される。点滴バッグ (i n t r a v e n o u s b a g) 溶液もま

50

た、用いてもよい。変化を観察するのに必要な処置の期間、および処置後に応答が生じる間隔は、所望の効果に依存して変わるようである。

本発明の分泌タンパク質を含有する薬学的組成物は、経口的に、経直腸的に、非経口的に、大槽内に、腔内に、腹腔内に、局所的に（パウダー、軟膏、ジェル、滴薬または経皮パッチによるものなど）、口腔に、または口腔もしくは鼻スプレーとして、投与される。「薬学的に許容し得る担体」は、無毒性の固形、半固形または液体賦形剤、希釈剤、カプセル用材料または任意のタイプの配合助剤を指す。本明細書で使用する「非経口」という用語は、静脈内、筋内、腹腔内、胸骨内、皮下、および関節内注射ならびに注入を含む投与様式を指す。

分泌ポリペプチドはまた、持続放出系により適切に投与される。持続放出性組成物の適切な例としては、造形品、例えば、フィルムまたはマイクロカプセルの形態にある半透性ポリマーマトリックスが挙げられる。持続放出性マトリックスとしては、ポリラクチド（米国特許第3,773,919号およびEP 58481）、L-グルタミン酸およびガンマーエチルーL-グルタメートのコポリマー（Sidman, U. et al., Biopolymers 22: 547-556 (1983)）、ポリ（2-ヒドロキシエチルメタクリレート）（R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981)、およびR. Langer, Chem. Tech. 12: 98-105 (1982)）、エチレンビニルアセテート（R. Langer et al.）、またはポリ-D-（-）-3-ヒドロキシ酪酸（EP 133,988）が挙げられる。持続放出性組成物はまた、リポソームエントラップポリペプチドを包含する。分泌ポリペプチドを含有するリポソームは、自体公知の方法により調製される（Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985)、Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980)、EP 52,322、EP 36,676、EP 88,046、EP 143,949、EP 142,641、日本国特許出願第83-118008号、米国特許第4,485,045号および第4,544,545号、ならびにEP 102,324）。通常、リポソームは、小さい（約200~800オングストローム）単層型であり、そこでの脂質含量は、約30molパーセントのコレステロールよりも高く、選択比率は、最適な分泌ポリペプチド治療に関して調節される。

【0166】

非経口投与に関して、一態様では、分泌ポリペプチドは、一般的に所望の純度にて、単位投与量の注射可能な形態（溶液、懸濁液、またはエマルジョン）で、薬学的に許容し得る担体、すなわち、使用する投与量および濃度にてレシピエントに無毒性であり、配合物の他の成分と併用可能（compatible）であるもの）と混合することにより配合される。例えば、配合物は、好ましくは、酸化剤、およびポリペプチドにとって有害であると知られている他の化合物を含まない。一般に、配合物は、液体の担体または微細固体の担体、あるいはその両方と、均一かつ緊密にポリペプチドを接触させることにより調製される。次に、必要な場合には、生成物を所望の配合物に成形する。好ましくは、担体は、非経口担体、より好ましくは、レシピエントの血液と等張である溶液である。かかる担体媒質の例としては、水、生理食塩水、リンゲル溶液、およびブドウ糖溶液が挙げられる。不揮発性油およびオレイン酸エチルのような非水性媒質、ならびにリポソームもまた、本発明に有用である。

担体は適切に、等張性および化学的安定性を高める物質などの、少量の添加剤を含有する。かかる物質は、用いる投与量および濃度にて、レシピエントに対して無毒性であり、リン酸塩、クエン酸塩、コハク酸塩、酢酸、および他の有機酸またはそれらの塩などの緩衝液、アスコルビン酸などの酸化防止剤、低分子量（約10残基未満）ポリペプチド、例えば、ポリアルギニンまたはトリペプチド、血清アルブミン、ゼラチンまたは免疫グロブリンなどのタンパク質、ポリビニルピロリドンなどの親水性ポリマー、グリシン、グルタミン酸、アスパラギン酸またはアルギニンなどのアミノ酸、セルロースまたはその誘導体、

グルコース、マンノース、またはデキストリンを含めた単糖類、二糖類、および他の炭化水素、EDTAなどのキレート剤、マンニトールまたはソルビトールなどの糖アルコール、ナトリウムなどの対イオン、および／またはポリソルベート、ポロキサマー、またはPEGなどの非イオン性界面活性剤を含む。

分泌ポリペプチドは、典型的に、約0.1mg/ml～100mg/ml、好ましくは1～10mg/mlの濃度で、約3～8のpHにて、かかるビヒクル中で配合される。先述の賦形剤、担体または安定剤のうちのあるものの使用は、結果としてポリペプチド塩の形成を生じる結果となろうことは理解されるであろう。

【0167】

治療用投与のために使用されるべきポリペプチドはいずれも無菌であることができる。無菌状態は、滅菌濾過膜（例えば、0.2ミクロンの膜）を通した濾過により容易に達成される。治療用ポリペプチド組成物は一般に、滅菌アクセスポートを有する容器、例えば、皮下注射針により突き刺すことが可能なストッパーを有する点滴溶液バッグまたはバイアルに入れられる。

ポリペプチドは通常、単用量または複数用量容器（例えば、密封アンプルまたはバイアル）中に、水溶液として、または再構成用凍結乾燥配合物として、保管されるであろう。凍結乾燥配合物の例として、10mlバイアルを滅菌濾過した1%（w/v）ポリペプチド水溶液5mlで充填し、得られた混合物を凍結乾燥する。注入溶液は、注射用静菌水を用いて、凍結乾燥ポリペプチドを再構成することにより調製される。

本発明はまた、本発明の医薬組成物の1または2以上の成分で充填された1または2以上の容器を含む医薬パックまたはキットを提供する。1または2以上のかかる容器には、医薬品または生物製品の製造、使用または販売を管轄する政府機関が定めた形式で書かれた、該機関による、ヒトへの投与に関する製品の製造、使用または販売の承認を示す通知を添付することができる。さらに、本発明のポリペプチドは、タンパク質の治療的化合物と組合わせて用いることができる。

【0168】

例9：ポリペプチドの減少したレベルを処置する方法

個体において分泌タンパク質の標準または正常発現レベルにおける減少により引き起こされる症状は、好ましくは分泌形態にある、本発明のポリペプチドを投与することにより処置できることが理解されるであろう。したがって、本発明はまた、ポリペプチドレベルの増加を必要とする個体を処置する方法であって、かかる個体においてポリペプチドの活性レベルを増加させるためのポリペプチドの量を含む医薬組成物を、かかる個体に投与することを含む方法を提供する。

例えば、減少したポリペプチドレベルを有する患者は、1日用量0.1～100μg/kgのポリペプチドを6日間連続で受けてもよい。好ましくは、ポリペプチドは分泌形態である。投与および処方に基づく投薬スキームの正確な詳細は、上記に提供されている。

例10：ポリペプチドの増加したレベルを処置する方法

アンチセンス技術を、本発明のポリペプチドの産生を抑制するために用いる。この技術は、癌などの様々な病因に起因するポリペプチド、好ましくは分泌形態のもののレベルを減少させる方法の一例である。

例えば、異常に増加したポリペプチドレベルを有すると診断された患者に、0.5、1.0、1.5、2.0および3.0mg/kg/日にて、21日間、アンチセンスポリヌクレオチドを静脈内に投与することができる。この処置が十分に耐容された場合、処置は7日の休止期間後に繰り返される。アンチセンスポリヌクレオチドの製剤化は、上記に提供されている。

【0169】

例11：遺伝子療法を用いた処置の方法

遺伝子治療の一方法では、ポリペプチドを発現することが可能な線維芽細胞を、患者に移植する。一般に、線維芽細胞は、皮膚生検により対象から得られる。得られた組織を組織培養培地に置き、小片に分離する。組織の小片を組織培養フラスコの湿った表面上へ載せ

、各フラスコにおいて約10片を載せる。フラスコを上下ひっくり返し、きつく閉めて、室温で一晩放置する。室温にて24時間後、フラスコをひっくり返し、組織片をフラスコの底に固定させたままで、新鮮な培地（例えば、Ham's F12培地、10% FBS、ペニシリンおよびストレプトマイシン加）を添加する。次に、フラスコを37℃にて約1週間インキュベートする。

この時点で、新鮮な培地を添加し、続いて数日毎に取り換える。さらに2週間の培養の後、線維芽細胞の単層が現れる。単層をトリプシン処理し、大きなフラスコへはがし取る。モロニーマウス肉腫ウイルスの長い末端反復配列に隣接したpMV-7 (Kirschmeyer, P. T. et al., DNA, 7: 219-25 (1988)) を EcoRI および HindIII で消化した後、仔ウシ腸ホスファターゼ (calf intestinal phosphatase) で処理する。アガロースゲル上で線状ベクターを分画し、ガラスビーズを用いて精製する。

本発明のポリペプチドをコードするcDNAを、例1に記載のとおり、それぞれ5'および3'末端配列に相当するPCRプライマー用いて増幅することができる。好ましくは、5'プライマーは、EcoRI部位を含有し、3'プライマーは、HindIII部位を含む。等量のモロニーマウス肉腫ウイルス線状主鎖ならびに増幅したEcoRIおよびHindIII断片を、T4DNAリガーゼの存在下、一緒に添加する。得られた混合物を、2つの断片の連結に適した条件下において保持する。次に、連結混合物を用いて、バクテリアHB 101を形質転換した後、ベクターが適切に挿入された所定の遺伝子を有することを確認する目的で、カナマイシンを含有する寒天上に蒔く。

【0170】

両栄養性pA317またはGP+am12パッケージング細胞を、10%の仔ウシ血清(CS)、ペニシリンおよびストレプトマイシン加ダルベッコ変法イーグル培地(DMEM)中で、コンフルエント密度に、組織培養で増殖させる。次に、遺伝子を含有するMSVベクターを培地に添加し、パッケージング細胞にベクターで形質導入する。すると、パッケージング細胞は、遺伝子を含有する感染ウイルス粒子を生産する(パッケージング細胞は、今度はプロデューサー細胞と称される)。

形質導入したプロデューサー細胞に、新鮮な培地を添加し、続いて、コンフルエントなプロデューサー細胞の10cmプレートから、培地を採取する。ミリポアフィルターを通して感染ウイルス粒子を含有する使用済培地を濾過し、剥離したプロデューサー細胞を除去し、続いてこの培地を用いて、線維芽細胞を感染させる。線維芽細胞のサブコンフルエントなプレートから、培地を除去し、プロデューサー細胞からの培地とすばやく入れ換える。この培地を除去し、新鮮な培地と入れ換える。ウイルスの力価が高い場合、実質的にすべての線維芽細胞は感染し、選択は必要ない。力価が非常に低い場合、neoまたはhisのような選択マーカーを有するレトロウイルスベクターを使用する必要がある。線維芽細胞がいったん効果的に感染したら、線維芽細胞を分析して、タンパク質が産生されるかどうかを決定する。次に、操作した線維芽細胞を、単独で、またはcytodex3マイクロキャリアビーズ上においてコンフルエントまで増殖させた後に、宿主に移植する。

【0171】

例12: *in vivo* 遺伝子治療を用いた処置の方法

本発明の他の側面は、障害、疾患および症状を処置するために、*in vivo* 遺伝子治療方法を用いることである。遺伝子治療方法は、ポリペプチドの発現を増加または減少させるための、裸の核酸(DNA、RNA、およびアンチセンスDNAまたはRNA)配列の動物への導入に関する。

本発明のポリヌクレオチドは、プロモーターに、または標的組織によって、ポリペプチド発現に必要な任意の他の遺伝要素に作動可能に連結されてもよい。かかる遺伝子治療ならびに送達技術および方法は、当該技術分野で既知であり、例えば、WO 90/11092、WO 98/11779、米国特許第5,693,622号、第5,705,151号、第5,580,859号、Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479、Chao J et al. (

10

20

30

40

50

1997) Pharmacol. Res. 35 (6): 517-522、Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318、Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (参照により本明細書に組込まれる)を参照されたい。

ポリヌクレオチド構築物は、組織(心臓、筋肉、皮膚、大腸、肝臓、腸など)の間質腔への注射などの、動物細胞へ注入可能な物質を送達するいかなる方法により送達されてもよい。ポリヌクレオチド構築物は、薬学的に許容し得る液体または水性担体中で送達することができる。

10

「裸の」ポリヌクレオチド、DNAまたはRNAという用語は、細胞への侵入を補助、促進または助長するよう作用する、ウイルス配列、ウイルス粒子、リポソーム配合物、リポフェクチンまたは沈殿剤などを含むいかなる送達ビヒクルも含まない配列を指す。しかしながら、本発明のポリヌクレオチドはまた、当業者に既知の方法により調製され得るリポソーム配合物(例えば、Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139、およびAbdallah B. et al. (1995) Biol. Cell 85 (1): 1-7に教示されているもの)にて送達されてもよい。

遺伝子治療方法において使用されるポリヌクレオチドベクター構築物は、好ましくは、宿主ゲノムに組み込まれない構築物であり、またそれらは、複製が可能な配列を含有しないであろう。当業者に既知のいかなる強力なプロモーターも、DNA発現を促進するために使用することができる。他の遺伝子治療技術と異なり、標的細胞へ裸の核酸配列を導入することの1つの主要な利点は、細胞におけるポリヌクレオチド合成の一過性の性質である。研究により、非複製DNA配列が細胞に導入されて、最大6ヶ月の期間、所望のポリペプチドの産生を提供することができることがわかった。

20

【0172】

ポリヌクレオチド構築物は、筋肉、皮膚、脳、大腸、肝臓、脾臓、骨髄、胸腺、心臓、リンパ、血液、骨、軟骨、膵臓、腎臓、胆嚢、胃、腸、精巣、卵巣、子宮、直腸、神経系、眼、腺、および結合組織を含む、動物内組織の間質腔へ送達され得る。組織の間質腔は、細胞間液、器官組織の細網線維間のムコ多糖マトリックス、管または房の壁における弾性線維、線維性組織のコラーゲン線維、または筋細胞を覆う結合組織内もしくは骨小腔中にある同じマトリックスを含む。循環血漿およびリンパ管のリンパ液により占有される腔も同様である。下記で論じる理由により、筋組織の間質腔への送達が好ましい。それらは、これらの細胞を含む組織への注射により利便性よく送達され得る。送達および発現は、未分化またはあまり完全には分化していない細胞、例えば、血液の幹細胞または皮膚線維芽細胞などにおいて達成され得るが、それらは、好ましくは、分化した永続性の非分裂細胞に送達され、そこで発現される。in vivoの筋細胞は特に、ポリヌクレオチドを取り込み、発現するそれらの能力に優れている。

30

裸のポリヌクレオチド注射に関して、有効投与量のDNAまたはRNAは、約0.05 μ g/体重kg~約50mg/体重kgの範囲であろう。好ましくは、投与量は、0.005mg/kg~約20mg/kgであり、より好ましくは約0.05mg/kg~約5mg/kgであろう。当然のことながら、当業者が理解するように、この投与量は、注射の組織部位により変化するであろう。適切かつ有効な投与量の核酸配列は、当業者により容易に決定することができ、それは、処置する症状および投与経路に依存し得る。好ましい投与経路は、組織の間質腔への注射の非経口経路によるものである。しかしながら、特に大腸もしくは気管支組織、咽喉または鼻の粘膜への送達用のエアロゾル配合物の吸入などの、他の非経口経路もまた使用してもよい。さらに、裸のポリヌクレオチド構築物は、血管形成中に、当該手順で使用されるカテーテルにより、動脈へ送達することができる。

40

【0173】

筋肉中に注射されたポリヌクレオチドのin vivoでの用量反応効果は以下のように

50

決定される。本発明のポリペプチドをコードするmRNAの産生のための適切な鋳型DNAを、標準的な組換えDNA方法論にしたがって調製する。鋳型DNAは、環状または線状のいずれであってもよく、それは裸のDNAとして用いられるか、あるいはリポソームと複合体を形成する。次に、マウスの四頭筋に、様々な量の鋳型DNAを注射する。

5～6週齢の雌および雄Balb/Cマウスを、2.5%アベルチン0.3mlの腹腔内注射により麻酔する。腹側大腿上に1.5cmの切開を施し、四頭筋を直接見えるようにする。筋肉の膝への遠位付着部位から約0.5cmかつ約0.2cmの深さで、1分にわたり、27ゲージ針を通じて、1ccの注射器内の担体0.1ml中の鋳型DNAを注射する。将来の位置確認のために注入部位上に縫合糸を設置し、皮膚をステンレス製クリップで閉じる。

適切なインキュベーション時間後（例えば、7日）、四頭筋全体を切除することにより、筋肉摘出物を調製する。個々の四頭筋のすべての5番目の15μm断面を、タンパク質発現のために組織化学的に染色する。タンパク質発現に関するタイムコースは、種々のマウス由来の四頭筋が異なる時間にて採取されることを除いて、同様の様式にて行なわれ得る。注射後の筋肉中のDNAの永続性は、注射したマウスおよび対照マウスから全細胞DNAおよびHIRT上清を調製した後に、サザンブロット分析により確定されてもよい。マウスでの上記実験の結果を用いて、裸のDNAを用いた、ヒトおよび他の動物における適切な投与量および他の治療パラメータを外挿することができる。

【0174】

例13：トランスジェニック動物

本発明のポリペプチドはまた、トランスジェニック動物にて発現させることもできる。マウス、ラット、ウサギ、ハムスター、モルモット、ブタ、小型ブタ（micro-pig）、ヤギ、ヒツジ、ウシ、および非ヒト霊長類、例えば、ヒヒ、サル、およびチンパンジーが含まれるが、これらに限定されないあらゆる種の動物を用いて、トランスジェニック動物を生み出してもよい。特定の態様では、本明細書中に記載された技術、またはその他当該技術分野で知られている技術が、本発明のポリペプチドを、遺伝子治療手法の一環として、ヒトに発現させるのに用いられる。

当該技術分野で既知の任意の技法を用いて、動物に導入遺伝子（すなわち、本発明のポリペプチド）を導入して、トランスジェニック動物の創始系統を生産してもよい。かかる技法としては、前核（pronuclear）マイクロインジェクション（Peterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994)、Carver et al., Biotechnology (NY) 11: 1263-1270 (1993)、Wright et al., Biotechnology (NY) 9: 830-834 (1991); およびHoppe et al., 米国特許第4,873,191号(1989)、生殖系列（Vander Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)）、胚盤胞または胚へのレトロウイルス媒介性遺伝子移入、胚性幹細胞における遺伝子ターゲティング（Thompson et al., Cell 56: 313-321 (1989)）、細胞または胚へのエレクトロポレーション（Lo, 1983, Mol Cell Biol. 3: 1803-1814 (1983)）、遺伝子銃を用いた本発明のポリヌクレオチドの導入（例えば、Ulmer et al., Science 259: 1745 (1993)を参照）、胚性多能性幹細胞への核酸構築物の導入および胚盤胞への幹細胞の移し戻し、および精子媒介性遺伝子移入（Lavitrano et al., Cell 57: 717-723 (1989)）などが挙げられるが、これらに限定されない。かかる技法のレビューとして、Gordon, "Transgenic Animals", Intl. Rev. Cytol. 115: 171-229 (1989)を参照されたい（それは、参照により全体が本明細書に組込まれる）。

当該技術分野において既知のいかなる技法、例えば、静止状態へ誘導された培養胚細胞、胎児細胞または成人細胞からの核の、脱核卵母細胞への核移入（Campbell et al.

10

20

30

40

50

l., Nature 380: 64-66 (1966), Wilmut et al., Nature 385: 810-813 (1997))を用いて、本発明のポリヌクレオチドを含有するトランスジェニッククローンを生産してもよい。

【0175】

本発明は、すべての動物細胞において導入遺伝子を持つトランスジェニック動物、ならびに、すべてではないが幾つかの細胞において導入遺伝子を持つ動物、すなわちモザイク動物またはキメラ動物を提供する。導入遺伝子は、単一導入遺伝子として、またはコンカテマー（例えば、頭-頭タンデム、もしくは頭-尾タンデム）のような多重コピーとして、組み込まれてもよい。導入遺伝子はまた、例えば、Lasko et al. の教示 (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)) にしたがって、特定の細胞タイプに選択的に導入されて、そこで活性化されてもよい。かかる細胞タイプ特異的な活性化に必要な調節遺伝子は、所定の特定細胞タイプに依存し、それは当業者には明らかであろう。

ポリヌクレオチド導入遺伝子が内因性遺伝子の染色体部位に組み込まれることが望ましい場合、遺伝子ターゲティングが好ましい。簡潔に述べると、かかる技法を利用する場合、内因性遺伝子に相同的な幾つかのヌクレオチド配列を含有するベクターを、内因性遺伝子のヌクレオチド配列へ、染色体配列との相同組換えにより組み込み、その機能を崩壊させる目的で設計する。導入遺伝子はまた、特定の細胞タイプへ選択的に導入されてもよく、したがって、例えば Gu et al. (Gu et al., Science 265: 103-106 (1994)) の教示にしたがって、その細胞タイプのみで内因性遺伝子を不活性化させる。かかる細胞タイプに特異的な不活性化に必要な調節配列は、所定の特定細胞タイプに依存し、それは当業者には明らかであろう。

トランスジェニック動物がいったん創出されれば、組換え遺伝子の発現は、標準的な技法を利用してアッセイされ得る。初期スクリーニングは、導入遺伝子の組み込みが起ったことを確認すべく動物組織を分析するために、サザンブロット分析またはPCR技法により達成され得る。トランスジェニック動物の組織における導入遺伝子のmRNA発現レベルはまた、動物から得られた組織試料のノーザンブロット分析、in situ ハイブリダイゼーション分析、および逆転写酵素PCR (reverse PCR) が含まれるが、これらに限定されない技法を用いて評価してもよい。トランスジェニック遺伝子発現組織の試料はまた、導入遺伝子産物に特異的な抗体を用いて、免疫細胞化学的に、または免疫組織化学的に評価されてもよい。

【0176】

創始動物がいったん生産されれば、それらを、繁殖、同系交配、異系交配、または交雑させて、特定の動物コロニーを生産してもよい。かかる繁殖戦略の例としては、別個の系統を樹立するために、1つよりも多い組み込み部位を有する創始動物を異系交配すること、各導入遺伝子の付加的発現の効果が理由で、高いレベルで導入遺伝子を発現する複合遺伝子導入をもたらすために、別個の系統を同系交配させること、発現を増大させ、かつDNA分析による動物のスクリーニングの必要性をなくすために、所定の組み込み部位に関してホモ接合性の動物を生産するために、ヘテロ接合性のトランスジェニック動物を交雑させること、複合ヘテロ接合性またはホモ接合性系統を生産するために、別個のホモ接合性系統を交雑させること、および当該実験モデルに適切な明瞭なバックグラウンドに導入遺伝子を配置させるために繁殖させることが挙げられるが、これらに限定されない。

本発明のトランスジェニック動物は、本発明のポリペプチドの生物機能を説明し、異常発現に関連した症状および／または疾患を研究し、かかる症状および／または疾患を回復するのに効果的な化合物をスクリーニングするのに有用な動物モデル系を含むが、これに限定されない用途を有する。

【0177】

例14：ノックアウト動物

内因性の遺伝子発現はまた、標的 (targeted) 相同組換えを用いて、遺伝子および／もしくはそのプロモーターを不活性化または「ノックアウト」することにより低減す

ることができる（例えば、Smithies et al., Nature 317: 230-234 (1985)、Thomas & Capecchi, Cell 51: 503-512 (1987)、Thompson et al., Cell 5: 313-321 (1989)を参照。それらはそれぞれ、その全体が、参照により本明細書に組込まれる）。例えば、内在性のポリヌクレオチド配列（遺伝子のコード領域または調節領域）に相同的なDNAに隣接した突然変異体である本発明の非機能性ポリヌクレオチド（または完全に非関連のDNA配列）を、選択マーカーおよび／またはネガティブ選択マーカーを用いて、あるいは用いずに使用して、in vivoで本発明のポリペプチドを発現する細胞をトランスフェクトすることができる。別の態様では、当該技術分野にて既知の技法を用いて、所定の遺伝子を含むが、発現はしない細胞におけるノックアウトを生成する。標的相同組換えを介したDNA構築物の挿入は、結果として標的遺伝子の不活性化を招く。かかるアプローチは、研究および農業の分野に特に適しており、そこで胚性幹細胞の改変を用いて、不活性化標的遺伝子を有する動物の子孫を発生させることができる（例えば、上述のThomas & Capecchi 1987およびThompson 1989を参照）。しかしながら、組換えDNA構築物が、当業者には明らかであろう適切なウイルスベクターを用いて、in vivoにおいて、必要とされる部位に直接的に投与されるか、または標的にされるという条件で、このアプローチは、ヒトにおける使用にルーチンで適合され得る。

本発明のさらなる態様では、本発明のポリペプチドを発現するように遺伝子操作された細胞、あるいは本発明のポリペプチドを発現しないように遺伝子操作された（例えば、ノックアウト）細胞を、in vivoにおいて患者に投与する。かかる細胞は、患者（即ち、ヒトを含む動物）、またはMHC適合ドナーから得てもよく、線維芽細胞、骨髄細胞、血液細胞（例えば、リンパ球）、脂肪細胞、筋細胞、内皮細胞などを含むことができるが、これらに限定されない。細胞は、組換えDNA技法を用いて、in vitroにおいて遺伝子操作され、細胞に本発明のポリペプチドのコード配列を導入するか、あるいは例えば形質導入（ウイルスベクター、好ましくは細胞ゲノムに導入遺伝子を組み込むベクターを用いて）またはプラスミド、コスミド、YAC、裸のDNA、エレクトロポレーション、リポソーム等の使用を含むが、これらに限定されないトランスフェクション手法により、本発明のポリペプチドに関するコード配列および／または内因性調節配列を崩壊させる。

本発明のポリペプチドのコード配列は、強力な構成もしくは誘導プロモーターまたはプロモーター／エンハンサーの制御下に置くことにより、本発明のポリペプチドの発現、好ましくは分泌を達成することができる。本発明のポリペプチドを発現、好ましくは分泌する操作細胞を全身的に、例えば、循環にて、または腹腔内にて患者に導入することができる。

あるいは、細胞をマトリックス中に組み込み、体内へ移植することができ、例えば、遺伝子操作した線維芽細胞は、皮膚移植片の一部として移植することができる。遺伝子操作した内皮細胞は、リンパ管または血管移植片の一部として移植することができる（例えば、各々その全体が参照により本明細書に組込まれるAnderson et al., 米国特許第5, 399, 349号、およびMulligan & Wilson, 米国特許第5, 460, 959号を参照）。

【0178】

投与されるべき細胞が非自系または非MHC適合細胞である場合、それらは、導入細胞に対する宿主免疫応答の発生を防ぐ既知の技法を用いて、投与され得る。例えば、細胞は、直に接している細胞外環境との構成成分の交換が可能であるものの、導入細胞が宿主免疫系により認識されることが不可能なカプセル化形態で導入されてもよい。

本発明のトランスジェニックおよび「ノックアウト」動物は、本発明のポリペプチドの生物機能を説明し、異常な発現に関連した症状および／または障害を研究し、かかる症状および／または疾患を改善するのに効果的な化合物をスクリーニングするのに有用な動物モデル系を含むが、これらに限定されない用途を有する。

10

20

30

40

50

本明細書に引用されたすべての特許、特許文献および他の出版された参考文献は、そのそれぞれが、本明細書に参照によって個別的かつ具体的に組込まれたかのように、その全体が参照によって本明細書に組込まれる。本発明の好ましい代表的な態様が記載されているが、当業者は、本発明が、限定のためではなく、例示の目的でのみ提示された、記載された以外の態様で実施できることを理解するだろう。本発明は、請求の範囲によってのみ限定される。

【国際公開パンフレット】

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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
20 June 2002 (20.06.2002)

PCT

(10) International Publication Number
WO 02/48370 A2(51) International Patent Classification: C12N 15/12,
C12Q 1/68, C12N 1/21, 5/10, C07K 14/47, 16/18, G01N
33/68, A61K 39/395, 38/17, 31/713LIU, Chenghua [CN/US]; 1125 Ranchero Way #14, San
Jose, CA 95117 (US).

(21) International Application Number: PCT/US01/51341

(74) Agents: LICATA, June, Mussey et al.; Licata & Tyrrell
PC, 66 E. Main Street, Marlton, NJ 08053 (US).

(22) International Filing Date: 30 October 2001 (30.10.2001)

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EP, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MY, NZ, NL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 60/244,717 31 October 2000 (31.10.2000) US

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/244,717 (CIP)
Filed on 31 October 2000 (31.10.2000)(84) Designated States (regional): ARIPO patent (GL, GM,
KH, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).(71) Applicant (for all designated States except US): DI-
ADENIX, INC., (US/CA), 443 Oyster Point Boulevard,
South San Francisco, CA 94080 (US).Published:
— without international search report and to be republished
upon receipt of that report

(72) Inventors; and

(75) Inventors/Applicants (for US only): SUN, Yongming
[CN/US], 551 Shou Drive, Redwood City, CA 94065
(US), RECIPOX, Horve [FR/FR], 85 Fortuna Avenue,
San Francisco, CA 94115 (US), GHOSIL, Mahavika, G.
[IN/US], 1733 Via Laganey, San Jose, CA 95129 (US).For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/48370 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic colon cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon tissue, identifying colon tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered colon tissue for treatment and research.

WO 02/48370

PCT/US01/51341

-1-

**COMPOSITIONS AND METHODS
RELATING TO COLON SPECIFIC GENES AND PROTEINS**

This application claims the benefit of priority from U.S. Provisional Application
5 Serial No. 60/244,717 filed October 31, 2000, which is herein incorporated by reference
in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and
10 polypeptides present in normal and neoplastic colon cells, including fragments, variants
and derivatives of the nucleic acids and polypeptides. The present invention also relates
to antibodies to the polypeptides of the invention, as well as agonists and antagonists of
the polypeptides of the invention. The invention also relates to compositions comprising
the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists
15 of the invention and methods for the use of these compositions. These uses include
identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-
cancerous disease states in colon tissue, identifying colon tissue and monitoring and
identifying and/or designing agonists and antagonists of polypeptides of the invention.
The uses also include gene therapy, production of transgenic animals and cells, and
20 production of engineered colon tissue for treatment and research.

BACKGROUND OF THE INVENTION

Colorectal cancer is the second most common cause of cancer death in the United
States and the third most prevalent cancer in both men and women. M. L. Davila & A.
D. Davila, *Screening for Colon and Rectal Cancer*, in Colon and Rectal Cancer 47 (Peter
25 S. Edelstein ed., 2000). Approximately 100,000 patients every year suffer from colon
cancer and approximately half that number die of the disease. Hannah-Ngoc Ha & Bard
C. Cosman, *Treatment of Colon Cancer*, in Colon and Rectal Cancer 157 (Peter S.
Edelstein ed., 2000). Nearly all cases of colorectal cancer arise from adenomatous
polyps, some of which mature into large polyps, undergo abnormal growth and
30 development, and ultimately progress into cancer. Davila & Davila, *supra* at 55-56.
This progression would appear to take at least 10 years in most patients, rendering it a

WO 02/48370

PCT/US01/51341

-2-

readily treatable form of cancer if diagnosed early, when the cancer is localized. *Id.* at 56; Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 125 (1998).

Although our understanding of the etiology of colon cancer is undergoing continual refinement, extensive research in this area points to a combination of factors, including age, hereditary and nonhereditary conditions, and environmental/dietary factors. Age is a key risk factor in the development of colorectal cancer, Davila & Davila, *supra* at 48, with men and women over 40 years of age become increasingly susceptible to that cancer, Burdette, *supra* at 126. Incidence rates increase considerably in each subsequent decade of life. Davila et al., *supra* at 48. A number of hereditary and nonhereditary conditions have also been linked to a heightened risk of developing colorectal cancer, including familial adenomatous polyposis (FAP), hereditary nonpolyposis colorectal cancer (Lynch syndrome or HNPCC), a personal and/or family history of colorectal cancer or adenomatous polyps, inflammatory bowel disease, diabetes mellitus, and obesity. *Id.* at 47; Henry T. Lynch & Jane F. Lynch, *Hereditary Nonglycopolis Colorectal Cancer (Lynch Syndromes)*, in Colon and Rectal Cancer 67-68 (Peter S. Edelstein ed., 2000).

In the case of FAP, the tumor suppressor gene APC (adenomatous polyposis coli), located at 5q21, has been either mutationally inactivated or deleted. Alberts et al., Molecular Biology of the Cell 1288 (3d ed. 1994). The APC protein plays a role in a number of functions, including cell adhesion, apoptosis, and repression of the *c-myc* oncogene. N. R. Hall & R. D. Madoff, *Genetics and the Poly-Cancer Sequence*, Colon and Rectal Cancer 8 (Peter S. Edelstein, ed., 2000). Of those patients with colorectal cancer who have normal APC genes, over 65% have such mutations in the cancer cells but not in other tissues. Alberts et al., *supra* at 1288. In the case of HNPCC, patients manifest abnormalities in the tumor suppressor gene HNPCC, but only about 15% of tumors contain the mutated gene. *Id.* A host of other genes have also been implicated in colorectal cancer, including the *K-ras*, *N-ras*, *H-ras* and *c-myc* oncogenes, and the tumor suppressor genes *DCC* (deleted in colon carcinoma) and *p53*. Hall & Madoff, *supra* at 8-9; Alberts et al., *supra* at 1288.

Environmental/dietary factors associated with an increased risk of colorectal cancer include a high fat diet, intake of high dietary red meat, and sedentary lifestyle. Davila & Davila, *supra* at 47; Reddy, B. S., *Prev. Med.* 16(4): 460-7 (1987).

WO 02/48370

PCT/US01/51341

-3-

Conversely, environmental/dietary factors associated with a reduced risk of colorectal cancer include a diet high in fiber, folic acid, calcium, and hormone-replacement therapy in post-menopausal women. Davila & Davila, *supra* at 50-55. The effect of antioxidants in reducing the risk of colon cancer is unclear. *Id.* at 53.

5 Because colon cancer is highly treatable when detected at an early, localized stage, screening should be a part of routine care for all adults starting at age 50, especially those with first-degree relatives with colorectal cancer. One major advantage of colorectal cancer screening over its counterparts in other types of cancer is its ability to not only detect precancerous lesions, but to remove them as well. Davila & Davila,
10 *supra* at 56. The key colorectal cancer screening tests in use today are fecal occult blood test, sigmoidoscopy, colonoscopy, double-contrast barium enema, and the carcinoembryonic antigen (CEA) test. *Id.*; Burdette, *supra* at 125.

The fecal occult blood test (FOBT) screens for colorectal cancer by detecting the amount of blood in the stool, the premise being that neoplastic tissue, particularly
15 malignant tissue, bleeds more than typical mucosa, with the amount of bleeding increasing with polyp size and cancer stage. Davila & Davila, *supra* at 56-57. While effective at detecting early stage tumors, FOBT is unable to detect adenomatous polyps (premalignant lesions), and, depending on the contents of the fecal sample, is subject to rendering false positives. *Id.* at 56-59. Sigmoidoscopy and colonoscopy, by contrast,
20 allow direct visualization of the bowel, and enable one to detect, biopsy, and remove adenomatous polyps. *Id.* at 59-60, 61. Despite the advantages of these procedures, there are accompanying downsides: sigmoidoscopy, by definition, is limited to the sigmoid colon and below, colonoscopy is a relatively expensive procedure, and both share the risk of possible bowel perforation and hemorrhaging. *Id.* at 59-60. Double-contrast barium
25 enema (DCBE) enables detection of lesions better than FOBT, and almost as well as colonoscopy, but it may be limited in evaluating the winding rectosigmoid region. *Id.* at 60. The CEA blood test, which involves screening the blood for carcinoembryonic antigen, shares the downside of FOBT, in that it is of limited utility in detecting colorectal cancer at an early stage. Burdette, *supra* at 125.

30 Once colon cancer has been diagnosed, treatment decisions are typically made in reference to the stage of cancer progression. A number of techniques are employed to stage the cancer (some of which are also used to screen for colon cancer), including

WO 02/48370

PCT/US01/51341

-4-

pathologic examination of resected colon, sigmoidoscopy, colonoscopy, and various imaging techniques. *AJCC Cancer Staging Handbook* 84 (Irvin D. Fleming et al. eds., 5th ed. 1998); Montgomery, R. C. and Ridge, J.A., *Semin. Surg. Oncol.* 15(3): 143-150 (1998). Moreover, chest films, liver functionality tests, and liver scans are employed to

5 determine the extent of metastasis. Fleming et al. eds., *supra* at 84. While computerized tomography and magnetic resonance imaging are useful in staging colorectal cancer in its later stages, both have unacceptably low staging accuracy for identifying early stages of the disease, due to the difficulty that both methods have in (1) revealing the depth of bowel wall tumor infiltration and (2) diagnosing malignant adenopathy. Thoeni, R. F.,

10 *Radiol. Clin. N. Am.* 35(2): 457-85 (1997). Rather, techniques such as transrectal ultrasound ("TRUS") are preferred in this context, although this technique is inaccurate with respect to detecting small lymph nodes that may contain metastases. David Blumberg & Frank G. Opelka, *Neoadjuvant and Adjuvant Therapy for Adenocarcinoma of the Rectum*, in *Colon and Rectal Cancer* 316 (Peter S. Edelstein ed., 2000).

15 Several classification systems have been devised to stage the extent of colorectal cancer, including the Dukes' system and the more detailed International Union against Cancer-American Joint Committee on Cancer TNM staging system, which is considered by many in the field to be a more useful staging system. Burdette, *supra* at 126-27. The TNM system, which is used for either clinical or pathological staging, is divided into four

20 stages, each of which evaluates the extent of cancer growth with respect to primary tumor (T), regional lymph nodes (N), and distant metastasis (M). Fleming et al. eds., *supra* at 84-85. The system focuses on the extent of tumor invasion into the intestinal wall, invasion of adjacent structures, the number of regional lymph nodes that have been affected, and whether distant metastasis has occurred. *Id.* at 81.

25 Stage 0 is characterized by *in situ* carcinoma (Tis), in which the cancer cells are located inside the glandular basement membrane (intraepithelial) or lamina propria (intramucosal). *Id.* at 84-85; Burdette, *supra* at 127. In this stage, the cancer has not spread to the regional lymph nodes (N0), and there is no distant metastasis (M0). Fleming et al. eds., *supra* at 85; Burdette, *supra* at 127. In stage I, there is still no spread

30 of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the submucosa (T1) or has progressed further to invade the muscularis propria (T2). Fleming et al. eds., *supra* at 84-85; Burdette, *supra* at 127. Stage II also involves

WO 02/48370

PCT/US01/51341

-5-

no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the subserosa, or the nonperitonealized pericolic or perirectal tissues (T3), or has progressed to invade other organs or structures, and/or has perforated the visceral peritoneum (T4). *Id.* Stage 3 is characterized by any of the T substages, no distant metastasis, and either metastasis in 1 to 3 regional lymph nodes (N1) or metastasis in four or more regional lymph nodes (N2). Fleming et al. eds., *supra* at 85; Burdette, *supra* at 127. Lastly, stage 4 involves any of the T or N substages, as well as distant metastasis. *Id.*

Currently, pathological staging of colon cancer is preferable over clinical staging as pathological staging provides a more accurate prognosis. Pathological staging typically involves examination of the resected colon section, along with surgical examination of the abdominal cavity. Fleming et al. eds., *supra* at 84. Clinical staging would be a preferred method of staging were it at least as accurate as pathological staging, as it does not depend on the invasive procedures of its counterpart.

Turning to the treatment of colorectal cancer, surgical resection results in a cure for roughly 50% of patients. Burdette, *supra* at 125. Irradiation is used both preoperatively and postoperatively in treating colorectal cancer. *Id.* at 125, 132-33. Chemotherapeutic agents, particularly 5-fluorouracil, are also powerful weapons in treating colorectal cancer. *Id.* at 125, 133. Other agents include irinotecan and floxuridine, cisplatin, levamisole, methotrexate, interferon-alpha, and leucovorin. *Id.* at 133. Nonetheless, thirty to forty percent of patients will develop a recurrence of colon cancer following surgical resection. Wayne De Vos, *Follow-up After Treatment of Colon Cancer, Colon and Rectal Cancer* 225 (Peter S. Edelstein ed., 2000), which in many patients is the ultimate cause of death. Accordingly, colon cancer patients must be closely monitored to determine response to therapy and to detect persistent or recurrent disease and metastasis.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of colorectal cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that

WO 02/48370

PCT/US01/51341

-6-

would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop colorectal cancer, for diagnosing colorectal cancer, for monitoring the progression of the disease, for staging the colorectal cancer, for determining whether the colorectal cancer has metastasized, and for imaging the colorectal cancer. There is also a need for better treatment of colorectal cancer.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon; identify and monitor colon tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered colon tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to colon cells and/or colon tissue. These colon specific nucleic acids (CSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the CSNA is genomic DNA, then the CSNA is a colon specific gene (CSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 148 through 250. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 147. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a CSP, or that selectively hybridize or exhibit substantial sequence similarity to a CSNA, as well as allelic variants of a nucleic acid molecule encoding a CSP, and allelic variants of a CSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a CSP or that comprises a part of a nucleic acid sequence of a CSNA are also provided.

WO 02/48370

PCT/US01/51341

-7-

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a CSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a CSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a CSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a CSNA.

Another object of the invention is to provide methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a CSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutain), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a CSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention.

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring colon tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered colon tissue for treatment and research.

WO 02/48370

PCT/US01/51341

-8-

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon. The invention provides methods of using the polypeptides of the invention to identify and/or monitor colon tissue, and to produce engineered colon tissue.

5 The agonists and antagonists of the instant invention may be used to treat colon cancer and non-cancerous disease states in colon and to produce engineered colon tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for
10 comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by
15 those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used
20 in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and
25 Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology - 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory
30 Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

WO 02/48370

PCT/US01/51341

-9-

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 10 A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and
- 15 "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.
- 20 The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes
- 25 any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included
- 30 are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are

WO 02/48370

PCT/US01/51341

-10-

known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may
5 comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a
10 nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

15 A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an
20 RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a
25 portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized
30 polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site,

WO 02/48370

PCT/US01/51341

-11-

recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, e.g. for use as probes or primers, or may be double-stranded, e.g. for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

WO 02/48370

PCT/US01/51341

-12-

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroamidothioate, phosphoraniladate, phosphoramidate, and the like. See e.g., LaPlante *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) *Oligonucleotides and Analogues: A Practical Approach*, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each

WO 02/48370

PCT/US01/51341

-13-

nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

5 The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

10 The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number
15 of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, e.g., the programs FASTA2 and FASTA3, provides alignments and percent sequence identity
20 of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent
25 sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

30 A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its

WO 02/48370

PCT/US01/51341

-14-

complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology"

5 interchangeably. In this application, those terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optionally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its
10 complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

15 Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at
20 least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species,
25 length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of
30 hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is

WO 02/48370

PCT/US01/51341

-15-

performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence

5 hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where l is the length of the hybrid in base pairs.

10 The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50$$

15 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours).

30 Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten

WO 02/48370

PCT/US01/51341

-16-

hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify

5 nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents

10 to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook

15 (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in

20 the particular hybridization assay indicates detection of a specific hybridization

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon

25 degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (e.g., for oligonucleotide probes) may be calculated by the formula:

$$T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/N),$$

wherein N is change length and the $[\text{Na}^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p.

30 11.45. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using

WO 02/48370

PCT/US01/51341

-17-

mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. See, e.g., Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

5 The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 µg of plasmid or DNA fragment
10 is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are
15 specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the
20 art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, e.g., Sambrook (1989), *supra*.

25 Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon
30 portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-

WO 02/48370

PCT/US01/51341

-18-

derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one

5 exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic

10 acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing

15 Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that

20 nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the

25 nucleic acid molecule comprises the wild type nucleic acid sequence encoding a CSP or is a CSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high

30 rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

WO 02/48370

PCT/US01/51341

-19-

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. See, e.g., Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

5 The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

10 The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. See, e.g., Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

15 The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

25 The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, e.g., Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

30 The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein

WO 02/48370

PCT/US01/51341

-20-

small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, e.g., Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are

WO 02/48370

PCT/US01/51341

-21-

introduced (e.g., bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors
5 are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that
10 serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding
15 generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in
20 its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

25 As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-
30 occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide

WO 02/48370

PCT/US01/51341

-22-

each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a CSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide
 5 that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be
 10 "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single
 15 species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,
 20 followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared
 25 to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40
 30 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

WO 02/48370

PCT/US01/51341

-23-

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modifications include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antigens (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antigens which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. See Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a

WO 02/48370

PCT/US01/51341

-24-

different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs.

- 5 The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide
10 analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

- The term "non-peptide analog" refers to a compound with properties that are
15 analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm
20 polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: $-\text{CH}_2\text{NH}-$, $-\text{CH}_2\text{S}-$, $-\text{CH}_2-\text{CH}_2-$, $-\text{CH}=\text{CH}-$ (*cis* and *trans*), $-\text{COCH}_2-$, $-\text{CH}(\text{OH})\text{CH}_2-$, and $-\text{CH}_2\text{SO}-$, by methods well-known in the art. Systematic substitution of one or more amino acids of a
25 consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may
30 add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (e.g., a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure,

WO 02/48370

PCT/US01/51341

-26-

Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), *Immunology - A Synthesis* 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, β -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate, ϵ -N,N,N-trimethyllysine, ϵ -N-acetyllysine, O-phosphoserine, N-acetyls erine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, ϵ -N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence

WO 02/48370

PCT/US01/51341

-27-

similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. See, e.g., Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutant thereof. See, e.g., GCG Version 6.1. Other programs include FASTA, discussed *supra*.

WO 02/48370

PCT/US01/51341

-28-

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. See, e.g., Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein

5 incorporated by reference. Preferred parameters for blastp are:

	Expectation value:	10 (default)
	Filter:	seg (default)
	Cost to open a gap:	11 (default)
	Cost to extend a gap:	1 (default)
10	Max. alignments:	100 (default)
	Word size:	11 (default)
	No. of descriptions:	100 (default)
	Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at
 15 least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms
 20 other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (e.g., FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be
 25 determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding
 portion thereof that competes with the intact antibody for specific binding to a molecular
 30 species, e.g., a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv,

WO 02/48370

PCT/US01/51341

-29-

dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')₂ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature 341: 544-546 (1989).*

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science 242: 423-426 (1988); Huston et al., Proc. Natl. Acad. Sci. USA 85: 5879-5883 (1988).* Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA 90: 6444-6448 (1993); Poljak et al., Structure 2: 1121-1123 (1994).* One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For

WO 02/48370

PCT/US01/51341

-30-

instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (e.g., BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than 1 μ M, preferably less than 100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "colon specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the colon as compared to other tissues in the body. In a preferred embodiment, a "colon specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "colon specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more

WO 02/48370

PCT/US01/51341

-31-

preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the colon or to colon cells or tissue or that are derived from such nucleic acid molecules. These isolated colon specific nucleic acids (CSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon, a colon-specific polypeptide (CSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 148 through 250. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 147.

A CSNA may be derived from a human or from another animal. In a preferred embodiment, the CSNA is derived from a human or other mammal. In a more preferred embodiment, the CSNA is derived from a human or other primate. In an even more preferred embodiment, the CSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a CSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a CSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a CSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 148 through 250. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 147.

WO 02/48370

PCT/US01/51341

-32-

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 148 through 250. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 147. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a CSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human CSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 148 through 250. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a CSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 148 through 250, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a CSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a CSP.

WO 02/48370

PCT/US01/51341

-33-

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a CSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 147. In a preferred

5 embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a CSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 147, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a CSNA, more preferably at least 95%,

10 more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a CSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one

15 that exhibits sequence identity over its entire length to a CSNA or to a nucleic acid molecule encoding a CSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the CSNA or the nucleic acid molecule encoding a CSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at

20 least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 148 through 250 or demonstrates

25 significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 147. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the CSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated

30 species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived

WO 02/48370

PCT/US01/51341

-34-

from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In
5 another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a CSNA. Further, the substantially similar nucleic acid molecule may or may not be a CSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a CSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a
10 CSNA or a nucleic acid encoding a CSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population.
15 Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic
20 variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a CSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a CSNA. In a more preferred embodiment, the gene
25 is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 147. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic
30 acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a CSP. However, in a preferred embodiment, the part encodes a CSP. In one aspect, the invention comprises a part of a

WO 02/48370

PCT/US01/51341

-35-

CSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a CSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a CSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a CSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a protein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (e.g., reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or

WO 02/48370

PCT/US01/51341

-36-

fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as ^{32}P -dATP, ^{32}P -dCTP, ^{32}P -dGTP, ^{32}P -dTTP, ^{32}P -3'-dATP, ^{32}P -ATP, ^{32}P -CTP, ^{32}P -GTP, ^{32}P -UTP, ^{35}S -dATP, α - ^{35}S -GTP, α - ^{32}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy5-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc., Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Hensgariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, e.g., from phage promoters, such as T7, T3, and SP6, for

WO 02/48370

PCT/US01/51341

-37-

RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301-305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. See, e.g., Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside

WO 02/48370

PCT/US01/51341

-38-

- bonds. See Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties.
- Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. See Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.
- Modified oligonucleotide backbones include, without limitation,
- phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'.
- Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.
- Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholine linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and

WO 02/48370

PCT/US01/51341

-39-

methylenedihydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (see, e.g., "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average).

WO 02/48370

PCT/US01/51341

-40-

Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyanide backbone with nucleobase sidechains. See, e.g.,

5 Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure

10 throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24:

15 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed,

20 partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by

25 reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Prascuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

30

WO 02/48370

PCT/US01/51341

-41-

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a CSNA, such as deletions, insertions, translocations, and duplications of the CSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. See, e.g., Andreeff *et al.* (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, e.g., Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify CSNA in, and isolate CSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. See, e.g., Schwarzbacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing

WO 02/48370

PCT/US01/51341

-42-

nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to CSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the

5 nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in

10 their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a CSP. In a more

15 preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 148 through 250. In another preferred embodiment, the probe or primer is derived from a CSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 147.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even

20 more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. See, e.g., Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide

25 probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter*

30

WO 02/48370

PCT/US01/51341

-43-

- alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, e.g., in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.
- PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.
- The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (See, e.g., U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).
- Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. See, e.g., Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. See, e.g., Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).
- Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as

WO 02/48370

PCT/US01/51341

-44-

hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, e.g., a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, e.g., a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polycetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, e.g. on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

WO 02/48370

PCT/US01/51341

-45-

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacsa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

WO 02/48370

PCT/US01/51341

-46-

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the

5 nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, e.g., the numerous derivatives of phage lambda, e.g., NM989, λ GT10 and λ GT11, and other phages, e.g.,

10 M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: e.g., typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

15 In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components,

20 e.g. through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (e.g., YIp5) and Yeast Replicating plasmids (the YRp and YEplac series

25 plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIp1ac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most

30 common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

WO 02/48370

PCT/US01/51341

-47-

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSFTTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, e.g., in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (e.g., vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (e.g., bovine papillomavirus), and retroviral vectors (e.g., murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL941.

WO 02/48370

PCT/US01/51341

-48-

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, e.g., promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, e.g., sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, e.g., *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of *fd* coat protein, or the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

WO 02/48370

PCT/US01/51341

-49-

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast α -mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the

5 transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the
10 Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the CSNA of interest. Often, expression is enhanced by
15 incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BOH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker
20 gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (e.g., ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted
25 in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product
30 information from manufacturers of biological, chemical and immunological reagents also provide useful information.

WO 02/48370

PCT/US01/51341

-50-

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PltetO-1* promoter. The *PltetO-1* promoter takes advantage of the high expression levels from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity

WO 02/48370

PCT/US01/51341

-51-

resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the VS epitope, detectable by anti-VS antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, e.g., the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. See Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, e.g. the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, e.g., the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like

WO 02/48370

PCT/US01/51341

-52-

chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. See Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. See Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (see, e.g., Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (see, e.g., Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (see, e.g., Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (see, e.g., Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entirety. See also Conn (ed.), Green Fluorescent Protein (Methods in

WO 02/48370

PCT/US01/51341

-53-

Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

5 Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable
10 expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin
15 C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney
20 murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphiPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency;
25 varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art
30 may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must

WO 02/48370

PCT/US01/51341

-54-

be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide CSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquitination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database <http://www.abrf.org/ABRF/Research/Committees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological

WO 02/48370

PCT/US01/51341

-55-

sources" Cooper et al. *Nucleic Acids Res.* 29: 332-335 (2001) and <http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html> (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translational modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

WO 02/48370

PCT/US01/51341

-57-

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically.

- 5 Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule
- 10 encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired
- 15 post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid
- 20 sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

- In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its
- 25 controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture
- 30 requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

WO 02/48370

PCT/US01/51341

-58-

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (See, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces*

WO 02/48370

PCT/US01/51341

-59-

pombe, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical

5 mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562

10 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from colon are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human colon cells.

15 Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel

20 (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

25 Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or

30 electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, e.g., with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80

WO 02/48370

PCT/US01/51341

-60-

- (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g., Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent B. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) ([http://www.biorad.com/LifeScience/pdf/](http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf)
- 10 New_Gene_Pulser.pdf).
- Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in
- 15 the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.
- For lithium-mediated transformation, yeast cells are treated with lithium acetate,
- 20 which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).
- 25 For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991).
- 30 The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

WO 02/48370

PCT/US01/51341

-61-

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. See, e.g., Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. See, e.g., Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification: Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by

WO 02/48370

PCT/US01/51341

-62-

means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

5 Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a colon specific polypeptide (CSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ
10 ID NO: 148 through 250. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide,
15 wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a CSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 148 through 250. A polypeptide that comprises only a fragment of an entire CSP may or may not be a polypeptide that is also a CSP. For instance, a full-length polypeptide may be
20 colon-specific, while a fragment thereof may be found in other tissues as well as in colon. A polypeptide that is not a CSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-CSP antibodies. However, in a preferred embodiment, the part or fragment is a CSP. Methods of determining whether a polypeptide is a CSP are described *infra*.

25 Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by
30 native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

WO 02/48370

PCT/US01/51341

-63-

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. See, e.g., Lerner, *Nature* 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

10 Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is 20 at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by 25 truncating the nucleic acid molecule, e.g., a CSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. See, e.g., Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a 30 fragment of polypeptide of the invention, preferably a CSP, may be produced by

WO 02/48370

PCT/US01/51341

-64-

chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a CSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be colon-specific. In a preferred embodiment, the mutein is colon-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 148 through 250. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino

WO 02/48370

PCT/US01/51341

-65-

acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Mutains with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is colour-specific, as described below. Multiple random mutations can be introduced into the

5 gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing mutains with targeted or random amino acid alterations are well-known in the art. *See, e.g.*,

10 Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the

15 polypeptide is homologous to a CSP. In an even more preferred embodiment, the polypeptide is homologous to a CSP selected from the group having an amino acid sequence of SEQ ID NO: 148 through 250. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a CSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity

20 to an comprising an amino acid sequence of SEQ ID NO: 148 through 250. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In a yet more

25 preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%,

30 99.7%, 99.8% or 99.9% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

WO 02/48370

PCT/US01/51341

-66-

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a CSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a CSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the CSNA is selected from the group consisting of SEQ ID NO: 1 through 147. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a CSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the CSP is selected from the group consisting of SEQ ID NO: 148 through 250.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 148 through 250. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the CSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, e.g., dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a CSP. Further, the homologous protein may or may not encode polypeptide that is a CSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a CSP.

WO 02/48370

PCT/US01/51341

-67-

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but
 5 also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in
 10 eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a CSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino
 15 acid sequence selected from the group consisting of SEQ ID NO: 148 through 250. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 147.

In another embodiment, the invention provides polypeptides which comprise
 20 derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a CSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 148 through 250, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been
 25 acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antigens that can serve as specific binding pair members for a labeled ligand.

30 Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of

WO 02/48370

PCT/US01/51341

-68-

glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.),

5 Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Scifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of

10 ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide,

15 including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic

20 processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on

25 the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), e.g., offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546,

30 Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa

WO 02/48370

PCT/US01/51341

-69-

Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

10 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMD, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCEs, DFDNB, DMA, DMP, DMS, DFDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCEs, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMAP, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

25 The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

30 The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-CSP antibodies.

WO 02/48370

PCT/US01/51341

-70-

- The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with troyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.
- In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a CSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 148 through 250. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a CSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂-, --CH=CH--(cis and trans), --COCH₂-, --CH(OH)CH₂-- and --CH₂SO--.
- In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a CSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis; peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Kole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are

WO 02/48370

PCT/US01/51341

-71-

described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

5 Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular
10 Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and tBOC derivatives of dabeyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabeyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for
15 pairing with the dabeyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-Fmoc-L-glutamic acid or the corresponding tBOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated Fmoc synthesis of peptides using
20 (Fmoc)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and
25 phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, e.g., Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-
30 bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-

WO 02/48370

PCT/US01/51341

-72-

trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-
 5 2-aminobenzoic acid (anthranilic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-
 10 hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-
 15 methylbenzoic acid, Fmoc-3-amino-2-naphthoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-
 20 1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed
 25 mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*,
 30 *Science* 292(5516): 498-500 (2001).

WO 02/48370

PCT/US01/51341

-73-

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a CSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 148 through 250, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 147, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 147.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so

WO 02/48370

PCT/US01/51341

-74-

included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), *The Yeast Two-Hybrid System*, Oxford University Press (1997); Zhu *et al.*, *Yeast Hybrid Technologies*, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595; Fabbri *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95,

WO 02/48370

PCT/US01/51341

-75-

- 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.
- Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.
- The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.
- Fusion partners include, *inter alia*, myc, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin trpE, protein A, β -lactamase, α -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast λ mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin, and the constant domain of IgG. See, e.g., Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (e.g., a Merrifield synthesis), or produced by chemical cross-linking.
- Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the CSP.

WO 02/48370

PCT/US01/51341

-76-

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize CSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to

5 assay for the polypeptides of the present invention, particularly CSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of CSPs, as for example by immunoprecipitation, and for use

10 as specific agonists or antagonists of CSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine

15 scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are

20 available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZ104KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of

25 one having ordinary skill in the art. See, *e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of

30 absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

WO 02/48370

PCT/US01/51341

-77-

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so

WO 02/48370

PCT/US01/51341

-78-

attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

5 In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies
10 are specific for a polypeptide that is a CSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 148 through 250, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes,
15 discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a CSP may be glycosylated in cancerous cells,
20 but not glycosylated in normal cells or *visa versa*. In addition, alternative splice forms of a CSP may be indicative of cancer. Differential degradation of the C or N-terminus of a CSP may also be a marker or target for anticancer therapy. For example, a CSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

25 As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-CSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-
30 fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine

WO 02/48370

PCT/US01/51341

-79-

the presence of the protein of the present invention in samples derived from human colon.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but

WO 02/48370

PCT/US01/51341

-80-

also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in

WO 02/48370

PCT/US01/51341

-81-

immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues

5 (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard et al. (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody

10 Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal

15 antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two

20 necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

25 Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies

30 of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein

WO 02/48370

PCT/US01/51341

-82-

- (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.,* Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al., Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al., Immunotechnology*, 4(1): 1-20 (1998); Rader *et al., Current Opinion in Biotechnology* 8: 503-508 (1997);
- 5 Aujame *et al., Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.,* 17: 453-455 (1996); Barbas *et al., Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al., Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas
- 10 (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.
- Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length
- 15 antibody in a further prokaryotic or a eukaryotic host cell.
- Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.
- For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al., Biosci.*
- 20 *Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al., J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al., Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al., Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al., J. Immunol. Methods*. 201(1): 67-75 (1997);, Frenken *et al., Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al., Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein
- 25 by reference in their entireties.
- Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al., Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al., Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al., Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al., Immunology* 91(1): 13-9 (1997);
- 30 and Nesbit *et al., J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

WO 02/48370

PCT/US01/51341

-83-

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, 5 *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present 10 invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Linonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody 15 fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

20 Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

25 The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and 30 protein fragments encoded by the isolated nucleic acids of the present invention.

WO 02/48370

PCT/US01/51341

-84-

Among such useful fragments are Fab, Fab', Fv, F(ab')₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that
5 bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or
10 one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human
15 mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.,* United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21):
20 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region,
25 Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody
30 complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

WO 02/48370

PCT/US01/51341

-85-

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BlueGal; iodonitrotriazolium

(INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are
5 luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic
10 compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such
15 enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, e.g., for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

20 There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein
25 (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy transfer fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc.,
30 Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY

WO 02/48370

PCT/US01/51341

-87-

630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, e.g., for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H , and ^{125}I .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{89\text{m}}\text{Tc}$, ^{87}Ru , $^{90\text{Y}}$, ^{90}Sr , $^{88\text{Y}}$, ^{72}Se , ^{67}Cu , or ^{47}Sc .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. See Hall (ed.), *Immunotoxin Methods and Protocols* (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), *Clinical Applications of Immunotoxins*, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention,

WO 02/48370

PCT/US01/51341

-88-

to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

WO 02/48370

PCT/US01/51341

-89-

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a CSP. In a preferred embodiment, the CSP comprises an amino acid sequence selected from SEQ ID NO: 148 through 250, or a fragment, mutant, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a CSNA of the invention, preferably a CSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 147, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human CSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g., Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, *Mouse Genetics and Transgenics: A Practical Approach*, Oxford University Press (2000); and Pinkert, *Transgenic Animal Technology: A Laboratory Handbook*, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g., Paterson et al., Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al., Biotechnology* 11: 1263-1270 (1993); Wright *et al., Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989) retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g., Van der Putten et al., Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g., Thompson et al., Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g., Lo, 1983, Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g., Ulmer et al., Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem

WO 02/48370

PCT/US01/51341

-90-

cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (see, e.g., Laviterno *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (see, e.g., Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (i.e., a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, e.g., the teaching of Lasko *et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines, inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to

both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

5 Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

10 Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous
15 gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. See, e.g., Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. See, e.g., Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable
25 marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications
30 to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. See, e.g., Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are

directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e.g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e.g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e.g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmid, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e.g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e.g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. See, e.g., U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

WO 02/48370

PCT/US01/51341

-93-

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 147 and SEQ ID NO: 148 through 250 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data

WO 02/48370

PCT/US01/51341

-94-

set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences
5 comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or
10 data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and
15 similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a
20 nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence
25 comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the
30 steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one

WO 02/48370

PCT/US01/51341

-95-

overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Colon Cancer

5 The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a CSNA or a CSP in a human patient that has or may have colon cancer, or who is at risk of developing colon cancer, with the expression of a CSNA or a CSP in a normal human control. For purposes of the present invention,
 10 "expression of a CSNA" or "CSNA expression" means the quantity of CSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a CSP" or "CSP expression" means the amount of CSP that can be measured by any method known in the art or the level of translation of a CSG
 15 CSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing colon cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of CSNA or CSP in cells, tissues, organs or bodily fluids compared with levels of CSNA or CSP in
 20 control, wherein an increase, or decrease in certain cases, in levels of a CSNA or CSP in the patient versus the normal human control is associated with the presence of colon cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in the structure of the mRNA of a CSG compared to the mRNA from a normal
 25 control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in a CSP compared to a CSP from a normal control. These changes include, e.g., alterations in glycosylation and/or phosphorylation of the CSP or
 30 subcellular CSP localization.

In a preferred embodiment, the expression of a CSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from

WO 02/48370

PCT/US01/51341

-96-

SEQ ID NO: 148 through 250, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the CSNA expression that is measured is the level of expression of a CSNA mRNA selected from SEQ ID NO: 1 through 147, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. CSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. CSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a CSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, CSNA expression may be compared to a known control, such as normal colon nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a CSP is measured by determining the level of a CSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 148 through 250, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of CSNA or CSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of colon cancer. The expression level of a CSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the CSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.*, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the CSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or

phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a CSP is prepared if one is not already available. In a preferred
5 embodiment, the antibody is a monoclonal antibody. The anti-CSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the CSP will bind to the anti-CSP antibody. The sample is removed, the solid support is washed to remove unbound
10 material, and an anti-CSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the CSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the
15 amount of a CSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure CSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-CSP antibody is attached to a solid
20 support and an allocated amount of a labeled CSP and a sample of interest are incubated with the solid support. The amount of labeled CSP detected which is attached to the solid support can be correlated to the quantity of a CSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential
25 separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly
30 square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

WO 02/48370

PCT/US01/51341

-98-

Expression levels of a CSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example,

5 reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

10 Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more CSNAs of interest. In this approach, all or a portion of one or more CSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the

15 RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid

20 molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any

25 other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of CSNA or CSP includes, without limitation, colon tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, colon cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment,

30 especially when metastasis of a primary colon cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without

WO 02/48370

PCT/US01/51341

-99-

limitation, needle biopsy, e.g., transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. See Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in

5 CSNAs or CSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a CSNA or CSP. In many cases, the use of another cancer marker
10 will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other CSNA or CSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular CSNA or CSP is measured. In a more
15 preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression
20 levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having colon cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP and then ascertaining whether the patient has colon cancer from the expression level of the CSNA or CSP. In general, if high expression
25 relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is
30 indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the

WO 02/48370

PCT/US01/51341

-100-

same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether colon cancer has metastasized in a patient. One may identify whether the colon cancer has metastasized by measuring the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a variety of tissues. The presence of a CSNA or CSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (e.g., the same tissue from another individual) is indicative of metastasis if high level expression of a CSNA or CSP is associated with colon cancer. Similarly, the presence of a CSNA or CSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a CSNA or CSP is associated with colon cancer. Further, the presence of a structurally altered CSNA or CSP that is associated with colon cancer is also indicative of metastasis.

In general, if high expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The CSNA or CSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with colon cancers or other colon related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of colon disorders.

30 *Staging*

The invention also provides a method of staging colon cancer in a human patient. The method comprises identifying a human patient having colon cancer and analyzing

WO 02/48370

PCT/US01/51341

-101-

cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more CSNAs or CSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more CSNAs or CSPs is determined for each stage to obtain a standard expression level for each CSNA and CSP. Then, the CSNA or CSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The CSNA or CSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the CSNAs and CSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a CSNA or CSP to determine the stage of a colon cancer.

Monitoring

Further provided is a method of monitoring colon cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, e.g., chemotherapy, radiotherapy or surgery, has decreased or eliminated the colon cancer. The method comprises identifying a human patient that one wants to monitor for colon cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more CSNAs or CSPs, and comparing the CSNA or CSP levels over time to those CSNA or CSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a CSNA or CSP that are associated with colon cancer.

If increased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a

WO 02/48370

PCT/US01/51341

-102-

cancerous lesion, then detecting an decrease in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of CSNAs or CSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples.

- 5 Monitoring a patient for onset of colon cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

- The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a CSNA and/or CSP. The present invention
- 10 provides a method in which a test sample is obtained from a human patient and one or more CSNAs and/or CSPs are detected. The presence of higher (or lower) CSNA or CSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly colon cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more CSNAs and/or
- 15 CSPs of the invention can also be monitored by analyzing levels of expression of the CSNAs and/or CSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

20 *Detection of Genetic Lesions or Mutations*

- The methods of the present invention can also be used to detect genetic lesions or mutations in a CSG, thereby determining if a human with the genetic lesion is susceptible to developing colon cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing colon cancer. Genetic lesions can be detected,
- 25 for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the CSGs of this invention, a chromosomal rearrangement of CSG, an aberrant modification of CSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a CSG. Methods to detect such lesions in the CSG of this invention are known to those having ordinary skill in the art following the teachings
- 30 of the specification.

WO 02/48370

PCT/US01/51341

-103-

Methods of Detecting Noncancerous Colon Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having or known to have a noncancerous colon disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP, comparing the expression level or structural alteration of the CSNA or CSP to a normal colon control, and then ascertaining whether the patient has a noncancerous colon disease. In general, if high expression relative to a control of a CSNA or CSP is indicative of a particular noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of a noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a CSNA and/or CSP is associated with a particular noncancerous colon disease by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining which CSNAs and/or CSPs are expressed in the tissue at either a higher or a lower level than in normal colon tissue. In another embodiment, one may determine whether a CSNA or CSP exhibits structural alterations in a particular noncancerous colon disease state by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining the structural alterations in one or more CSNAs and/or CSPs relative to normal colon tissue.

30

Methods for Identifying Colon Tissue

In another aspect, the invention provides methods for identifying colon tissue.

These methods are particularly useful in, e.g., forensic science, colon cell differentiation
5 and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is colon tissue or has colon tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising colon tissue or having colon tissue-like characteristics, determining whether the sample expresses one or more CSNAs
10 and/or CSPs, and, if the sample expresses one or more CSNAs and/or CSPs, concluding that the sample comprises colon tissue. In a preferred embodiment, the CSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 148 through 250, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 147, or a
15 hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a CSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a CSP is expressed. Determining whether a sample
20 expresses a CSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the CSP has an amino acid sequence selected from SEQ ID NO: 148 through 250, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two CSNAs and/or CSPs is determined. In a more preferred embodiment, the
25 expression of at least three, more preferably four and even more preferably five CSNAs and/or CSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is colon tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are
30 recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into colon tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, e.g., in producing new colon tissue by tissue engineering. These agents include,

WO 02/48370

PCT/US01/51341

-105-

e.g., growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

5 Methods for Producing and Modifying Colon Tissue

In another aspect, the invention provides methods for producing engineered colon tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a CSNA or a CSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of colon tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal colon tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered colon tissue or cells comprises one of these cell types. In another embodiment, the engineered colon tissue or cells comprises more than one colon cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the colon cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more CSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode CSPs having amino acid sequences selected from SEQ ID NO: 143 through 250, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 147, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a CSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial colon tissue may be used to treat patients who have lost some or all of their colon function.

30 Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives,

WO 02/48370

PCT/US01/51341

-106-

antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a CSNA or part thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 147, a nucleic acid that hybridizes

5 thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a CSP or fragment thereof. In a more preferred embodiment, the CSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 148 through 250, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of

10 the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-CSP antibody, preferably an antibody that specifically binds to a CSP having an amino acid that is selected from the group consisting of SEQ ID NO: 148 through 250, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the

15 polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in

20 Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties,

25 and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation,

30 topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transnasal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

WO 02/48370

PCT/US01/51341

-107-

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or

WO 02/48370

PCT/US01/51341

-108-

suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

5 Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

15 For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

25 Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

WO 02/48370

PCT/US01/51341

-109-

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, e.g., 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

WO 02/48370

PCT/US01/51341

-110-

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for

5 treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

10

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

15

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example CSP polypeptide, fusion protein, or fragments thereof, antibodies specific for CSP, agonists, antagonists or inhibitors of CSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof, as would be understood in the medical arts, cure, although desired, is not required.

20

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

25

For example, the ED₅₀ (the dose therapeutically effective in 50% of the population) and LD₅₀ (the dose lethal to 50% of the population) can be determined in one or more cell culture or animal model systems. The dose ratio of toxic to therapeutic

30

WO 02/48370

PCT/US01/51341

-111-

effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After
5 administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

10 The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting
15 pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody
20 agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (e.g., 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will
25 employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to
30 the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

WO 02/48370

PCT/US01/51341

-112-

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of colon function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

10 *Gene Therapy and Vaccines*

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a CSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a CSP are administered, for example, to complement a deficiency in the native CSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid

WO 02/48370

PCT/US01/51341

-113-

molecule encodes a CSP having the amino acid sequence of SEQ ID NO: 148 through 250, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a CSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allogenic rejection, and are administered to complement defects in CSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a CSP having the amino acid sequence of SEQ ID NO: 148 through 250, or a fragment, fusion protein, allelic variant or homolog thereof.

10 Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a CSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a CSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a CSG. For example, oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to CSG transcripts, are also useful in therapy. See, e.g., Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the CSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. See, e.g., Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

WO 02/48370

PCT/US01/51341

-114-

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a CSP, preferably a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule
5 having a nucleotide sequence of SEQ ID NO: 1 through 147, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a CSP, a
10 fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant CSP defect.

Protein compositions are administered, for example, to complement a deficiency in native CSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to CSP. The immune response can
15 be used to modulate activity of CSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate CSP.

In a preferred embodiment, the polypeptide is a CSP comprising an amino acid
20 sequence of SEQ ID NO: 148 through 250, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 147, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is
25 administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of CSP, or to target therapeutic agents to sites of CSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a

WO 02/48370

PCT/US01/51341

-115-

CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 147, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a CSP or have a modulatory effect on the expression or activity of a CSP. Modulators which decrease the expression or activity of CSP (antagonists) are believed to be useful in treating colon cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a CSP can also be designed, synthesized and tested for use in the imaging and treatment of colon cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the CSPs identified herein. Molecules identified in the library as being capable of binding to a CSP are key candidates for further evaluation for use in the treatment of colon cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a CSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of CSP is administered. Antagonists of CSP can be produced using methods generally known in the art. In particular, purified CSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a CSP.

In other embodiments a pharmaceutical composition comprising an agonist of a CSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP encoded by a

WO 02/48370

PCT/US01/51341

-116-

nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 147, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Colon Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the colon or to specific cells in the colon. In a preferred embodiment, an anti-CSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if colon tissue needs to be selectively destroyed. This would be useful for targeting and killing colon cancer cells.

In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting colon cell function.

In another embodiment, an anti-CSP antibody may be linked to an imaging agent that can be detected using, e.g., magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring colon function, identifying colon cancer tumors, and identifying noncancerous colon diseases.

EXAMPLES

Example 1: Gene Expression analysis

CSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer"

WO 02/48370

PCT/US01/51341

-117-

disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be

5 considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the CSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), highest differential expression for a given cancer (CLASP 4);
 10 differential expression in cancer tissue (CLASP 5), and. cDNA libraries were divided into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with
 15 rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are
 20 incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were
 25 analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

(a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a
 30 gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.

WO 02/48370

PCT/US01/51341

-118-

- (b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.
- (c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

The CLASP™ scores for SEQ ID NO: 1-147 are listed below:

SEQ ID NO: 1	DEX0253_1	CLASP2
SEQ ID NO: 2	DEX0253_2	CLASP2
15 SEQ ID NO: 3	DEX0253_3	CLASP2
SEQ ID NO: 4	DEX0253_4	CLASP2 CLASP1
SEQ ID NO: 5	DEX0253_5	CLASP2 CLASP1
SEQ ID NO: 6	DEX0253_6	CLASP5 CLASP1
SEQ ID NO: 7	DEX0253_7	CLASP5 CLASP1
20 SEQ ID NO: 8	DEX0253_8	CLASP2
SEQ ID NO: 9	DEX0253_9	CLASP2
SEQ ID NO: 10	DEX0253_10	CLASP2
SEQ ID NO: 11	DEX0253_11	CLASP2 CLASP1
SEQ ID NO: 12	DEX0253_12	CLASP2
25 SEQ ID NO: 13	DEX0253_13	CLASP2
SEQ ID NO: 14	DEX0253_14	CLASP5 CLASP1
SEQ ID NO: 15	DEX0253_15	CLASP2
SEQ ID NO: 16	DEX0253_16	CLASP2
SEQ ID NO: 17	DEX0253_17	CLASP2
30 SEQ ID NO: 18	DEX0253_18	CLASP2
SEQ ID NO: 19	DEX0253_19	CLASP2
SEQ ID NO: 20	DEX0253_20	CLASP2
SEQ ID NO: 21	DEX0253_21	CLASP2
SEQ ID NO: 22	DEX0253_22	CLASP2
35 SEQ ID NO: 23	DEX0253_23	CLASP2
SEQ ID NO: 24	DEX0253_24	CLASP2 CLASP1
SEQ ID NO: 25	DEX0253_25	CLASP2 CLASP1
SEQ ID NO: 26	DEX0253_26	CLASP2 CLASP1
SEQ ID NO: 27	DEX0253_27	CLASP2 CLASP1
40 SEQ ID NO: 28	DEX0253_28	CLASP2
SEQ ID NO: 29	DEX0253_29	CLASP2
SEQ ID NO: 30	DEX0253_30	CLASP2

WO 02/48370

PCT/US01/51341

-119-

	SEQ ID NO: 31	DEX0253_31 CLASP2
	SEQ ID NO: 32	DEX0253_32 CLASP2
	SEQ ID NO: 33	DEX0253_33 CLASP2
	SEQ ID NO: 34	DEX0253_34 CLASP2
5	SEQ ID NO: 35	DEX0253_35 CLASP2
	SEQ ID NO: 36	DEX0253_36 CLASP2
	SEQ ID NO: 37	DEX0253_37 CLASP2
	SEQ ID NO: 38	DEX0253_38 CLASP2
	SEQ ID NO: 39	DEX0253_39 CLASP2
10	SEQ ID NO: 40	DEX0253_40 CLASP2
	SEQ ID NO: 41	DEX0253_41 CLASP2
	SEQ ID NO: 42	DEX0253_42 CLASP2
	SEQ ID NO: 43	DEX0253_43 CLASP2
	SEQ ID NO: 44	DEX0253_44 CLASP2
15	SEQ ID NO: 45	DEX0253_45 CLASP2
	SEQ ID NO: 46	DEX0253_46 CLASP2
	SEQ ID NO: 47	DEX0253_47 CLASP2
	SEQ ID NO: 48	DEX0253_48 CLASP2 CLASP1
	SEQ ID NO: 50	DEX0253_50 CLASP2 CLASP1
20	SEQ ID NO: 51	DEX0253_51 CLASP2 CLASP1
	SEQ ID NO: 52	DEX0253_52 CLASP2 CLASP1
	SEQ ID NO: 53	DEX0253_53 CLASP2
	SEQ ID NO: 54	DEX0253_54 CLASP2 CLASP1
	SEQ ID NO: 55	DEX0253_55 CLASP2 CLASP1
25	SEQ ID NO: 56	DEX0253_56 CLASP2
	SEQ ID NO: 57	DEX0253_57 CLASP2
	SEQ ID NO: 58	DEX0253_58 CLASP2
	SEQ ID NO: 59	DEX0253_59 CLASP2
	SEQ ID NO: 60	DEX0253_60 CLASP2
30	SEQ ID NO: 61	DEX0253_61 CLASP2
	SEQ ID NO: 62	DEX0253_62 CLASP5 CLASP1
	SEQ ID NO: 63	DEX0253_63 CLASP5 CLASP1
	SEQ ID NO: 64	DEX0253_64 CLASP5 CLASP1
	SEQ ID NO: 65	DEX0253_65 CLASP5 CLASP1
35	SEQ ID NO: 66	DEX0253_66 CLASP2
	SEQ ID NO: 68	DEX0253_68 CLASP2
	SEQ ID NO: 69	DEX0253_69 CLASP2 CLASP1
	SEQ ID NO: 70	DEX0253_70 CLASP5 CLASP1
	SEQ ID NO: 71	DEX0253_71 CLASP5 CLASP1
40	SEQ ID NO: 72	DEX0253_72 CLASP2
	SEQ ID NO: 73	DEX0253_73 CLASP2
	SEQ ID NO: 74	DEX0253_74 CLASP2
	SEQ ID NO: 75	DEX0253_75 CLASP2
	SEQ ID NO: 76	DEX0253_76 CLASP2 CLASP1 CLASP4
45	SEQ ID NO: 77	DEX0253_77 CLASP2
	SEQ ID NO: 78	DEX0253_78 CLASP5 CLASP1
	SEQ ID NO: 79	DEX0253_79 CLASP5 CLASP1
	SEQ ID NO: 80	DEX0253_80 CLASP5 CLASP1

WO 02/48370

PCT/US01/51341

-120-

	SEQ ID NO: 81	DEX0253_81	CLASP5 CLASP1
	SEQ ID NO: 82	DEX0253_82	CLASP2
	SEQ ID NO: 83	DEX0253_83	CLASP2
	SEQ ID NO: 84	DEX0253_84	CLASP1
5	SEQ ID NO: 85	DEX0253_85	CLASP1
	SEQ ID NO: 86	DEX0253_86	CLASP2
	SEQ ID NO: 87	DEX0253_87	CLASP2
	SEQ ID NO: 88	DEX0253_88	CLASP2
	SEQ ID NO: 89	DEX0253_89	CLASP2
10	SEQ ID NO: 90	DEX0253_90	CLASP2
	SEQ ID NO: 91	DEX0253_91	CLASP2
	SEQ ID NO: 92	DEX0253_92	CLASP2
	SEQ ID NO: 93	DEX0253_93	CLASP2
	SEQ ID NO: 94	DEX0253_94	CLASP2
15	SEQ ID NO: 95	DEX0253_95	CLASP2 CLASP1
	SEQ ID NO: 96	DEX0253_96	CLASP2 CLASP1
	SEQ ID NO: 97	DEX0253_97	CLASP2
	SEQ ID NO: 98	DEX0253_98	CLASP2
	SEQ ID NO: 99	DEX0253_99	CLASP2
20	SEQ ID NO: 100	DEX0253_100	CLASP2
	SEQ ID NO: 101	DEX0253_101	CLASP2
	SEQ ID NO: 102	DEX0253_102	CLASP2
	SEQ ID NO: 103	DEX0253_103	CLASP2
	SEQ ID NO: 104	DEX0253_104	CLASP5 CLASP1
25	SEQ ID NO: 105	DEX0253_105	CLASP2
	SEQ ID NO: 106	DEX0253_106	CLASP2
	SEQ ID NO: 107	DEX0253_107	CLASP2
	SEQ ID NO: 108	DEX0253_108	CLASP2
	SEQ ID NO: 109	DEX0253_109	CLASP2 CLASP1
30	SEQ ID NO: 110	DEX0253_110	CLASP2
	SEQ ID NO: 111	DEX0253_111	CLASP2
	SEQ ID NO: 112	DEX0253_112	CLASP2
	SEQ ID NO: 113	DEX0253_113	CLASP2
	SEQ ID NO: 114	DEX0253_114	CLASP2
35	SEQ ID NO: 115	DEX0253_115	CLASP5 CLASP1
	SEQ ID NO: 116	DEX0253_116	CLASP2
	SEQ ID NO: 117	DEX0253_117	CLASP2
	SEQ ID NO: 118	DEX0253_118	CLASP2
	SEQ ID NO: 119	DEX0253_119	CLASP2
40	SEQ ID NO: 120	DEX0253_120	CLASP2
	SEQ ID NO: 121	DEX0253_121	CLASP2
	SEQ ID NO: 122	DEX0253_122	CLASP2
	SEQ ID NO: 123	DEX0253_123	CLASP2
	SEQ ID NO: 124	DEX0253_124	CLASP2
	SEQ ID NO: 125	DEX0253_125	CLASP5 CLASP1
45	SEQ ID NO: 126	DEX0253_126	CLASP5 CLASP1
	SEQ ID NO: 127	DEX0253_127	CLASP2
	SEQ ID NO: 128	DEX0253_128	CLASP2
	SEQ ID NO: 129	DEX0253_129	CLASP2

WO 02/48370

PCT/US01/51341

-121-

	SEQ ID NO: 130	DEX0253_130	CLASP2
	SEQ ID NO: 131	DEX0253_131	CLASP2
	SEQ ID NO: 132	DEX0253_132	CLASP2
	SEQ ID NO: 133	DEX0253_133	CLASP5 CLASP1
5	SEQ ID NO: 134	DEX0253_134	CLASP2
	SEQ ID NO: 135	DEX0253_135	CLASP2
	SEQ ID NO: 136	DEX0253_136	CLASP2
	SEQ ID NO: 139	DEX0253_139	CLASP2
	SEQ ID NO: 140	DEX0253_140	CLASP5
10	SEQ ID NO: 141	DEX0253_141	CLASP2
	SEQ ID NO: 142	DEX0253_142	CLASP2
	SEQ ID NO: 143	DEX0253_143	CLASP2
	SEQ ID NO: 144	DEX0253_144	CLASP2
	SEQ ID NO: 145	DEX0253_145	CLASP2
15	SEQ ID NO: 146	DEX0253_146	CLASP2
	SEQ ID NO: 147	DEX0253_147	CLASP2

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The

WO 02/48370

PCT/US01/51341

-122-

results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the CSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the CSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the CSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 147 being a diagnostic marker for cancer.

DEX0095_71

25 **cln150-sqcln033**

Sequence

Sequence ID #

Dex0095_71 (cln150-sqcln033)

DEX0253_104(SEQ ID No.:104)

Semi quantitative PCR was done using the following primers:

30

Primer	DexSeqID	From	To	Primer Length
Sqcln033F	DEX0253_104	36	56	21
Sqcln033R	DEX0253_104	256	236	21

WO 02/48370

PCT/US01/51341

-123-

The relative levels of expression of sqcln0033 in 12 normal samples from 12 different tissues were determined. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Normal
Breast	0
Colon	1000
Endometrium	0
Kidney	0
Liver	0
Lung	0
Ovary	0
Prostate	0
Small Intestine	1000
Stomach	0
Testis	10
Uterus	0

10 Relative levels of expression in the table above show that none of the normal tissues examined here except colon and small intestine exhibit appreciable levels of sqcln033 gene expression.

The relative levels of expression of sqcln033 in 12 cancer samples from 12 different tissues were determined. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Cancer
Bladder	1
Breast	0
Colon	1000
Kidney	1
Liver	0
Lung	0
Ovary	0
Pancreas	0

WO 02/48370

PCT/US01/51341

-124-

prostate	1
Stomach	0
Testes	1
Uterus	0

Relative levels of expression in Table 2 show that sqcln033 is expressed in very low levels in bladder, kidney, prostate and testicular cancer. High levels of expression were observed for colon cancer.

- 5 The relative levels of expression of sqcln033 in 6 matching pair colon samples were determined. Each matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

- 10 Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Sample ID	Tissue	Cancer	NAT
9609E019	Colon	10	100
9709C074RA	Colon	1	10
9705F002D	Colon	10	100
9608B012	Colon	0	100
4004709A1	Colon	10	10
9707C004GB	Colon	100	10

- 15 Relative levels of expression in the table above show that sqcln033 is expressed in low levels in four and in moderate levels in one of the six colon cancer samples. Moderate levels of expression was observed in matching normal adjacent tissue (NAT).

From the semi-quantitative PCR experiments exhibited above especially from the results in the tables above high degree of specificity of sqcln033 for colon tissue is

- 20 observed.

Experiments are underway to design and test primers and probe for quantitative PCR experiments.

Example 3: Protein Expression

- 25 The CSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the CSNA is subcloned in pET-21d for expression in *E. coli*. In

WO 02/48370

PCT/US01/51341

-125-

addition to the CSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of CSNA, and six histidines, flanking the COOH-terminus of the coding sequence of CSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

5 An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of CSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickel chelating resin. The column was packed and washed with five column volumes of wash buffer. CSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

15 Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. See, e.g., WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

30 In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such

WO 02/48370

PCT/US01/51341

-126-

cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin. The

5 splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*,

10 *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes

15 use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability

20 to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, *CABIOS*, 4(1), 181-186, 1988, the contents of which are incorporated by

25 reference).

30	DEX0253_151	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	68-108	1.02	41
30	DEX0253_154	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	13-31	1.00	19
35	DEX0253_155	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	8-22	1.25	15
	67-98	1.06	32

WO 02/48370

PCT/US01/51341

-127-

	DEX0253_156	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	14-33	1.13 20
5	DEX0253_161	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	32-46	0.98 15
	DEX0253_170	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	6-48	1.02 43
10	DEX0253_172	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	19-28	1.00 10
	DEX0253_174	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
15	13-29	1.01 17
	DEX0253_178	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	2-17	0.99 16
	DEX0253_179	Antigenicity Index(Jameson-Wolf)
20	positions	AI avg length
	7-20	0.96 14
	DEX0253_184	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	18-33	1.00 16
25	DEX0253_187	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	42-54	1.09 13
	DEX0253_188	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
30	14-41	1.03 28
	81-97	0.96 17
	DEX0253_192	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	25-75	0.97 51
35	DEX0253_197	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	31-58	1.06 28
	DEX0253_198	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
40	73-82	1.22 10
	154-165	1.21 12
	213-224	1.06 12
	38-58	1.05 21
	314-323	1.03 10
45	250-267	0.99 18
	275-295	0.98 21
	137-149	0.96 13
	DEX0253_203	Antigenicity Index(Jameson-Wolf)

WO 02/48370

PCT/US01/51341

-128-

	positions	AI	avg length
	3-28	0.99	26
	DEX0253_204	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
5	29-39	1.10	11
	52-65	0.97	14
	DEX0253_208	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	3-14	1.04	12
10	DEX0253_211	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	33-54	1.22	22
	59-76	1.08	18
	DEX0253_212	Antigenicity Index(Jameson-Wolf)	
15	positions	AI	avg length
	22-31	1.00	10
	DEX0253_214	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	7-22	1.24	16
20	DEX0253_217	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	34-46	1.06	13
	DEX0253_218	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
25	23-32	1.16	10
	DEX0253_219	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	19-30	1.24	12
	DEX0253_221	Antigenicity Index(Jameson-Wolf)	
30	positions	AI	avg length
	58-69	1.08	12
	42-51	1.04	10
	DEX0253_222	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
35	58-69	1.08	12
	42-51	1.04	10
	DEX0253_226	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	8-18	1.01	11
40	DEX0253_229	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	8-36	1.02	29
	DEX0253_237	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
45	14-31	0.91	18
	DEX0253_245	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	3-27	1.12	25

WO 02/48370

PCT/US01/51341

-129-

DEX0253_246	Antigenicity Index(Jameson-Wolf)
positions	AI avg length
23-32	0.93 10

5 Examples of post-translational modifications (PTMs) of the BSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the LSPs of the invention (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html)
 10 of the invention (most recently accessed October 23, 2001). For full definitions of the PTMs see <http://www.expasy.org/cgi-bin/prosite-list.pl> most recently accessed October 23, 2001.

DEX0253_148	Tyr_Phospho_Site 3-9;
DEX0253_150	Ck2_Phospho_Site 19-22;23-26;
15 DEX0253_151	Ck2_Phospho_Site 83-86;128-131; Myristyl 28-33;49-54; Pkc_Phospho_Site 77-79;83-85;105-107;139-141;
DEX0253_152	Leucine_Zipper 6-27; Pkc_Phospho_Site 16-18;
DEX0253_153	Pkc_Phospho_Site 16-18;24-26;
DEX0253_154	Ck2_Phospho_Site 2-5;23-26;
20 DEX0253_155	Amidation 12-15; Myristyl 12-17;31-36;34-39; Pkc_Phospho_Site 75-77;
DEX0253_156	Ck2_Phospho_Site 8-11; Pkc_Phospho_Site 16-18;
DEX0253_158	Pkc_Phospho_Site 2-4;
DEX0253_159	Myristyl 26-31; Pkc_Phospho_Site 9-11;14-16;
25 DEX0253_160	Myristyl 13-18; Pkc_Phospho_Site 31-33;
DEX0253_161	Pkc_Phospho_Site 8-10;17-19;
DEX0253_162	Ck2_Phospho_Site 28-31; Myristyl 48-53; Pkc_Phospho_Site 28-30;
DEX0253_164	Asn_Glycosylation 21-24; Myristyl 17-22;
30 DEX0253_165	Pkc_Phospho_Site 44-46;
DEX0253_166	Asn_Glycosylation 7-10; Camp_Phospho_Site 3-6; Ck2_Phospho_Site 19-22; Myristyl 32-37; Pkc_Phospho_Site 6-8;45-47; Tyr_Phospho_Site 28-34;
DEX0253_167	Asn_Glycosylation 14-17; Myristyl 23-28; Pkc_Phospho_Site 9-11;
35 DEX0253_168	Asn_Glycosylation 38-41; Ck2_Phospho_Site 40-43; Myristyl 49-54; Pkc_Phospho_Site 16-18;35-37;
DEX0253_170	Camap_Phospho_Site 43-46; Myristyl 36-41;
DEX0253_171	Asn_Glycosylation 37-40; Ck2_Phospho_Site 79-82;86-89;
40 DEX0253_172	Myristyl 22-27;34-39;47-52; Pkc_Phospho_Site 7-9;15-17; Asn_Glycosylation 7-10;11-14;21-24; Pkc_Phospho_Site 52-54;55-57;
DEX0253_173	Ck2_Phospho_Site 27-30;
DEX0253_174	Myristyl 56-61;

WO 02/48370

PCT/US01/51341

-130-

	DEX0253_175	Myristyl 24-29;
	DEX0253_176	Pkc_Phospho_Site 8-10;
	DEX0253_177	Pkc_Phospho_Site 9-11;
5	DEX0253_178	Ck2_Phospho_Site 52-55; Myristyl 24-29; Pkc_Phospho_Site 45-47;
	DEX0253_179	Asn_Glycosylation 63-66; Ck2_Phospho_Site 25-28;
		Pkc_Phospho_Site 27-29;
	DEX0253_180	Ck2_Phospho_Site 26-29;
	DEX0253_181	Amidation 51-54; Pkc_Phospho_Site 6-8; 51-53;
10	DEX0253_183	Ck2_Phospho_Site 17-20; Pkc_Phospho_Site 36-38; 39-41;
	DEX0253_184	Pkc_Phospho_Site 31-33;
	DEX0253_186	Pkc_Phospho_Site 18-20;
	DEX0253_187	Ck2_Phospho_Site 42-45; 77-80;
	DEX0253_188	Ck2_Phospho_Site 21-24; 86-89; Myristyl 25-30;
15		Pkc_Phospho_Site 73-75;
	DEX0253_189	Amidation 25-28; Camp_Phospho_Site 10-13; Ck2_Phospho_Site 37-40;
	DEX0253_191	Ck2_Phospho_Site 11-14; Myristyl 16-21;
	DEX0253_192	Ck2_Phospho_Site 48-51; 61-64; Myristyl 57-62;
20		Pkc_Phospho_Site 22-24; 58-60;
	DEX0253_193	Ck2_Phospho_Site 43-46; Pkc_Phospho_Site 18-20; 34-36;
	DEX0253_196	Asn_Glycosylation 3-6; Ck2_Phospho_Site 8-11;
	DEX0253_197	Myristyl 4-9;
25	DEX0253_198	Asn_Glycosylation 46-49; 54-57; 99-102; 123-126; 135-138; 182-185; 195-198; 223-226; 229-232; 272-275; 305-308;
		Camp_Phospho_Site 259-262; Ck2_Phospho_Site 75-78; 82-85; 114-117; 197-200; 208-211; 247-250; 308-311; 382-385; 395-398;
		Myristyl 173-178; 316-321; 320-325; 334-339; 339-344;
		Pkc_Phospho_Site 63-65; 82-84; 101-103; 114-116; 157-159; 187-189; 197-199; 212-214; 225-227; 293-295; 395-397;
30		Receptor_Cytokines_1 126-138; Receptor_Cytokines_2 304-310;
	DEX0253_199	Camp_Phospho_Site 6-9;
	DEX0253_202	Amidation 38-41; Asn_Glycosylation 2-5; 45-48;
	DEX0253_203	Camp_Phospho_Site 24-27; Pkc_Phospho_Site 20-22;
35	DEX0253_204	Asn_Glycosylation 65-68; Pkc_Phospho_Site 54-56;
	DEX0253_205	Ck2_Phospho_Site 6-9; Myristyl 46-51; 47-52; 59-64; 63-68; 78-83;
		Pkc_Phospho_Site 6-8; 19-21;
	DEX0253_207	Ck2_Phospho_Site 57-60; 61-64; Myristyl 69-74;
		Pkc_Phospho_Site 34-36;
40	DEX0253_208	Camp_Phospho_Site 5-8; Ck2_Phospho_Site 37-40; Myristyl 26-31;
	DEX0253_209	Pkc_Phospho_Site 52-54;
	DEX0253_210	Camp_Phospho_Site 34-37; Ck2_Phospho_Site 9-12;
		Pkc_Phospho_Site 9-11; 33-35;
45	DEX0253_211	Amidation 67-70; Camp_Phospho_Site 69-72; Ck2_Phospho_Site 5-8; 10-13; 23-26; Myristyl 35-40; 54-59; Pkc_Phospho_Site 67-69;
	DEX0253_212	Prokar_Lipoprotein 6-16;

WO 02/48370

PCT/US01/51341

-131-

DEX0253_213	Asn_Glycosylation 44-47; Camp_Phospho_Site 37-40; Ck2_Phospho_Site 2-5; Pkc_Phospho_Site 8-10;
DEX0253_215	Asn_Glycosylation 3-6; Ck2_Phospho_Site 28-31;
DEX0253_216	Ck2_Phospho_Site 9-12; Pkc_Phospho_Site 32-34;
5 DEX0253_217	Camp_Phospho_Site 28-31; Myristyl 12-17; Rieske_2 35-40;
DEX0253_218	Ck2_Phospho_Site 21-24; Myristyl 29-34;
DEX0253_221	Camp_Phospho_Site 51-54; Ck2_Phospho_Site 29-32;98-101; Myristyl 21-26;70-75;75-80; Pkc_Phospho_Site 50-52;57-59;104- 106; Prokar_Lipoprotein 66-76;68-78;
10 DEX0253_222	Camp_Phospho_Site 51-54; Ck2_Phospho_Site 29-32;98-101; Myristyl 21-26;70-75;75-80; Pkc_Phospho_Site 50-52;57-59;104- 106; Prokar_Lipoprotein 66-76;68-78;
DEX0253_223	Ck2_Phospho_Site 75-78;
DEX0253_224	Ck2_Phospho_Site 96-99; Myristyl 72-77;76-81;105-110;
15 DEX0253_225	Pkc_Phospho_Site 63-65;87-89;
DEX0253_226	Asn_Glycosylation 3-6;8-11;9-12; Pkc_Phospho_Site 12-14;
DEX0253_228	Camp_Phospho_Site 14-17;
DEX0253_229	Ck2_Phospho_Site 14-17;
20 DEX0253_230	Asn_Glycosylation 41-44; Ck2_Phospho_Site 20-23;39-42;
DEX0253_231	Myristyl 25-30;30-35;37-42;45-50; Pkc_Phospho_Site 9-11;20-22;
DEX0253_232	Ck2_Phospho_Site 15-18;
DEX0253_233	Camp_Phospho_Site 25-28;
DEX0253_235	Ck2_Phospho_Site 7-10; Rgd 8-10;
25 DEX0253_236	Ck2_Phospho_Site 21-24; Myristyl 4-9; Pkc_Phospho_Site 21-23;
DEX0253_237	Ck2_Phospho_Site 12-15;26-29; Myristyl 10-15;24-29;30-35;
DEX0253_238	Ck2_Phospho_Site 41-44;
DEX0253_239	Ck2_Phospho_Site 2-5; Pkc_Phospho_Site 46-48;
DEX0253_240	Pkc_Phospho_Site 9-11;17-19;
30 DEX0253_241	Camp_Phospho_Site 2-5; Ck2_Phospho_Site 5-8;
DEX0253_242	Pkc_Phospho_Site 20-22;
DEX0253_243	Ck2_Phospho_Site 30-33; Myristyl 4-9;27-32;36-41;
DEX0253_245	Asn_Glycosylation 40-43; Pkc_Phospho_Site 12-14;48-50;
35 DEX0253_246	Asn_Glycosylation 7-10; Glycosaminoglycan 9-12; Myristyl 10- 15;19-24; Pkc_Phospho_Site 23-25;
DEX0253_247	Ck2_Phospho_Site 3-6;35-38;
DEX0253_248	Ck2_Phospho_Site 12-15;27-30;
40 DEX0253_250	Glycosaminoglycan 30-33; Myristyl 18-23; Pkc_Phospho_Site 22- 24;
	Asn_Glycosylation 21-24; Ck2_Phospho_Site 14-17;28-31;
	Myristyl 26-31;

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have
 45 a phenotype of interest. cDNA is then generated from these RNA samples using
 protocols known in the art. See, Sambrook (2001), *supra*. The cDNA is then used as a

WO 02/48370

PCT/US01/51341

-132-

template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 147. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5066): 706-9 (1991). See also Sidransky *et al.*, *Science* 278(5340): 1034-9 (1997).

5 PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then
10 cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holtan *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are
15 nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium
20 iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical
25 Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological

30 Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific

WO 02/48370

PCT/US01/51341

-133-

antibodies, at a final concentration of 0.2 to 10 $\mu\text{g/ml}$. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphatic (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 $\mu\text{g/kg/day}$ to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day , and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 $\mu\text{g/kg/hour}$ to about 50 mg/kg/hour , either by 1-4 injections per day or by continuous

WO 02/48370

PCT/US01/51341

-134-

subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

5 Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation
10 auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable
15 polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl
20 acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP
25 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

30 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable

WO 02/48370

PCT/US01/51341

-135-

carrier, i. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

WO 02/48370

PCT/US01/51341

-136-

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 µg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the

WO 02/48370

PCT/US01/51341

-137-

treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of
 5 expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a
 subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and
 separated into small pieces. Small chunks of the tissue are placed on a wet surface of a
 tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned
 upside down, closed tight and left at room temperature over night. After 24 hours at room
 10 temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of
 the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and
 streptomycin) is added. The flasks are then incubated at 37°C for approximately one
 week.

At this time, fresh media is added and subsequently changed every several days.
 15 After an additional two weeks in culture, a monolayer of fibroblasts emerge. The
 monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al.,
 DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine
 sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf
 intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using
 20 glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified
 using PCR primers which correspond to the 5' and 3' end sequences respectively as set
 forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer
 includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear
 25 backbone and the amplified EcoRI and HindIII fragment are added together, in the
 presence of T4 DNA ligase. The resulting mixture is maintained under conditions
 appropriate for ligation of the two fragments. The ligation mixture is then used to
 transform bacteria HB 101, which are then plated onto agar containing kanamycin for the
 purpose of confirming that the vector has the gene of interest properly inserted.

30 The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture
 to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf
 serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then

WO 02/48370

PCT/US01/51341

-138-

added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

5 Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

10 If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or
15 after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences
20 into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO 90/11092, WO 98/11779; U. S. Patent 5,693,622;
25 5,705,151; 5,580,859; Tabata H. et al. (1997) *Cardiovasc. Res.* 35 (3): 470-479, Chao J et al. (1997) *Pharmacol. Res.* 35 (6): 517-522, Wolff J. A. (1997) *Neuromuscul. Disord.* 7 (5): 314-318, Schwartz B. et al. (1996) *Gene Ther.* 3 (5): 405-411, Tsurumi Y. et al. (1996) *Circulation* 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers
30 injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

WO 02/48370

PCT/US01/51341

-139-

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention
 5 may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain
 10 sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide
 15 production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue.
 20 Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to
 25 the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin
 30 fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

WO 02/48370

PCT/US01/51341

-140-

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 µg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 mm cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection

WO 02/48370

PCT/US01/51341

-141-

may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be used to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

5 Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals.

Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific
10 embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection
15 (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)); blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell.
20 Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993)); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723
25 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated
30 oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campbell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810-813 (1997)).

WO 02/48370

PCT/US01/51341

-142-

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than

WO 02/48370

PCT/US01/51341

-143-

one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

WO 02/48370

PCT/US01/51341

-144-

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including
5 human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence
10 associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve
15 expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the
20 body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible
25 cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the
30 introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function

WO 02/48370

PCT/US01/51341

-145-

of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

5 All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The
10 present invention is limited only by the claims that follow.

WO 02/48370

PCT/US01/51341

-146-

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes
5 an amino acid sequence of SEQ ID NO: 148 through 250;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID
NO: 1 through 147;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid
molecule of (a) or (b); or
 - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is a cDNA.
- 15 3. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of a colon specific nucleic acid
(CSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule according to claim 1
under conditions in which the nucleic acid molecule will selectively hybridize to a colon
specific nucleic acid; and
 - 30 (b) detecting hybridization of the nucleic acid molecule to a CSNA in the
sample, wherein the detection of the hybridization indicates the presence of a CSNA in
the sample.

WO 02/48370

PCT/US01/51341

-147-

7. A vector comprising the nucleic acid molecule of claim 1.
8. A host cell comprising the vector according to claim 7.
- 5 9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.
- 10 10. A polypeptide encoded by the nucleic acid molecule according to claim 1.
11. An isolated polypeptide selected from the group consisting of:
- (a) a polypeptide comprising an amino acid sequence with at least 60%
15 sequence identity to of SEQ ID NO: 148 through 250; or
- (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 147.
12. An antibody or fragment thereof that specifically binds to the polypeptide
20 according to claim 11.
13. A method for determining the presence of a colon specific protein in a sample, comprising the steps of:
- (a) contacting the sample with the antibody according to claim 12 under
25 conditions in which the antibody will selectively bind to the colon specific protein; and
- (b) detecting binding of the antibody to a colon specific protein in the sample, wherein the detection of binding indicates the presence of a colon specific protein in the sample.
- 30 14. A method for diagnosing and monitoring the presence and metastases of colon cancer in a patient, comprising the steps of:

WO 02/48370

PCT/US01/51341

-148-

- (a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and
- (b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the colon specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of colon cancer.
15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.
16. A method of treating a patient with colon cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the colon cancer cell expressing the nucleic acid molecule or polypeptide.
17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

20

WO 02/48370

PCT/US01/51341

1

SEQUENCE LISTING

<110> Sun, Yongming
Recipon, Herve
Ghosh, Malavika
Liu, Chenghua
DiaDexus, Inc.

<120> Compositions and Methods Relating to Colon Specific
Genes and Proteins

<130> DEX-0254

<140>

<141>

<150> 60/244,717

<151> 2000-10-31

<160> 250

<170> PatentIn Ver. 2.1

<210> 1

<211> 421

<212> DNA

<213> Homo sapiens

<400> 1

```
cttaaaaaata attctagat tcttggcatt attaaaaacc taaatccttt taggaactat 60
tgcgaagaaa gaatatgata ttctgaagag ctccagtgcta atattagcat tggttatggg 120
agtgaagagc cagataaate ttttagttgg gaagtatgct ttgaggtata ctctcttata 180
atcattaagt aataagtaa aactatatta catagataat gtgtaactct ctgtattaca 240
tagaattgct gcagaatgta gataggaaaa ataaagtgtg tcaataattt tcaacatott 300
tattgagata cagttaatct gccatgaaga ttgctactat ataaagtgtt catttcagtg 360
tgtttagcta gtgtatttgc agagttgtgc agtcatcacc acagtaactt ccctaacact 420
c 421
```

<210> 2

<211> 426

<212> DNA

<213> Homo sapiens

<400> 2

```
agaaaccat tcttaagtga actgccactg ctctagtcta acttaggttg gcagagagcc 60
agcactttct tcagcattca gggcagggag cactagggat attggcattg cttattacta 120
agcacacaga tacaagtatg tgcctgatat gtaaccaaag taagttaaac tctttattta 180
atottagcac ctgtctaaag gctgggtgac tgtatttata gataggaaa actgaaatt 240
ggggggccaa gggcagtgaa gtgaagtgaac ttgtttatg atacacagct agtaggaata 300
ttagcactgg aatttgaatt tcatgccatc ccattccaac ctgggtgttt actactccc 360
actatctccc aagcatgggt attttaggaa atatagaaca ttttctcagc aatcacagct 420
tatttc 426
```

<210> 3

<211> 1015

<212> DNA

<213> Homo sapiens

WO 02/48370

PCT/US01/51341

2

<220>
 <221> unsure
 <222> (469) .. (574)
 <223> a, o, g or t

<400> 3
 agaaacccat tcttaagtga actgccactg ctctagctga acttaggttg gcagagagoc 60
 agcaactttct tcagcaattca gggcagggag cactgaggat attggcatty cttattacta 120
 agcacacaga tacaagtatg tgccttgatg gtaacccaaag taagttaaac tctttattta 180
 atcttagcac ctgtctaaag gctgggtgac tgtatttata gatgaggaaa actgaaatt 240
 gggggccaaag gggcagtgaa gtgaagtgac ttgtttctatg atacacagct agtaggaata 300
 ttagcaotgg aatttgaatt toatgocato caattccaaac cctgggtggt tactacttcc 360
 caatatctcc caagcatggg tatttttagga aatatagaac attttctcag caatacagac 420
 ttattctctc attctctctt ccacatactc tottttccct taacaacann nnnnnnnnn 480
 nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 540
 nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 600
 tgaatgcttt tcatltgatg tctttataga catagtgaag taaaagttta tctaggatat 660
 atgttgggag gtgaggaaga cttaggtaga gagggtccaa accagttggt actgcttagc 720
 tcaatttcag acatacttcc tccagccctc tctaaactac ccaccagctc tgcctctctc 780
 ttcttaggtt ctgtggcaat tgcctgggtg gcaataactg tatggcatgc tgttctcacc 840
 agtcaggttg agactagcat cgaagggcac atcaacaaga aggagagacg toggctacag 900
 gccaagggca gagtgaagtag ggttgaaggc togggtggg taggtgggta aotgaattg 960
 ctctccctgt aaacagaggg catgggcagg gctgactagg gcaagcatta taaaag 1016

<210> 4
 <211> 1358
 <212> DNA
 <213> Homo sapiens

<400> 4
 ctctctggggc tcgtttttct caggagggct cattctgac dataaacctt ctctctgggg 60
 tttagggctcg agctgttctc gatgtttatc ggagactggg atcaaaagcta tccaggtcat 120
 aaatctctctc ctgtggtctgt tgggcccacg ggcagctgaa gagggttgac agccctttgg 180
 aectcaaaag aaaaaatgtg ctctactcca ccaactccca gctctgcaaa gaagctgtcc 240
 totgagaagc catggctggg cgttctcatt ctggggagct gctgaaaaga gctgggaggg 300
 cgaagaagaac ttgctgttgc tgggggagag gaagcctggc cttagaggag ggttgcaagg 360
 gtggtctctg tctgtgtggg ggttggggga ccttgtgtgc cttttctctg tggctgtgaa 420
 atgctttatg agtacttcca tagggggatg gacagggagt cggggagata aactcagcca 480
 caaggcccga gggcctcagg aaacttgca ccaacctctc cattttacag aagaaaactg 540
 tgcctgggaag gttgaagggt ttgttccacg tcacacaaac agggatcctt aggcacagca 600
 gaccaggaaa ccatttccaa actgtccaaag catggcagag tatcaagacc taaggaaaca 660
 tccagacacc atggaagcat tgggaaaagc ctctcttagct ttggaagctc ctcatgttcc 720
 ttgagtgctg atggagccca tgaactcggg gttttgtaga caactcaggg attacatgac 780
 tggtaacccct gacaaagtca aggtctgtgg acaaaatgag tccgaggatt taaggggcac 840
 gctggggcga ggaactggtg ggcgtgtggg agtgccctct tactgggag gcttctctcc 900
 tctctggtgat ggggggttcc tcagcacaaa agtgaaaggg tggaggggct ggaggagcag 960
 gaatctctct tgttgatagg tatgagccct tgaagtcctt ttctttgtcc caggattcat 1020
 ggcgctctcg gggctgatct ttgagtttcc aagcatgggg tgcagagacg tttaggtaaa 1080
 ctcttaccgt cctctctctt cgtcagggct tcccaggaa caacaatgac caagaaggaa 1140
 gggattgtag aaatagctta accctttcat ttaccaacgt ggaaattgaa gcccaaggaa 1200
 gggaaaggac cgtctgtgga agggagagcc atccagagaa agagaacctg agatcttccg 1260
 ctgggattcc caggaaagtc agcccgagct gattccacga acaaatgcat gcaaaccttg 1320
 ctatcaataa attacacatg caattacgta aaacacat 1358

<210> 5
 <211> 2375

WO 02/48370

PCT/US01/51341

3

<212> DNA
 <213> Homo sapiens

```

<400> 5
cttttctctt gttgagtga aatggagaa agctgtctac gctcgtctgc tgacatdago 60
tattttctag gatgaccctc cgagacaggc cagggtcatt agacccaatt tgggttctag 120
caaatatgtg ttatttctct catgctgtgg ccacaggctg gtttcttggg tgcaatgaat 180
agctgcaggc ttattagggt gtottttttg atggaatgat gtttcccgat gtcctatagaa 240
cactccggac ccgggagagt gaagactctg cctgtcggac ttgcttttag aagatctctc 300
tcacacctcc catggcagaa gttgcttacc agaggggagc agttttatgg atgtggctga 360
gaccttaaac ttgaggcaac ccatctgagg tggcatccag aggagaotgg ctgggcccctc 420
cttcacacct gatgtagtgc tgtttctagg atctcttttc aatcagcaaa acaggggagt 480
ttccaaagag gttgtgattc cctgcatcc cactgggtca agtggagggg acgggaaaaa 540
gctatgagag gtttgtgacc acacagactc tctgggcccc ctgtcctttt ggaagaagaa 600
cagggatgaa atataatcaa gcaatttaac acccccatca tcaccaagaa caacagtatc 660
aacaaagaga acagggaaca caaaaccac ggaagaaaca ttccctttctc agctcagatc 720
ttactctgtg cgttctctct ctgctctgtc ttggtgtgtg gtttagagaa acatggacaa 780
cgctgtttgg aagaacaggc gagcgagggt ggggaatttc agagggctgg gcccaacgpc 840
tcacacacctt cccagttta acctttgaca ggaatttacc ctctctctga tcagcatgpc 900
ttctgttcaa aaggcttaag ccaaccagct gtgtcccttt cccagaaag caagggcaga 960
tcgcagtggt tctgttgatg agagaacttt aaggggccaa tcagtccctg ggcaccacct 1020
cctgagctgc ttttctccag gaggtctgat tctgacatc aaacctctc ctcggggttt 1080
agggtcgagc tgttctgat gtttatcaga gaotggagat aagctatcc aggtatataa 1140
tctctctctg tggcgttgg gccccagggc agctgaagag ggttgaaagc cctttggacc 1200
tcacaagaaa aaatgtctc tactccacc actccagct ctgccaaaga cgttctctc 1260
gagaagccat ggcgtggcgc ttccattctg ggggagctgt gaaaagagct gggaggccga 1320
gaagaacttg cgtgtgctgg gggagaggaa gctbaggctt gaggaggagg tgcaggtgtg 1380
gctctctgct gtgtgggggc tgggggacot tgtgtgctt ttcttctg ctgtgaaatg 1440
ctttatgagt acttccatag gaggtggagc aggaagtcgg ggagataaac tcagccacaa 1500
ggcccaaggg cctcagaaac ctgccacca accctctcat ttacagaag aaaactgtgc 1560
ctggaaagtt gaagggtttg ttccagtag cacaaaccag gatccttagc acagccagac 1620
caggaaaacca ttccaaact gccaaagcat ggcagagtat caagacctca ggaaccatcg 1680
agacaccatg gaagcattgg gaaaagctc cttagctttt gaagctctc attgttctg 1740
agtgtgcatg gagcccatga ctgggggttt ttgtagacac ctoagggatt acatgactgg 1800
taccctgac aaagtcagg ctgctggaca aaatgagtc gaggatttca ggggacagct 1860
gggggcagga gotggtggg tgttgggagt gccctttac tgggaggtt tcttctctc 1920
tgggtgctgg ggttctctca gcacaaaagt gaagggttgg aggggctgga gggagcaggaa 1980
tctctcttgc cgtatggtat gaggccttga agtcttttc ttgtcccag gattcatgga 2040
cgcttcgggt ctgactttg agtttctaac catgggtgc agagacgttt aggtaaactc 2100
ttacgctcct ctctctctgt cagggtctcc caggaatcaa caatgccaa gaagaaagg 2160
attgtagaaa tagcttaacc ctttcattta ccaagtgga aattgaagcc cagggaagg 2220
aaggaccggc tegtggagg gagagccatc agcagaaaga gacctgaga tcttgcctg 2280
ggattccag gaagtcagc ccgagctgat tcacagaaca aatgcatgca aaccttgcta 2340
tcaataaatt aacatgcac ttacgtaaaa cact 2375

```

<210> 6
 <211> 410
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (34)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (56)

WO 02/48370

PCT/US01/51341

4

<223> a, c, g or t

<220>

<221> unsure

<222> [108]

<223> a, c, g or t

<400> 6

```

cagagtcgaag gccccaaggc cgtgggtctt tgaggagggt gtttttgaga catgtncagg 60
gacaaaacctt gcaacaagag aactcttaat cccatagctg atattgcnaa ttagcttttc 120
ctttacaaaa tatgttcacac cctaagtatg tttactataa tgttagctgt taaagacccc 180
tccatccccc aaaccattta ccttcaata aaaaatgggc caagttgcnaa gggtragaca 240
ggatgtatatt gaaattttag aagtttgaat aatttcttta acacaaaagc attttttct 300
tatttctcat aattttgaat ctatttaaat acaacttcag tgcgtattaa totactaaat 360
tggaagattt aagatttata gctgggtgca gtggctacac ctgtaactct 410

```

<210> 7

<211> 416

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> [114]

<223> a, c, g or t

<400> 7

```

ctcagcgaga gtcaaggccc caaggccgtg ggtctttgaa ggaggggttt ttgagacatg 60
tacaggagaca aacctagcaa caaggagaact cttaatccca tagtgatatt tgcnaattag 120
cttttctcttt cacaatatatt gtcacacctt agtatgttta ctataatggt agctgttaaa 180
gacctctctct acccccaaaa caatttacct tcaataaaaa tgggtgccag ttgcaagggt 240
tagacaggta tgtattgaaa tttagaaagt ttgaataatt tctttaacac aaagcattt 300
ttttcttatt tctcatactt ttgaatctat ttaatacaaa ctttcagtgt gattaactta 360
ctaaatgtga aagtttaaga tttatagctg ggtgcagtgg ctacacctgt aatctt 416

```

<210> 8

<211> 786

<212> DNA

<213> Homo sapiens

<400> 8

```

atgttctctag tagaacacaa agtttgccta ggtaacacac aagtaagcat taaatgcctt 60
cctgttgttat ctgagaagtt tgttatgaaa tattttggtt accgtgcgat agtcagtgta 120
ggaggagcag atgmatttta gctgtgttta tgtgtgctgt aaagactat acgtgcttgt 180
attagtcaga atgagtacac cactaatttt tgtatggtta gagatttata ctaagctcat 240
catcagtttc tataetttcag tgagataaaa ctgagtcaga ttgattttta ggttagcaca 300
gtagaacacag ctaattttat tccctgatt tgatctcat ctattgatta tataaactaa 360
agaagctaaag aacaattaac ccttacaggg ttacacagtc aggagatgct gaactgagat 420
tcagtgtaga aagtcgtctt tcagagccta tgcctttagt ctttatgcta agtttaactt 480
gtttaaatag caagattatg aagcactata cagtgacctc gtatagaaca aaatatagta 540
tattgattat tagagaaact acatattaga ctgttgaca tacgtgggca agtatitgtt 600
aaatcatttc agttgcctaa atttaagcaa ctgtgctgtt taaaacatgc tcattoacat 660
tttttcttaa tctagaaagt cactctgaa taattgcttg tttagatttt ctcatttggg 720
gtgggaattt tatattaaaa ttttaactaa tattotaaca atacagagtc tgaacctaaa 786
gtccac

```

WO 02/48370

PCT/US01/51341

5

<210> 9
 <211> 1509
 <212> DNA
 <213> Homo sapiens

<400> 9
 atcagacctt gtcggttagc ttctggatct cagaatcact tatacttaag tccaggctgt 60
 tctcaaatat ggcagaagac abctgctgtt aatagctgac agtaaatatc acaagtaaaa 120
 acatggaaaa ttaaaagtcag aaagctagg aagcttttct atcattttta attttorgca 180
 aaatacaga cataatcagg tttagatctt gcttgtagtg gataaatatc atctgtatct 240
 cttctttttt catattactg cattcagacg ataatttgct tccagatata ttgctcatct 300
 aatcgtrcat agactggaaa taagtagtaa catctcccaa tcttaggaag cattataac 360
 tagcttttgc ctttttgggt gttgatagac tagtggtagt tataagcttt ctagctcttg 420
 aaaaagcaaa ggaagattaa aataatcata ggaataaaa atacttttaa acccttctag 480
 tctttaattt taaaatgttc cagtagaaca caaatttgct caggttaaac acaagtaagc 540
 attaaatgco tctctgtgta tctgagaagt ttgttatgaa atatttttga aacogctgca 600
 tagtcagcgt aggaggagca gatgaatttt agctgtggtt atgtgtgctg taasagacta 660
 tacgtgcttg cattagtcag aatgagtaaa ccaataattt ttgtatggtg agagatttat 720
 tcatcagttt ccatcagttt ctataattta gtgagataaa actgagtcag attgattttt 780
 aggtagcaca tctagaaaca gctaatatta ttccctgat ttgatcttca tctattgatt 840
 atataaaata aagaagctaa gaacatttaa cctctcagag gttaacacgt caggagatgc 900
 tgaactgaga ttcagtgtag aaagtctgtc ttcaagacct atgcttttag totttatgct 960
 aagtttaact tgtttaata gcaagattat gaagcactat acagtgaact cgtatagaca 1020
 aaaaataagt atattgatta tttagaagac tacatattag actgttgtad ataogtgggc 1080
 aagttattgt taaatcatct cagttgccta aatttaagca actgtgtgct ttaaaacatg 1140
 ctcatcaca ttttttctta atctagaag tcaattctga ataattgctt gtttagattt 1200
 tctcatttgg tgtgggaat tatattiaaa attttacta atattctaac aatcacagat 1260
 ctgaacctaa agtccagaag aattttaagt catgcccag acaggatgaa cagtatagca 1320
 aatcagaata atagactgtg aggggggcta ggggggaacc catgagaatt ttaggatgtc 1380
 aagataaagc ttggaattga ggtaaaggca tcagataagg aagtgatcat ttcatatctt 1440
 gttttgtctt gaaatatatt atattttaca tcaaaaagt agtataactg ttattttgct 1500
 aatgcacag 1509

<210> 10
 <211> 283
 <212> DNA
 <213> Homo sapiens

<400> 10
 ctaagtaato cttgtcaggg gaggtggttc ccaattcgtg actcttggac cttggggcat 60
 cttatgattt attgttatca ctacacaatag ctgcctatgt gtcattgtctt ctgctacata 120
 ttttatgttt tatttcaagc ttttaaaaga ttttctatgt tcatgatgtt tgttaagcag 180
 gactaggctg tatgtacata ttgaaatga aagtttcaaa aaacatcatt tacctttaat 240
 atgtgtgaca caacttgcata tttttoaatt aatctatttt att 283

<210> 11
 <211> 736
 <212> DNA
 <213> Homo sapiens

<400> 11
 gtctttctga agggagcaco tcggaatcct tccgaacttt ccaagtcact ccatgattca 60
 gagatactgc ctctctctctc tctgggattt tatgtgttct tgatagttaa tbtgtgatgt 120
 atttgcatac ttgctctctt tctctttcaa gaattgatca ttttatatgc tgtttggaga 180
 aaaaagaagc ttttctttagc aaggaggttt cagaatctat tttggatttt ctgttaagct 240
 ttaatttttag tctaggggac agcatctctc atcccgagct aaattttctg ctttgacctg 300
 catggattat tttttcagge tgggaattt ctgggaactt acctgtatga tggggcaact 360

WO 02/48370

PCT/US01/51341

6

```

ggtttgggtg cagagtaaga aggtggaaga atgagctgta cttgggttaag cagttgaaac 420
cttttttggag caggatctgt aaaagcataa ttgaatttgt ttcaaccccg tggattccag 460
tgggcccagag agcgcaacag gtttgcagat ttcttttgaa attccttttt ccccccctcc 540
tttgccctcag caaaagaaaa gaatccatata aacaggttca tggtaatttg cttggctttt 600
cagcaatttat tctgaagact ttataatatt ttlaaacttg acctgggaa acagagggtt 660
ttgtgggtga ggtgtattta tattacttta aggggtgcac ttttaaaaat cttattctgt 720
gtttgtacaa agaagc 736

```

```

<210> 12
<211> 547
<212> DNA
<213> Homo sapiens

```

```

<400> 12
ccggttagaa tagagcttcc acaagctcct accttgatat ctgcccctct agcaactggg 60
ccaatglttc ctgctttccc tctatgtgaa ccttcctgtt tctcaatata acttggatta 120
atcaactcct cttctggcct ctaaaagata gtaactctaa caacttttcc cttcttttca 180
tgcaattcct accttggcct tctctgtctg accttttctc atcgacatct aaacatgctg 240
ttatgtcttc caaccaaaa aatgcacaaa accttttccg cctatgcttc acctatctc 300
cagctgtagt cctcttctct ccttttactc tcttttatta tagcttaatt tcttgaagg 360
atggaatgct caactcctct cctcccatcc ttctctgaac ctaccccaat ctgctttttg 420
tcccactgtt gccagtgaga gggctcttga taagctctcc ctccattgac ttcagtttgc 480
tcaatgaat ggcagttct cagtctctat cttaactgac ttccagcag ctttagtag 540
taccaga 547

```

```

<210> 13
<211> 1559
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> {1337}
<223> a, c, g or t

```

```

<400> 13
gtttctacgt taaaacaacc tottccctct aactttaaaa tcagatacag taaaagcctc 60
ttgttgagga tgtggttata ttggtagatg agagtgtgtg agaaacaggt agaaacttac 120
ctggcaaaag aactagtact gtatcttgac ttgttacatg gcaacaatca attagatgat 180
aatttctatt taaaagcatt ctatatgggg aaagacatgt tcaatttgat aagtaagac 240
aaaatctagg tttttagttg atgtgtgttg taacatgtgt ctttggaaag caaacctaac 300
tatgtattat tgacattaaa aatgatgact taatgtgtgg taaatcctgt actcagaaga 360
tactcaatga tgatccatcc ctggctataa cctatgaact aaacgaattt tttaatcttg 420
gtgcttatta ttacttctag ctgctctctc taataatccc aacacdttgt gctctcatcc 480
gtgctctcag ttacttctct gccccgtttt tcaatgagaa gacagaagca gttagaatag 540
agcttccaca agctcctact ttgatatttg cctctctagc actggggcca ctgtttcttg 600
ctttccctct atgtgaactc tccgtgtttc taatctcttc tggatttaac acctctctc 660
tggcctactc aaagatagta actctaacaa cttttccctc tctttcatgc aattcctact 720
ttgctctctc ctgtctggac ttttctcttc gatataaaa catgtcttta tgtctcccaa 780
ccaaaaaaa tgcacaaacc ctttcagccc tatgctcacc catcatccag ctgtagtctc 840
cttctctctc tttactctcc tttattatag cttaattttc tgaaggatg gaatgtccac 900
ttctctctc cccatctctt cctgaacctc ccccaatctg ccttttctcc ccactgtgac 960
agtgagaggg cctctgataa gctctcctct cactgaactc cagtgtctca atgaatggg 1020
cagttctcag tctctatctt acctgaactt ccagcagcat tttagtactac bagcagtcac 1080
tcactcttga aatactctct tttcccatat ctctaatgct ttaagtcaaa aggggtccat 1140
gacccagttc ttacataact taaactcttt ggctaagctc attatctggg acctcatcca 1200
gtcttggggc ttttaatact atatggggac aactacagcc gagaaccttt cctgaactt 1260

```

WO 02/48370

PCT/US01/51341

7

```

tagactcttt  tgtccagaag  attatacaaa  ttctctgttt  ggttatagaa  tttagaatgc  1320
cccaaatcaa  gataatnctc  cctcaattct  gtctctccta  taagcttccc  caatcggtaa  1380
atgaaaactg  tgtccttcta  gttaatcata  ccaaaatcct  aaaaatcctc  cttaactcct  1440
ctcatctctg  atatccatat  ccaaccatg  agcaaatact  gtcaatctgc  cagaatccaa  1500
acatctctcc  agccccatg  ccaccacct  ggtccagcc  accaccaggc  cttgcttag  1559

```

```

<210> 14
<211> 1455
<212> DNA
<213> Homo sapiens

```

```

<400> 14
ggagtgtgaa  ggtggtgagt  catggggagt  ccaagggaa  ggggtatgaa  gggaggctc  60
aaatgagga  caagtggaga  aggtagcttg  ggaaggaga  aggtagcttc  tccattatg  120
atgggaaga  cagaggaaga  ggtcagaat  acagtgaat  aggtgata  tggagtgag  180
ttgaggaac  tcaactctg  gtctgaaac  cctaggttt  ggggggttt  gagatggga  240
agaggttaa  agtcagttgt  tctagcaat  atggtttga  attatttgt  gatgtttaa  300
aatattgctg  aagagaagtg  aagtctatcc  tagagttgga  tggtagagat  atttagtgga  360
cctaccagat  ccatgttgtg  attctttcca  gtatcattca  gcagcccttg  ggcagttgcg  420
agggcaagca  tcaagtgtgt  atggagattt  tcccaggtgg  gtgtggttga  aggcaggaa  480
gaacaggttc  agggcacat  tacaagaaga  agtgactgt  aaggtccagg  ctgagcagga  540
aggtaaagca  agaaggaac  atgagttgt  gaaggaagt  tttaggggat  gagggagcag  600
gagagatgaa  cagttgcagg  atgtagctag  agtggcagtg  tttagatctg  gggccagaga  660
tctttacaat  gattatgaag  atcaagggc  attagaatca  agtatagaag  agccactgtt  720
tgatgttggg  atgtgagcat  gctgcaggtg  gatgtctgca  cattgatggg  gagaacatgg  780
tcactctggc  cctgctgggt  ctttgtaaa  gagaactgtc  tctgtcttg  gggccgtttt  840
catcatctga  tttagacagt  ggtcccaaca  tgggtgttct  tggacacatc  gtataaaatg  900
ttcatgggtc  aaggataaaa  tggaaaaaca  gagaaaaatg  cacagaaatg  tggccattgt  960
tgaacagcca  cagctgttcc  tttttggagg  attgttcttt  attotaaaaa  tgtatatatt  1020
ctattctatt  aaaaattttt  tttattggca  tttttttctc  ttttatgaaa  tggccatggg  1080
tagaaatttg  taatgtatcc  aattctcctg  tottcatgta  ttgcccctgt  gtgggggagg  1140
ggatgttggc  agtactggcc  aagaggtctg  gggcagaggt  gcaatgttag  acttctagcc  1200
tggagcattt  aattcttagt  acaagactct  ctaacattct  tctccctctg  ttccctgctt  1260
gggtgatact  gaggtattgc  aaccccatc  aacottagto  tttagggcag  tttagtgaga  1320
aacagagcac  cccacaccto  cctgcagatg  aagcatgagt  gagaaaaaca  acttctgatg  1380
tttgaagtta  ccaagatttg  ggagtgtgtt  gttattgcag  caaaaacctca  cctattctga  1440
ccaatcatgg  tggaa  1455

```

```

<210> 15
<211> 904
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (281)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (329)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (469)
<223> a, c, g or t

```


WO 02/48370

PCT/US01/51341

8

<220>
<221> unsure
<222> (471)
<223> a, c, g or t

<220>
<221> unsure
<222> (539)..(540)
<223> a, c, g or t

<220>
<221> unsure
<222> (662)
<223> a, c, g or t

<220>
<221> unsure
<222> (692)..(693)
<223> a, c, g or t

<220>
<221> unsure
<222> (701)..(703)
<223> a, c, g or t

<220>
<221> unsure
<222> (776)
<223> a, c, g or t

<220>
<221> unsure
<222> (785)
<223> a, c, g or t

<400> 15
tggaaaacca aaaattgat atgctaaagt atactcttaa aggtcttaat actttaaaag 60
tatatagatc tcatgaacat aattcatttg aggaaaaaaa tacaatcat ttcttgtecc 120
aggaaaacag caaatcttta atggaactct tttagcaatta tgacaaaag aatggaaaaa 180
tgtttaacca tatataaaag gctagacgtt tatgcgcaa tagtatctaa aggtcataga 240
atagttagga attctgtcat ttgtttttgt gtaataaata ncccttctct ttaccttcto 300
accctaataa tagatatcca ccattttgnt gtgattatcc aactatagag taccttttcc 360
aagaactcat tatataccaa agtaggagct tgcagacact gataatgctt tatttagttt 420
tgtagtgaac tacaattacc atttgcttag gaaaaaaaat aaagaadana nacaagtaaa 480
ttttttaaaa ctatgggtgt gtatatataa gttgataaaa atcctttggg agaaaactnn 540
tgtctgtgtg gtaagagca ttaaatagtc ataccctta gccctagtgtg tcttctatcc 600
tgaaaaaaa ttacaaaagc aaatactaac ttasgaaaaa aaactacagc actgaaaaga 660
tntgtgttaa tattgtttat gctaacataa amatgtaaa nnttatata ttgtttatcc 720
tgacttataa ttattacta tacatagtg aaattatgat acattggctt tggtagcag 780
ttttntaacc gctaataata taataccat actattaaca atctagaaaa atgattctgg 840
tataggttat gtgaaaaggc acaaaaataa attgtatata gtacactagc aatgaacagt 900
ctga 904

<210> 16
<211> 984
<212> DNA
<213> Homo sapiens

WO 02/48370

PCT/US01/51341

9

```

<400> 16
acagatttac  tctctgaat  ttccagaaa  tctagatct  tttaaatca  aggaaggctg  60
tattttgttt  tgttcagaa  ttttctattc  cagaaaatca  tctcaattga  cagcaaaagc  120
actttgtgtc  attgagcctc  ctgtgtaaag  caccgacgtc  attctgtagt  tgtctacact  180
gtatttcagg  tttatctaca  cgtaggagtg  agcatttgac  agcttccatg  tcttctagtg  240
gggttgagaa  tttacatatt  aagatcacaca  ttattttatta  tcaattactt  tcttctttca  300
atgtccattt  agagcactaa  aatatctttt  gtaggtagtt  gatattactt  atgaatttta  360
tttcaggaga  gcaaaaggaaa  atacaagata  gtgtatgaa  aaggggggac  cgggtgtgtc  420
agagtggctc  accaccgccc  tacacagtg  gctaatggc  tggagagtag  agctgactct  480
gcacagttgc  atgctgaccc  tctgaagaat  ttttttaca  aagcgtgacg  tctgctgaag  540
accttgacag  aattagcaaa  gcggttgaga  tgcatacttt  ggagtcagac  agactccagt  600
tcacatcttg  gctttttata  ttacagctgt  ataaccttag  acaatctatc  taccctctgg  660
ccgactccat  tctctcaatt  ataatatagg  ataatctgtg  aaatgcttcc  cacaagatta  720
ctattgcatt  tatttctctc  accactctta  atgaagagag  tcttgaacca  gataactcta  780
attgtcttca  gagtccaggt  ccccaagaaa  gattatgctt  tctaaaagct  agtctgttcc  840
cttccagtg  gagccatttc  attcatctg  ctctactctt  tacttggact  gctagcaaac  900
atggagctaa  gtactcatgc  ctatttctg  tggcttcttc  caaatagggt  ttcaatacta  960
tagtttgccc  tcactccatt  cccc  984

```

```

<210> 17
<211> 429
<212> DNA
<213> Homo sapiens

```

```

<400> 17
cgtgataaaa  atagtttgct  ctgagttttt  gcctttcttg  aatttaaatg  caagaaaaat  60
atgttcccta  cctctccagc  cccactctta  cctccctgtg  gcttgttaag  ccttcttctc  120
gcctcctgca  tcaacttctc  gatggagagt  gtatgaatgc  aaaagctcct  cctttagcac  180
ttaactagtg  cttaactctc  tgggtctctg  ccaatgggtc  ccagatgaag  gagtttgatt  240
ttaaaaatca  gagtttatgg  ctttttataa  ataacctctc  acctatttat  caaaagctcc  300
ttctaaataa  tatttacaac  aacaacaaatg  ataatggcta  ctatctagta  ttcccatctc  360
tccagacaat  gtgctgggtc  ctttccaaac  actgttttaa  tctttaccaa  caccagctcc  420
gccgctcta  429

```

```

<210> 18
<211> 734
<212> DNA
<213> Homo sapiens

```

```

<400> 18
cttttggacc  ataagcctca  ggaagctata  aggtattatt  gcattcttad  aactggggac  60
tcttctcttt  tgcrgaatca  caghttttca  atcttttcta  tttttgaaat  aggttaagaa  120
agaaataaat  ttctagaat  ttgaagaaaa  atcttaaac  atttgaatt  otttgttatg  180
atgactaata  taacgaatag  cactcaggtt  tatcaaatat  taacattttt  ccatatttgt  240
tatagaattt  ttttccatct  ttgtcacaga  aataatttct  ttatatatat  aatacatatt  300
tgaacactga  ttttacttga  tacatttaata  taatgtgat  gtgctgagat  gaataaatca  360
aagaacctct  tggagctctt  ggtgtgcaat  aagcactgtt  aacgaatata  aataaagtga  420
tattttctag  aaaaataata  ctggtctaca  atgctctatc  tgtcatttca  aagtctctaa  480
aangacttga  aaatccaatg  ccttttataa  ataaaattac  ggtaatctca  ttgggccaca  540
aaacctgttc  agaattgatg  tgaggctatt  aagatattta  tttctcttat  ttattagtga  600
atattcatct  ttaactacag  aaatactaac  gagtttgatt  acagggtgct  ttgacttccc  660
ctcaaggtgt  adattattgc  taactttctc  taaaaatcca  aacatctctg  attctgaaac  720
acattctaaa  cccc  734

```

```

<210> 19

```

WO 02/48370

PCT/US01/51341

10

```
<211> 1184
<212> DNA
<213> Homo sapiens

<400> 19
attctaaactc tgtgacatgc agtctgtgac actgagagtt acttgcacct tccctcggac 60
tggagatcctt ttctagtgcg gacattttat aattctattc tgtatcgtgt tcatttaagt 120
agtcctgtctt atcattacat taacatttat gaaagacttg ctggtatcat tggcttagcg 180
attatcttttc catctagatg ctttttttaa agaatatgta atgtttttaa 240
tgtacattttt agtttgattt aaattttaat caaggatttt tattttatac attacatact 300
gatcaactgtt ttatgttaac tctggctcta ataaaagaaa aataacaatt tggatatact 360
acaacaatga gagctcaggg taaaatcatg cataaataag acatatatgt gtatgaaactg 420
agttatatatg aaataattaa atgtaacaat cttttggacc ataagcctca ggaagctata 480
aggattatttt gcattcttcc acctggggac tcttcccttt tgcgtaatad cagtttttca 540
atcttttctca tttttgaaat aggttaagaaa agaaaataat tttctagaat ttgaagaaaa 600
atcttaaaac atttgaaatt ctttgttatg atgactaata taacgaatag cactcaggtt 660
tatcaaatatc taacattttt ccatatttgt tatagaattt ttctccatat ttgtacaga 720
aatcaattctt ttatcatatt aatacatatt tgaacactga ttttacttga tccattata 780
taattgtgatg ttgctgagat gaataaatca agaacctctc tggagctctt ggtgtgaact 840
aagctatgatt aacgaatata aaataagtgat tattttctag aaataaata ctggtctaca 900
atgctctatc ttgcatattca aagctctata aaagatctga aaatcaatg cttttttaa 960
ataaaattao ggttaactoa ttggccaca aaactcttcc agaattgatg tgaggctatt 1020
aagatattta ttctctttat ttattagtga atattcatct ttcaatcag aaatactaac 1080
gagtttgatt acagggtgtc ttagaactcc ctcaaggtgt acatatttgc taactttctc 1140
taaaatccca aacatctcgg attctgaac acatctaac cccc 1184
```

```
<210> 20
<211> 550
<212> DNA
<213> Homo sapiens
```

```
<400> 20
ctttcccgct cccggcccca gtgccttgca tgcagcaagg tcttgcatg tgcagcttc 60
cttaaggagc ctgcagcttt gctccaagc acacactggc agactctggc cagatgcctg 120
gcacaggggc tggggaggga aggcctgccc aaccccgctt ttccctttgc agatgagcat 180
ctccaaatc catgtttacc cagtcctctc taatgtctgc ttccaaactg tccagggggtg 240
ctaaaaagca cacattagga tgaattagaa catgcacagg tgcaggggcg ggtgtcatcc 300
cagaactcac agagcactgt gagggtcagc ccgctcagcc acatctttag gtcccaccag 360
catctccccc caggcatgga cctcccaat ttacctgtg agggctgcat ggagaaagatg 420
caggtctttag gaacagccag catcaccaga ggtgccactt agtgagtacc cagtgggctc 480
ccaacacogt gctgagctcc cagtgggaga accggaaccg tctgctgtgt ctctgttgtg 540
ttccagcatc 550
```

```
<210> 21
<211> 599
<212> DNA
<213> Homo sapiens
```

```
<400> 21
tactatgtgc cagacacagg agttttcagg atgagtcact aagataataa acacaaagtc 60
cgggcccagc tgccttgcat gcagcaaggt ctggcctgt gcaagcttcc ttaaggagcc 120
tgcagctttg ctccaaagca cacactggca gaccttggcc agatgcctag caccaggggtc 180
ggggaggggaa aggcctgcca acccccgctt tccctttgca gatgagcatc ctccaaatcc 240
atgttttacc agtctctctt aatgctgctt tccaaaatgt cagcgggtgc taaaaagcac 300
acattaggat gaattagAAC atgcctaggt gcaaggcggt gttgcatccc agaaactcac 360
gagcagcttg agggctcagc cgtctagcca catctttagg tcccaccagc atctccccc 420
aggcatggac ctcccaatt taacctgtga aggtgtcatg gagaagatgc aggtcttagg 480
```

WO 02/48370

PCT/US01/51341

11

```

aacaccagc atcaccagag gtgccactta gtgagtagcc agtgggctcc caaacccgtg 540
ctgagctccc agtgggagaa ccggnaaccgt ctgectgttc tctgtttgat tccagcctc 599

```

```

<210> 22
<211> 618
<212> DNA
<213> Homo sapiens

```

```

<400> 22
gaaaaactac tcttttttgt gtaagatat tttttatatt ttctttgctt gtaagagatt 60
attatcaatt tgttaagtata aaaactgcac gtatagtggg tagttgataa gaaaggtaga 120
taataaaact taaaagggat ggacacagat tgaaaaagcc cttgagtgc aagacaagag 180
ctctgacatt taacaggccc tggaaacogt cataggtctt aggtaggat atgctgtgct 240
cccaccatct taattaggtc ttatggaggt ttgatagcaa gagggtagga atatcattta 300
gcaggctact gcaagtatcc aggtgaatgt tacagaggtt ttgaactagg ctgctgggga 360
gggtgcagag aagaaatatt ttggaaataa aatggacaga aagtgtataa atggataaag 420
agagggaatag aactgaacac aggcctcaag cctgatgccc gagaaataaag gtgttaattat 480
gaagggaate cagggaagaca tggaaagagt ggttgagata aggttaaaag gatagtttta 540
gattgggtta ttttgacgtt gaagtgttga ccaacttctt aagtgaatac gtgcaacagt 600
cattgaaat atgagttt 618

```

```

<210> 23
<211> 711
<212> DNA
<213> Homo sapiens

```

```

<400> 23
gaaaaataag tttttgttaa tgggttggat ttctttaetg gcctcgtgac aagttttgtt 60
atctcttatt atatabatct taacttttta tgggaaaaac taactttttt ggtgtaaaag 120
tattttttat atttctcttg ctgttaaaga gttattatca atttgtaagt ataaaaactg 180
caggtatagt tggtagttga taagaaaggt agataataaa acttaaaagg gatggacaca 240
gattgaaaaa ggccttgagt gccaaagcaa gacgtctgaa ctttaacagg cactggaaac 300
cgtatagagt cttaggtagg aatatcgtgt gctccaccca tcttaattag gtcttatgga 360
ggtttgtagt caagagggtg ggaatatact ttacgaggct actgcaagta tccaggtgaa 420
atgtcacagag gttttgaact aggcctcgtg ggagggtgca gagaagaat attttggaaa 480
taaaatggac agaagtgta taaatggata aagagaggaa tagaactgac accaggttcc 540
aagcctgatg cctgagataa aaggtgtaat tatgaaggga atccagggaag acatgggaaag 600
agtgggttga gtaagggtta agtgatagtt ttgatttggg ttattttgac gttgaagtg 660
tgaccaactt cttaagtga aatgtgcaac agtcattgaa aatatgagtt t 711

```

```

<210> 24
<211> 547
<212> DNA
<213> Homo sapiens

```

```

<400> 24
aaccaaggtaa gcatagccgg ttttcattgg cttattttct catggaaatg attctgtgta 60
gaattgatta ttcatgaaga cacaaatgaa catcaagttt ggggttaagt tctcagtgcc 120
aacacaagag acgtatttgt aatcactccc atgagttctac ttgacagcaa gaacatgcat 180
tttggaaatta ttcccatcct gtgtcctgaa tactgtatgt gactcttagt cagctctgtg 240
acctttgtca agtaacttaa gctctttgat catcagcttt gtcatctgta aaatgggcat 300
ctgctact tcaaaagaga gttgaaggga ttaaaagaga taacctacaa agagcaccac 360
gcacaatgac ctaaaaaagg aaggcaatga atcattctca ctccctacc ttcagttctga 420
tctgtctctt attgtcaaaa ggataaattc aattttaata gatctgagat cctgtttttt 480
aataataatt ttatagaatt ttccatttta tggcaggcac agggctcatg cctgtaatcc 540
cagcact 547

```

WO 02/48370

PCT/US01/51341

12

<210> 25
 <211> 549
 <212> DNA
 <213> Homo sapiens

<400> 25
 gcaaaagacot catgagggtt caacgagggg aagccctcgt gggtcagagt acgccacggg 60
 acagactatg ctggcagcct ctagatcgtt gaactctgtt cttgaagact gggcagaatc 120
 taggaagaac ggaaggcact gaggtaacca ggtgggacga acctggcctt agcacggcat 180
 gtggcattta ggtgcttaag ttgtttgttt tttttaatt aaagtgggtg acctggagag 240
 ctgggtgtgga aatgtagcag gagggtctatt tggaaagaag gatggagtag attatgaag 300
 ttcttaataa tcataatgag gcttctggat ttatctctgt ggtttggatg ctctctctct 360
 ccattcccttg gatgcaaca ggcattgcact gtttaatttt ggaattcaaa cgggtggcctc 420
 aaacagtgag gctgagtagt tggcctcatt agcttcagac ccagcagggc tgggctcaca 480
 ggcgtgtcat ttatcaaggg ctggaatctc tgcacagtaa tttatctaag acaactctat 540
 gagatgggg 549

<210> 26
 <211> 350
 <212> DNA
 <213> Homo sapiens

<400> 26
 ctttaagata gatgggtaca catattatga atatacttct cttttgccag acctgacat 60
 tctgtagact tttaaggaa tattatttgc ctctttcttc ttactctgac gtatgggtg 120
 gatgggttac gtgcaggata atgtatgaac ctcccaagc tctgtacaaa tataacttgt 180
 cttctgtaga gacgtatgta ttatatactg tgcattcagt cttatttcta gattttcttc 240
 ccatttgctt aatactgac gctatggct agatgtgaaa tttaaccagt actactcata 300
 gcaggagtg aaacctgga ctacgtgct cttctctct ttctctccca 350

<210> 27
 <211> 627
 <212> DNA
 <213> Homo sapiens

<400> 27
 ccacgcgtcc ggtttcaaaa aagaagagta agtcaagggt taaacttttg gggcggagga 60
 aaaaggataa gaagaggat acagagttaa atcagagttg gcattcagata gactaaccat 120
 ggacattttg aagctgtaac ctctctcata ttctgccaaag gataactgct tctgtatta 180
 tcatgtaatg agttttatgc gtgatggaaa atgtaaaagt aactttaacc caaacctgca 240
 tttaaatgcc acatggaccg gctgtaattt atggcattct taagatagat gggtaacacat 300
 attatgaata taactttcct ttgccagacc ttgacattct gtacactttt aatggaatat 360
 tatttgcctc ttcatctta ccttcagcta tgaggtggat ggcttacgtg cagggtaatg 420
 tatgaacott cccaagctct gtacaaatat aacttgctat tcttagagac gtatgtattt 480
 atatgtgtgc atgaggtctt atttctagat ttctctccca tttgttaaat actgaacgct 540
 atggcctaga tgtgaaattt accaggtact actcatagca ggcagtgaat ccgtggactc 600
 agctgctctt tctctctttc ctcccca 627

<210> 28
 <211> 548
 <212> DNA
 <213> Homo sapiens

<220>

WO 02/48370

PCT/US01/51341

13

<221> unsure
 <222> (132) .. (348)
 <223> a, c, g or t

<400> 28
 gttgcabtg ttggggatat ttotccatta gcaagaagtt tccaaacctt accagtgitt 60
 tgaatgaatc aggaacagat ctggcagtg gacctacac cattttcccc acggaacgca 120
 tcttgctgga gtnnnnnnnn mnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
 atgtgatag aaaaatgggt tgatggcag atatatccag attgtagatt tcataatatt 420
 aaeggggagt gggcaaataa taaaatgcaa gaaatgaaag catttgaaaa ttttagaggac 480
 agaatgact ttaagtaag tgattttagg tgtactggaa tgaatcatct agaattttg 540
 atctgaga 548

<210> 29
 <211> 988
 <212> DNA
 <213> Homo sapiens

<400> 29
 aaatccacaa ataataattt acattttgaga aaatccccca gtactttctat gaataagatc 60
 aaggccaana gtgtgctctt ttacatgcca gaaacctcaa gaatttttcg taaggtacag 120
 ttragggaana accaagcagc tcttgactca acaataaaaa atgtaagtct gbtgagaaa 180
 ttatggaacc agggcaccoca gtccagctttc tcttaaaaaa aatttggaga gctgaagat 240
 atggatgggg tccagatttct aaaaaatcag tatacacaca gtgttttaag aataaaaaac 300
 agattgacta aagggaaaaa taatttgtaa ataacacaga ccataactta gagataaaaa 360
 taactgtcct ctgattaaca gaacttttag aatgatgaga aaaaataata aacacgttaa 420
 agatctcaca gtgattttta aaaaatttc aaggttgag aaaaatatt cctatgagaa 480
 taccaggtga aaaaagatcaa agtaaaatga atcaggtcgg tatcagaaat ttcatgtata 540
 taccaggtga gaataaaatg gagcagcagc tatagttttg aaacaaaatg tattttccaa 600
 gggtttctga cccaacocaa ttataactta tgtgttagga caatagagaa gtaattttag 660
 ccaagaanaa atctggaat tatagcatct atgcacattt attgaaacaa gaaactdaga 720
 aatcaaaaaa gcagagaat taataaaata ttcaaaagga ggaaataaca ttttagaata 780
 aagcataatg aggaataaaa tcaactatgac tttttgaaag tataaaaatt gttattttt 840
 tctatgaata cttgctcaaa ttttaagtag tggatttaat gttgtagcgc taagtattoa 900
 gccaagaggt agaactaata aataaaatg atagttcttt taaaaaaaca taaaaaatat 960
 tatctatga gtacgctaag aaaaaagc 988

<210> 30
 <211> 651
 <212> DNA
 <213> Homo sapiens

<400> 30
 acaccaata aggtaatgga gataaacttt agaaatcatg tttttaact gatgtttaaa 60
 agggatggaa ctacactat ttaaaagggt aagactgcaa cgtcagtggt aaattgttta 120
 aaaaagttca acacatttgg ggtggacac accagtcaaa tgggtgaat tagaagatgg 180
 ggaataaata tgtcaggtta atacttttatt tcatgtgatt tatgacttcc cctgaagaa 240
 gcattattat ttatataaaa tacccaaaaa aaaaaaacaa caaaggcagc taattctga 300
 aatttaattgc atctgcataa tgatttcaaa tatatttaac tctgaaaaaa gtgcgttaaa 360
 atggttaagc acataaatca aaataaagtt tgtatagcaa tatcaatc acataaata 420
 taatttagaa caaaaaagca cttataggga taagagaaa caacagagaa aaacaaagaa 480
 aaaaatctaa gaaaatataa ctttcaataa cttatattgt tttaacagcaa agcccgtaga 540
 ctgtttaata taggaagcgc aaacgtgact gaagttacaa gagactgaga caactttcaa 600
 aactcatggg gggagaattt tatcacttca acagaaactt acaatttaa c 651

WO 02/48370

PCT/US01/51341

14

<210> 31
 <211> 553
 <212> DNA
 <213> Homo sapiens

<400> 31
 actggacttc ctctttcttc catcaaaagac taagatggcct tttttcccttg atgtacttta 60
 ttttggggag catattatct accttccctga aaaaatgggt tatgggagat aaatcataaa 120
 aagggttttat tagattctac atctcatgat tgaaccaaaa gacgttttaa aaacaaaaaa 180
 aaaaaaggcc ttgtagggtct taactcttac ttagccctac atttatttga tagtttgagt 240
 gagtatctta aaaaattgaag atgattataa aaattttaat gtagacatta tttttcttaa 300
 gaattttgaa ggcaactgctc tgccttttgc agttggagag tctgatgaca ttctgattct 360
 taaatctttt atacaaaaaa cgtttttgct tttggcagga agctttacct ttctttctct 420
 tcaagtgctc tgaaccttca ctgagatgta tcatggtata ggtccacttt gatccactgt 480
 cctggacact tgcaggccct tticagtctc gaagctcatg accttcaggc aagagaaatt 540
 tacytctaag acc 553

<210> 32
 <211> 2159
 <212> DNA
 <213> Homo sapiens

<400> 32
 ggcccgcttaa ttaaagatct tttttttttt tttttttttg tgcgaataa tagtccattg 60
 tctttatgta ccacagttta tccactcacc taactgaagga catcttagtt gcttcaatgt 120
 tttaggggtt acagataatg ctactataaa catccatggt cagggttttg tctgaatgta 180
 aagtttccaa ttcatcttgag taaataccaa agcatgcgat tgcacatca tataaagta 240
 tggtttggtac tataaagaac tgcctaaactg tctctttaag tggctatgca tttttcact 300
 tccaccagca ataaatggag ttctgttgc tcccatgct cactagcatt tgggtgtgtc 360
 agtgtttctg attttgttca ttctaatag taactagta tatctggtt ttttaattta 420
 caattcccta atgacaratg atgttgaaac tcttctcata tgcctatttg ccatcgttat 480
 atctactttg gtgaggtatc tgttcagatc ttttgcctt ttttttctt tgagacagag 540
 tctcactctt gtacccagg ctggaagtga gtggccgat ctaagctac tgcacactct 600
 gctgtctggg ttcaagcaat tcttctgct cagccctcca agtagctggg attacaggca 660
 cccaccacca cggccaggta atttttatat ttttcataga gatgggttt cggcatattg 720
 gccaggctgg tctcaactc ctgaactcag gtgatccac tgcctcagcc tccgaagtg 780
 ctgggattac agggctgaga caccacaccc ggctcttggt caactaatc ttttttatt 840
 gagatggagt ctgtctctgt tgcacaggct ggagtgcagt ggcctgatc cggctcacc 900
 caactctcgc ctctcaggct caagagattc ttgtgctca gccctccagg tagctgggac 960
 tgtgcacac catgtcgggc taatgtttgt atttttagta gagtggggg ttcacttagc 1020
 oaggctggtc ccgaactctt gcccycaaaa gatctgccc cctcggttc tcaagtgcc 1080
 ttggatcccc aaagtgtctg gattacaggt gtgaaccatc atgactggca aagcatatgc 1140
 ttttgaagcc catgtctctt cctaatttgt tgaatcata ctacatgagt atcttcaaac 1200
 actgagcaac tacyaaattt ttgtgaaat gccagtga atactaataa gtatbtatb 1260
 tccaggtaaa atgagacacg ggttttttaa agtcaatgaa tgtgcatgga agtatatttg 1320
 agactcacta aggaatatga ggcacacga ctctctgtaa tttttagtaa aagactccta 1380
 tctgaggga tctgggattc ccccaaaaag gatctcagtt tgaatccct acagtgaagg 1440
 tcaacagtc ctacccaaga attcaaaaac cctgtcagtc tttagttccc tagtcttgaa 1500
 gttttgagcag agtccatcat taccagagaa ttogaggata gtatctcoga gaagcgggga 1560
 aaaaactcag ttaagagaga agggatgctt taaaaaaaa aaagaggtc tttagagtaa 1620
 attttcttta cctgaaagtc atgagcttcc agactgaaa ggcctagcaa gtgtccagg 1680
 cagtgagatca aagtgagcct ataccatgat acatctcagt gaagtctcag gacacttgaa 1740
 aaaaagaaaa ggtcagctt cctgcaaaa gcaaaaacat gttttgtata aaagatttaa 1800
 gaatcagaat ggaatcagac tctccaaatg caaaagacag agcagtgctc tcaaaattct 1860
 gagaaaaaat aatgtctaca tcaaaattt tataatctc ttcactttt aagatactca 1920
 ctcaactat caaataaatg tggagtaag taagagttaa gacctacaag gcttttttt 1980

WO 02/48370

PCT/US01/51341

15

```

gttttgtttt taaaaagttt ttggatcaa tcatgagatg tagaatctaa taaaaccttt 2040
ttatgattta tctcccataa accatttttt caggaaagta gataatatgc tccacaaaat 2100
aaagtacatc aaggaaaaaa ggcatcttag tctttgatgg aagaagagag aagtcagat 2159

```

```

<210> 33
<211> 450
<212> DNA
<213> Homo sapiens

```

```

<400> 33
agaaaaaag atccagatag aaaaatcgat tgtattttta ctatgcta atattagcaga 60
tattgaaact ttttaaacat acaattttat atagcatcag aaaaatggaa tgccttaagta 120
taaatctgac aaaaaatgtg agctacctgt acactggacc actaaacct agtgaacaa 180
aatgaagag ctacttaatt ggaatbcagt ttccccccag atttatctat agagtccagt 240
aatcccaat caaaatctca gcaaggctct taagaatttg acaatcttat ttaaaatttt 300
aagtgagagat gcgaaataac taagcaatt ctctgacaaa aacaagaaaa aagctagaag 360
gctaacaccc acactgattg caagatttat cagaacaggt ataataatca ggcagctgtc 420
atatcgagat acacagataga ccaggagatc 450

```

```

<210> 34
<211> 584
<212> DNA
<213> Homo sapiens

```

```

<400> 34
ctagacttat ggatttgagg gagctgtgtg aaactcatca tggcaaatat gottatgtgt 60
atatatcctt tgcacatcat gtgctgcaaa ctgtaaatga atgttatbta taagactgggt 120
aaggcatgtg ttattagact ggaacacaaa aagcccttga ttatctagga agcaatcctc 180
tagggtccag atgtgttttg gaatgtgggt gtttagtato actgtacttc attactgatt 240
tttatttcta tgcgttttga ctgttttagc tctttgttat tattggggag gttagccagag 300
gtccccagat tcccataatg aatttacagg tctgatotta tggacaagga ggaatcagct 360
gtattagttg ggggttcaat ctgtcctgat aagcttttcc tagttggttt taacagatcc 420
agccctgac taccctctgc tgcacctgtc tgtttcttat atgcatgtca ccatgatata 480
tgagtatgta tgaataata tttaggctaa ttttaactag aatatggaaa ggaaaaagtt 540
ctattgtctt gcattgtctt gttttcagca atcactgttt ttca 584

```

```

<210> 35
<211> 642
<212> DNA
<213> Homo sapiens

```

```

<400> 35
gctagactta tggatttgag ggagctgtgt gaaactcatc atggcaataa tgcattatgt 60
tatatatcct ttgccataca tgtgctgcaa actgtaatga aatgttatct ataatgctgg 120
taaggcatgt gtlattagac tggacacaca aagcccttg attatctagg aagcaatcct 180
ctagggctca gatgtattt ggaatgtggg tgttttagtat cactgtactt cactactgat 240
ttttattctt atgtgttttg actgtattag ctotttgcta ttattgggga ggttagccaga 300
ggctccaga ttccataat gaatttcag gtgtgatctt atggacaagg aggaatcagc 360
tgtattagtt ggggttcaa tcttgctga taagdttttc ctagtgtgtt ttaacagatac 420
gagccctgat ctactccctg ctgccactgt ctgtttctat gatgcattgc accatgatat 480
ctgagtatgt atgaaaatat atttagcta attttaacta gaatatggaa aggaaaaagt 540
tctattgtct tgcatttgc ctgttttcaa gcaatcactt tttttcacc acatatagaa 600
agtttgaaag ctctctctga tgtctggcaa ccagatctcc ca 642

```

```

<210> 36

```


WO 02/48370

PCT/US01/51341

16

<211> 669
 <212> DNA
 <213> Homo sapiens

<400> 36
 ccaaaattta ctgaatgtc ctgaaccaca tcttccataa tgttgctgac tcaagactc 60
 ttgaaggctc ctgaccacat tattcgcaat tctaacctcc ttgccacccc tcccccatga 120
 cccatgtaca attacatgct ctgacatctc tctccaaaga tgaacataag tctgcaaatat 180
 caacaccttg gcagccctat taccaattgc tgatctgtag tccccatgta agtacgacct 240
 ttttagcaac cagtttttgt cccagccata ttaatacttg tggctcagtg ttaacaattgt 300
 tgaagcttaa attatattca gatgaattt taaaaggaa tcaattgtgt tctctgtgt 360
 taacacagga atccagcat gtgtttctct tccaggaaac cataattata tgcacaata 420
 tctacccgga caattaggga cataatcatg ctctaaatag aagtggtcaa acaagtcac 480
 acctctctct cagttattcc tcttctctct tctctctaga tgcctaggtt tctgtgtctc 540
 aagacattta tgatttgatt tctctaaccc tttctaggtt ctattagagt caattagaca 600
 acatattctt tctttctaag aatctggaca aggaggtata cttttctaaa ttttaattct 660
 attaatgcc 669

<210> 37
 <211> 1006
 <212> DNA
 <213> Homo sapiens

<400> 37
 tcttaaaatg agcaccccca ggaactgttag gtaggagagg tgttagattt caagtagata 60
 caaataggctc cagaaggtaa aatgaggacc caaggataga agaggcacag tgatttcagc 120
 tgaagctctag tcccaagcac agaacttttc agaaacagaa tgggttgcat aatattgtcc 180
 ctttcaaaag acacttttga gacctggatg cctgtgtgtt ggcattggac atagaggctt 240
 cctgtccctg gtaaacatgc tgtgtgggac taggttctct ctgaaagtct ctccctgctt 300
 caggagctga gaattctaa gttctctcca ggaactctca aaattacta gaattgtctg 360
 aaccacatct tctataatgt tgcctgaacta aagactcttg aaggctcctg accacattat 420
 tgcctattct aactctcttg cccccccttc cccatgacc atgtacacat accatgtcta 480
 gatcttctcc tcaagatga acataagctc gaatatcaa cactctgcca gccctattat 540
 caattgtctg tctgttagtc ccatgtaagt acgctttttc tagcaaccag tttttgtccc 600
 agccatatta atacttgttg tcaagtgtta ccaatgttga agcttaactt atatccagat 660
 gaatttttaa aaaggaaatc cttgtgttcc totgtgttaa cccaggaaac ccagcatgtg 720
 ttctctctcc aggaaccat aattatatgt acaaatatct acccgacaaa ttaggggcat 780
 aatcatgtct taatagaag tgttcaaca agtcaacacc ttctctccag ttattctctc 840
 tctctctctc tcttagatgt catgtgttct gtgtctcaag acatttatga ttgtattttt 900
 ctaccccttt ctaggttcta tttagatcaa tttagcaaca tattctctct tcttaagaat 960
 ctggacaagg aggtataact tcttaaatc taatcctatt aatgcc 1006

<210> 38
 <211> 589
 <212> DNA
 <213> Homo sapiens

<400> 38
 aggaagctggg ttttgcttaa cagaaggagc actgacccat gttatagaaa atcgacgaat 60
 ttcatactcc catctataaa atgaaaacac aatacttctc accaacactt atacagcacc 120
 tactatgtgc taggttagag atcataaact ggtgatatgt aagtggaata taacccctag 180
 acttggtctg tgtgtctctc gcagttgac tgcaccagcc tttgttaaaa ttggaaggaa 240
 attgtcaata ttttaaaatc ggaattttcc cagcaaaatc taattttcta gtatctcaga 300
 aaaaatcata tttggcagca ctgggcccaga atttctgcag ggcattgtt gtctgaatt 360
 ggggtggctgg tggaaatggg cgtgtactcc taagtttgtc caaattgoba cggctgatt 420
 acttcatctt ttaattgtca ctactcttg cctgtggga tttttgaggc tgagattctt 480
 atattagggt ctgaaggcaa aacacacaca gaaaagaatg atttcaggcc ctctctgagc 540

WO 02/48370

PCT/US01/51341

17

ctactcatga tgtataactt ttatgacagt aatagtagta ttagcaat 589

<210> 39
 <211> 528
 <212> DNA
 <213> Homo sapiens

<400> 39
 aagacccgtgtc tttattttta gaagtaagaa taaaagagat tgtggtggag tatcacaggc 60
 agcggggggag cactgaggga gccctgacc caccctagga gtggatcagg atgactttctg 120
 aaaggccaaa ctgattaata agggataaat aaagtcatgc aaatgaaaag gttgtataig 180
 ttttggggga aagcattcca gacagaagga ccagtggtgtg caaaggccct ggggtgagag 240
 gtgcotaatc agtaotgaat atacaagag gttagagctgg gactaaacca ctgtgctcac 300
 ttgtccctgtc tgaattccga ttccaaggag tggaaatagac ttcaaatgtc tccaagtcca 360
 ctgttttctg ccaagttctc atttttgttc catgaaggca gacaccttc ttattttcat 420
 ccactgatga ctctcagcc cttagaatc tgccttatga tggattttctc agaaatatgt 480
 ttgtgtaatg aagacaagga cagtggtag agttacatt ctactggg 528

<210> 40
 <211> 673
 <212> DNA
 <213> Homo sapiens

<400> 40
 caaaaaatna aaacaaaac attagttggg cgtggtagtg tgtcccaggt actcaggaag 60
 ctgagggtggg aggatgtctt gagtcccgga gttggatgct gcagtgagct atgattgtgc 120
 cactgcagcc tgggtgacag aacaagaccc tctcttaaa aacaagaagt aagaataaaa 180
 gagattgttg tgggtatca taggcagcgt gggagcactg agggagccc tgacccaccc 240
 tagggatgga tcaagatgac ttctgaaagg ccaactgat taataagga taataaagt 300
 catgcaaaag aaaaggttgt atatgtgttg ggggaaagca ttccagacag aaggacagt 360
 gtgtgcaag gccctgggt gagaggtgc taatcagta tgaatatca aagggtaga 420
 gctgggacta aaccactgt ctcaatttgc ctgcttgat tccgattcca aggagtgga 480
 tagacttcaa atgtcttcaa gtccacttgt ttctgcaag ttctctattt tgttccatga 540
 aggcagagca cttcttttat ttcatcact gatgaattct cagcctctag aattctgct 600
 tatgatggat ttctcagaa tatgttgttg taatgaagc aaggacagt gtttagagttt 660
 acattctact ggg 673

<210> 41
 <211> 447
 <212> DNA
 <213> Homo sapiens

<400> 41
 ctcaagcagg gctagcacct ccaatctaga gcacccgtca cttccggctc caccggctct 60
 ctgtgccctt cactgccttg cctagggttg ccttctctc ctctcttaag ctgagtacaa 120
 gtgataatat agtgattaac acaatgctgt agtgtttcc tgttaaacag ggaatggttg 180
 attttccagg agaatagaat atgaattgt cattggagga cctcctcagt tgaatcatt 240
 ctgtggctga ttctctcta ttttgtttt tgttgggttg ttggttttg ctttttcagt 300
 agctaaccag gtatacaaat agcttctttg cagttctgat catctttagg ggcgcatttg 360
 ggcataattg gaataataat actagctaac ctgcttgacg ggttctctt gtgtgtgca 420
 ctttgtgagc acttaataa taggagc 447

<210> 42
 <211> 562
 <212> DNA

WO 02/48370

PCT/US01/51341

18

<213> Homo sapiens

```

<400> 42
ctcaagcagg gctagcacct ccaatctaga gcacccctga cttccggctc caccggctct 60
cttgctccctt cactgccttg cctagggggt ccttctcttc ctctcttaag ctgagtaaaa 120
gtgataatat agtgattaac acaatgctgt agtgctttcc tgttaaacag ggaatggttg 180
attttccagg agaataaaaa atgaattgtt ccttgaggga cctctctcag tgaatcattt 240
ctgtggctga tttctccta ttttggtttt tgttggtttg ttggtttttg ctttttcagt 300
agctaccctag gtatacaaat agcttctttg cagttctgat catcttttgg ggcgcagttg 360
ggcataattg gaataataat actagctaac ctgcttgcag ggtttgtctt gtgctgttga 420
ctttgtgagc actttaataa taggagccaa acctctcttt ccaaaagcct gaagggcagg 480
tgtctctgca gttccctatc catagatcac cactcttcca tggaaagtao tctgtggact 540
gttaacttgcc atctagaact tt

```

<210> 43

<211> 948

<212> DNA

<213> Homo sapiens

```

<400> 43
gggtctttct agctttcttg tcttttgtga agctggactg gtgatgtgca gttgaagaca 60
gcataatcgg ggcctctctg ctccatgtgt acctccagct atttgcaaaa gattgaacct 120
acaagatacg ttattagggt aagtatttac atggaaaggt cctgagttct ccaagacttt 180
ggctattttt tacaagatga tgaactaccc tgatgatttg tggaaatctt ttagggaaccg 240
tgactgtggtt gctttcttga tcatgggtac agggccatct ttgttgaggc ttcccatgtg 300
tgtgggcaca gagcttctgt ggcattccag cagtagatta atggagctgt catctcttga 360
agctcatcgg gttgtgcatg caaacctggt cctgtgaact gcctggaggt ctcttaaaag 420
ggcaggggga ttctctctct tgtgaaaggt ttagaatggc acatattttg aatttccaga 480
ctcatctttt cccactctca cttcaactct gtaattggcc gtactaaatt gttgacagtt 540
ctccaaatac aacagcattg ctattctgtc gcttttgtac atgcccgtta cattactgtc 600
acattgtcca ggaattcctc cctgccttga ctgcagtgc cctcttgaga gctcccggtg 660
ccctgtgctt gccgtgttca gagcttccag catgctgggc tgtggaggtg ttggtctgtt 720
tgccccccca gcaagcctct aagctcttca aagaacccaa ctgtccagca tatctggagc 780
agcacctggt acctaacggg tootaaatg cnggtggaat gaatgatgtc ttctgtctct 840
ttaaaccc

```

<210> 44

<211> 1111

<212> DNA

<213> Homo sapiens

```

<400> 44
gggtctttct agctttcttg tcttttgtga agctggactg gtgatgtgca gttgaagaca 60
gcataatcgg ggcctctctg ctccatgtgt acctccagct atttgcaaaa gattgaacct 120
acaagatacg ttattagggt aagtatttac atggaaaggt cctgagttct ccaagacttt 180
ggctattttt tacaagatga tgaactaccc tgatgatttg tggaaatctt ttagggaaccg 240
tgactgtggtt gctttcttga tcatgggtac agggccatct ttgttgaggc ttcccatgtg 300
tgtgggcaca gagcttctgt ggcattccag cagtagatta atggagctgt catctcttga 360
agctcatcgg gttgtgcatg caaacctggt cctgtgaact gcctggaggt ctcttaaaag 420
ggcaggggga ttctctctct tgtgaaaggt ttagaatggc acatattttg aatttccaga 480
ctcatctttt cccactctca cttcaactct gtaattggcc gtactaaatt gttgacagtt 540
ctccaaatac aacagcattg ctattctgtc gcttttgtac atgcccgtta cattactgtc 600
acattgtcca ggaattcctc cctgccttga ctgcagtgc cctcttgaga gctcccggtg 660
ccctgtgctt gccgtgttca gagcttccag catgctgggc tgtggaggtg ttggtctgtt 720
tgccccccca gcaagcctct aagctcttca aagaacccaa ctgtccagca tatctggagc 780
agcacctggt acctaacggg tootaaatg cnggtggaat gaatgatgtc ttctgtctct 840
ttaaacccac ctctcaactc gctaacataa tggatatttc ttctaactgg caattttaaa 900

```

WO 02/48370

PCT/US01/51341

19

```

gatcctgctg  tggccttttg  tcaggctttt  gagcagggtt  tggcaaatcc  gtggcctatg  960
gaaccaggctc  ggcgcgcggc  ctgatggtaa  tccttgcgct  ggccgtttca  ggtatgaatt  1020
acagttactg  aacccaattc  ctgtggaaaa  taataaaga  ctgcgggctt  tcaatcag  1080
tagcttaaaa  aggggaacag  gggacaaact  g  1111

```

<210> 45
 <211> 626
 <212> DNA
 <213> Homo sapiens

```

<400> 45
tatttctgaa  tcaacaggaa  aaatgggtaca  agaattttt  cagatcatgc  aaaaagcag  60
cacttcgtta  aaaggaaaga  aaaatttcaa  gtaaaacata  aacagggttt  tagattgctc  120
gataactcaa  ttatggaata  aaacaatgat  aaagctata  tatttcttgc  tgatttgcca  180
ggaaactag  acactgaaca  agalagcatt  acctaaagat  ataaaagcaa  agatagcgtt  240
gcaacagact  gcttaatgtg  tgcattctat  caaggggcta  tatgtgatga  gaagaaaaac  300
tgaatatgac  ctcaaatgtt  tcaactatca  gaaactgaaa  aaactcttac  tagtgtgttc  360
cgacataattg  tgagcaatat  tctaaagata  gacgtttctt  cagttatgat  ttcttgagg  420
ctacatcaga  gaacttcott  aaactgtgc  gtaatacaaa  atcagtgagt  catggcaag  480
gggagacatt  atctatctgt  tottgactat  gnaaaataat  gttgcagat  ctttgtcttg  540
tgtgtgaaga  agcagatgag  acaggaccag  aactgtccg  aagacgtatt  tcaggagacg  600
cactggcag  tcgggcgcgc  ctctag  626

```

<210> 46
 <211> 185
 <212> DNA
 <213> Homo sapiens

```

<400> 46
gaagaaactg  tgaggtcaca  atacttttga  ttcatatgt  gaataacat  acacactcac  60
atctctatta  ctgtatccat  ctctatata  ttgaactoca  tatgctatat  attaacttgc  120
cacaactcaa  ccaacaaaac  agggttcata  ttgattttt  cccccatat  ttatgattct  180
cagac  185

```

<210> 47
 <211> 268
 <212> DNA
 <213> Homo sapiens

```

<400> 47
atggatttgc  cacaagctgg  ctttgaaagc  agtggtagag  tgtgaagaa  gttaccttaa  60
gactttcttg  cagttgcact  gtaggtagca  tgtactgttt  gttgtgattt  gactttcttc  120
caccacccc  ctgcccacgg  aagatgtgat  cttgtgcac  ttgtgttcaa  gcagagtagg  180
tagttggat  ctttgtcaag  tctcagtgat  ccacatgcgt  gcattctatt  tgtcagtcgt  240
cttgtctttg  tatccatgtc  atactgtc  268

```

<210> 48
 <211> 108
 <212> DNA
 <213> Homo sapiens

```

<400> 48
gtcgaagacg  acagcaatgc  cgatccgcgt  caagcccgca  acgggctgcg  gctgcaggtg  60
atgctctgcc  tgcgcgaggc  ctcccgag  gcgcgcgtgg  cgtggcc  108

```

WO 02/48370

PCT/US01/51341

20

```

<210> 49
<211> 63
<212> DNA
<213> Homo sapiens

<400> 49
gacgcagatc ggccggctgc cgcctggtgca tctgcccgcc gaggcggctgc gcgcgccctg 60
gocgcctcgac gagcgcgagg tgc                                     83

<210> 50
<211> 475
<212> DNA
<213> Homo sapiens

<400> 50
aaagaaacaa gcaacaaata ggaanaatcaa attttttagaa gtaggctgat aataggggaa 60
tagottaagg ggagaaactat gatgttaatt ctttgaaagt gagtaatgta attagaacaa 120
taaacactatg agtttttcta taaacaaaat atagcaagat taagttgata acatacattt 180
ctaaaatttt ggcttcccta gagaaagcca accaaatata aaattttaga gcagagtcac 240
gttttttcag ttggccctat attttctttg gtaacactgt tctgaatgta tatgcagtg 300
ttatttcaca acttccctct gaatgacctt tcaaaaatta atgattcttc acattcatga 360
ccagatgttt tctctgatgg aagcatctga tgtttgcagt catcaaatat gattcaaaat 420
gtctgtttca agcaaatcaa gtaaaacttc tccatcacat caaagtaag gctttg 475

<210> 51
<211> 607
<212> DNA
<213> Homo sapiens

<400> 51
aaagaaacaa gcaacaaata ggaanaatcaa attttttagaa gtaggctgat aataggggaa 60
tagottaagg ggagaaactat gatgttaatt ctttgaaagt gagtaatgta attagaacaa 120
taaacactatg agtttttcta taaacaaaat atagcaagat taagttgata acatacattt 180
ctaaaatttt ggcttcccta gagaaagcca accaaatata aaattttaga gcagagtcac 240
gttttttcag ttggccctat attttctttg gtaacactgt tctgaatgta tatgcagtg 300
ttatttcaca acttccctct gaatgacctt tcaaaaatta atgattcttc acattcatga 360
ccagatgttt tctctgatgg aagcatctga tgtttgcagt catcaaatat gattcaaaat 420
gtctgtttca agcaaatcaa gtaaaacttc tccatcacat caaagtaag gcttttatg 480
gttcacaagt agctatatga aataaacaga atttaaacga tottaataat ttttttcttt 540
aaacaaagggtg acaaaataac aatgcacata tataaaact cttcattaat gataagtgct 600
agatgga                                     607

<210> 52
<211> 590
<212> DNA
<213> Homo sapiens

<400> 52
ctccbcatta atgataaatt ctatagtgac accatgtaaa gtatggaaaa tgcctgtctg 60
aaoaaatgct ttgtctaaat tctctgaatt tttttttgtt ttctctcacc agttagcttt 120
gatgttttga tccaggtttt tagaaaaatt ctaggatctg ttgcctttgg accttagagc 180
ttcttggagc cacatgtcag taataaaacg ttttcttaag cctctgcttt ccatagcaaa 240
aacatgttat gtccattatc caactaaact atacttaaaa acaaacccca agatgctcta 300
ttttgttttc aaagtccagag aagaaaaatag aggggaagta tttttatggt cttttccctg 360
aatttggtcga agctagttag ttcaaaaaag atacaaata tggaaatcca cctattttat 420

```

WO 02/48370

PCT/US01/51341

21

```

ttcctggcaa ctgtttcatt caaatcatag agtaacatat gatttactac actcctttat 480
gaatatataat ctctgtatctt cadagaatga cttatatatca ttgatcagct agaaccatga 540
cctcaaccgt ctgtgtgtttt taacgaatg tttatctcta gtcaaacccac 590

```

<210> 53
 <211> 217
 <212> DNA
 <213> Homo sapiens

```

<400> 53
agtctgctaa ctcattccag tggtttttct caactgcato tcagttatct tacatagact 60
gcaagaagtg agaaagacaa gaggttatct agtccagcct tgcctatttta tagtttaaat 120
cctccacca catccctgat gaacttttgc cagtgcgggt aattaaccaat atcccaagge 180
tgtcttgatt gtctgtatct ctctagtgtt gtttagag 217

```

<210> 54
 <211> 430
 <212> DNA
 <213> Homo sapiens

```

<400> 54
aataaagata agaatgacaa cagattttct ttgggaaca atgagagtgg gaagacaatg 60
agcaacatct ttaagttact gaaaggtatc agcagaccga tgcctacaaa aatgtaaaag 120
aacatcacca ggcagaagga aaaaaatagt atcagattga agtctgttct acacaaagta 180
atgaatacca gaaatgataa ctacctgggt aaatatataa gattatttct ttctttatta 240
aagtaagagt gagattctta tcaacaatag cataaaggct gaaggggaga aatggagtc 300
tattagtgtg atcttatata tgatgttgta tgatgtcact tgaatgtaga attataaaga 360
taaacagcat aaactcttaa agcaaccacc aaataacaa agagttataa ctaataattc 420
agcaaaaggag 430

```

<210> 55
 <211> 2956
 <212> DNA
 <213> Homo sapiens

```

<400> 55
gtgtgtgttg ttttttttga gacagagtct tctctctctg tctaggctag agtgcaagtg 60
cgccacatcg gctcaactga acctccacct cctgggttca agtgatttct ctgcctcagc 120
ctcccgagta gctgggttta cagggtctcg ccaccacgac cggctaattt ttgtttctctt 180
agtagggttt cactgtgttg gccaggctgg tctcgaactg ctgacctctg gatctgcaca 240
ccttggcctc ccaaaagtgt gagattacag gctgagacca ctgcaactgg ctttttattt 300
ttttaacttt gtatacggta ttttcttttt ctgtatagaa gtcaaaactat ttctcttcat 360
ggatctcgtt ttttctctct tcatttccag acaattttaa aaaaatgtgt cacttttctc 420
tctgatactt ttaaggtgtc tttctgaaga taaaacctga tgtgtctgca atgctagagt 480
gaggcttgag tatgggcaag ctctcctgag gcactgtgta gctgaggaca gcctggcgtg 540
tgaggaaagga tcagtccaca cagctcctgt aagctcaca gagaggctac tggcttccact 600
gcactgtctc actgggtgtt ttgacaacgt ggagtgaaata ctctcatgtcc tcacaaatcc 660
aaatgctggt tttatcatgt ataatatta tattggaaa aatataaatc ataatgaagt 720
tattgtctca cttatcttga agaaaaaac ataatgttg cacttctgaa ttaccttaa 780
cctgtttaat acctactgag aaagtctact attcagaatg cagaaaaagg tggaggaggt 840
ggttagggtc ctaaaagtca aactgggtcc ccgcagccca gagatcaaca ttatttaaaa 900
actcaccatg caaagctaat agagaacgaa ccatgtaac ctttttgaa tattacattt 960
tcaactcaaa gcttggcctc atcttccagt taactgtcta taaatgtcaa ctacgaagcc 1020
tttcagaggt cctacacttt gcaaatgaag taagtggaac ctctcctgac acagacagag 1080
cccaaaaggc aggagtgcaag ctggcagtcg agcccttggt ggggccaagg ggcaggtcac 1140
atggaagggt cgggttctct cccatgtcaa taccgtgacc ctctactcat gctccagac 1200

```

WO 02/48370

PCT/US01/51341

22

```

ccctctggac accgtgctgc tggcagatgc tgtgctcctg ggagggtggga tgcgaagctga 1260
acotttgctca ctccctttgg gctaaatgac aggtgagcac tgggcacagc aaatgtgact 1320
ggccacagcc toatctgcag gggcaacaag ttcccaacac aagatccctg taccatocca 1380
cacaccocgt ctccatctct ctggatcctt gttcagacac agtgttttta tcaacaccca 1440
cagaggaaaa tgggtaaatg cgtggtctgg cgtggtctgg aaggttata atacatgcac tgcacacact 1500
agaatgtagc agaattcaca gctggtctgg aaggttata aaggttctct accgccaagt 1560
aacggttttg aatataata agcgtatctt taagtctctt aaggttctct accgccaagt 1620
sgaataaaga caccacacct ttttctcatg aggtctcaag tctcctctgg ataccgttct 1680
ataatcbgaa gtaattctagg aaacttcaat ctggtctcat tggaaattaa ttttaagctt 1740
ttcaacattt ttctcaccac aagtgtctga taogcttcca gttctgcaga atcaataact 1800
atcaaggaca ccaagaaga aagcaatggg naatgtatcc caatatccat aaactatgat 1860
gttaaatgct aacactttcc ctttttggct tgtatttttg agtgcatttg ttctctctct 1920
aactacacct ttacacacac aaacacacag taccgttttg tatctatctt ggagccaagt 1980
ccttcacatta ggtgctccat tctgcatgaa goacagtttg aatcctgggc tgggaacata 2040
aggggcaatt ggtggttatt gaatttatcc caggagcctg aagcaggcca caggagccag 2100
taactatgaa gctgcaagca aaatatacaa gtagaalta aaacaaatga aacagagac 2160
cacttgactc catttaaatg taggtcatgt tgcctagaga ggcacatgtc tctctctctt 2220
tttttttta agatggagt tctgtctgtc acccaggctg gtgtgcagta gtggatctcg 2280
gctcatgaca acctctgctt cctgggttca agcaattctc ctggccacag ctctgtagta 2340
gctggactca caggcatggg coacacagcc cagctaattt ttttgtatcc ttagtagaga 2400
tgggtgttca ccaagtggc caggctgctc cggaaactct gacctcaagt gatccactg 2460
ccttagcttc ccaaatgctt gggattacag gctgagacca cctcaactgg cctaatctca 2520
ttttatctcc ttgtctgaat tattagtat aactctttgt tattttggcg gttgttttaa 2580
gagtttatgg tgtttatott tataaattota cattoaatgt acatcatacc adaatatgta 2640
taagattaca ctaatagact tccattcttc ccttcacgcc tttatgctat tgttgataag 2700
aatctcactc ttaacttaaa taagaagaaa ataattttat atattttacc aggtagttaa 2760
catttttggtt attcattact ttgtgtagaa cagacttcaa tctgatacta ttttttctt 2820
tctgctgtat gatgttcttt tacatttttt gtacatggg tctgtgata ctttctagta 2880
ctttaaagat gttgctcatt gtcttccac tctcattgtt cccaaaaaga aatctgttgt 2940
cattcttato tttactt 2956

```

<210> 56

<211> 517

<212> DNA

<213> Homo sapiens

<400> 56

```

cctggctgga gccgacacgg tcaagaacct cctccctacc ttctcccttc aaccocagct 60
caactcaacc aaaaatggcc cctctgtccc catgctctgat aggaagtcga ggggaagtc 120
tgtctcgatta ctgtcaaaaga agcagaggag taagggtcag agtggaccac tgactgaata 180
tgagctgcag aagtgttaga ggcagaagtc cagggccatt tctttaatat cgaagtgtct 240
ctgctggagg ttgggtatgg atttttgccc tgcatttaga agttctgggg tctctggaga 300
ggggagagaa gcccaatagc agaggagaca gagtgtgggc ggggcagacc ggaaggggtg 360
atcctgggag agcaccaggg tgaggagaga gtgaagatga gcccgctcag ggaagggctg 420
gcgagtgttg gaagtcacct gcccctcggc ctgtgagctg ctctgcttgg agtgaactaag 480
gctcgggagg tccaggctcg gccagaggca gctcata 517

```

<210> 57

<211> 1490

<212> DNA

<213> Homo sapiens

<400> 57

```

ggggacccag acgcccagtc acaggcgaga gccctgggat gccccggcca gaggccatgc 60
tgctgtgtgt caggtttgac ctctctgggg gccccacctg ggcaggagat aagtcagtcg 120
ggtctgcccc caatctcccc tgcctccctc caggagagcc agggactcac ccggcccttg 180
tcccagacta actctggtca cagaacccat ctgtctgctt ggagggggcg ggtccccctg 240

```

WO 02/48370

PCT/US01/51341

23

```

tctggcagag gtcaccccca taccacccga tggggatttt cttccctttg ggtctctctt 300
ttcttcagag atgtatggcc ctggaggagg caagtatttc agcaccactg aagactacga 360
ccatgaatc acagggctgc ggtgtctctg aggtctcttc ctggtgaaaa ggtgagtagg 420
gctatgggta tggggccagc gccatgtccc ctccctcccc acagtttccg gaactcaggg 480
cagggggtaa gcaacggtag ccaattttgc caacacatgc tggctactgt cgtatgctcc 540
tggctccgcg tcatgcttcc tggctggagc ggacacggtc agacggctct ccttaacttc 600
tccctcaac ccaagctcaa ctcaaccaaa aatggccctt ctgtcccat gctgatagg 660
aaagtccagg gaaagtctgt ccgattactg tcaaaagaag caggaggtaa ggtccagagt 720
ggaccactga ctgaatctga ctgcagagag tgttagaggt asaagtcag gccattttcc 780
ttaatactga agtctctctg ctgcaggtct gggatggatt tttagctctg atttagaagt 840
tctgggggtcc tggggagagg gagagaagcc caataggcga ggagacagag tgtggggagg 900
gagagccgga ggggtgctc ctgggagagc acccgggtga gggaggggtg aagatgagcc 960
dogtccaggg agcgtctggc agtctgggaa gtcacctgcc cctggcctg tgagctgctc 1020
tgcttggagt gactaaggct cgggaggtcc aggtctggcc agaggcagct catatgtggg 1080
ccacagtgac ggcagctggg gcttctctgg taccggagac ctggcctgag aagcagctct 1140
cctcaccagg atctcagtg ctctctccaa aagtcacacc ccttttgtag acggggaaac 1200
tgagtcgga gaggctgggt aacgagctca agatcacagg gcccaaaagt ggtagaatca 1260
gggttggtag ccagtgagtc tgtgtcagag aoccaaagtc tgatggctgt ggactctctg 1320
catccgggga aggagatgg gggagctgag gaacgggag gtgctgggac atccagatc 1380
tggagctcca aagcttttgc tctctccag tgtccaggtg aaacttgag actcctggga 1440
cgtgaactg gggacottag gtgggaatc ccaggaagtc aacctgcagc 1490

```

<210> 58
 <211> 436
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (197)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (432)
 <223> a, c, g or t

```

<400> 58
ctctgtctcc tcataggaat ttcttagttt cttggcttcc gaatgtgact caaccctcc 60
cttggctctg ctgtctgctg tgtcgtttt aggtctctgt gccacggcta actatgtttc 120
cctgtgtttc cagataaact tctgagggtc agaagctgac agaccaaagt catthttcaa 180
gccaatctgt gtcatacaga gaccacgggt ttctcttggg ttgggtctct ctacctgttt 240
cagtcagctg tgaacaaaac ttgtggaatt tggctatttt ccttaaaatg gagatcagag 300
agatcaccat ggtctggctg aaactagttc tggatctgat tctctttcca attgtttgtc 360
catcagggtg aoccaactct gagggaattt tggtaacatt ttcccaaaa taaagatcat 420
taactaatta tnaaaa
436

```

<210> 59
 <211> 458
 <212> DNA
 <213> Homo sapiens

```

<400> 59
ctctgtctcc tcataggaat ttcttagttt cttggcttcc gaatgtgact caaccctcc 60
cttggctctg ctgtctgctg tgtcgtttt aggtctctgt gccacggcta actatgtttc 120
cctgtgtttc cagataaact tctgagggtc agaagctgac agaccaaagt catthttcaa 180
gccaatctgt gtcatacaga gaccacgggt ttctcttggg ttgggtctct ctacctgttt 240

```


WO 02/48370

PCT/US01/51341

24

```

cagtcagctg tgaacaaaaa ttgtggaatt tggtcatttt ccttaaaaatg gagatacagag 300
agatcacccat ggtcggcgtg aaactagttc tggatctgat tgtcttttca attgtttgc 360
catcaggtga accactctg aaggaattt tggtaacatt ttcccaaaaa taaagatcat 420
taattaatga taaaaaaa aaaaaaaat gagcggcc 456

```

```

<210> 60
<211> 359
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (336)
<223> a, c, g or t

```

```

<400> 60
cggcgcgtg ggaacacaaa actgcatcat ccaaaaatc acetttggtc cccggtgccc 60
actggaagac atctgaattt tagacctcca gagagaagat ctgggtggct agctccagag 120
tggagcctg ctgtcttttt cttaacactt gtgaagagga atggatccgg acatctgcaa 180
tctgggtaga ggaaggcagg cagcaagctt agccactcgg ccaggcttct cagccttacc 240
tctagaaatg tgatctttcc tcaactgat atacttcaaa actttcttcc ggctactcaa 300
ggcatcccaa gttaaaagga aggtcagatg tgattnatca ctttattatg ataaaaaa 359

```

```

<210> 61
<211> 932
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (161)..(180)
<223> a, c, g or t

```

```

<400> 61
tggccagaga catatgaaaa gatgcottag acatatagca totttttctc tccacttact 60
aggagaatag ctactaaaaa ttatccctga atgccattta aaaaaatctc agattgttga 120
agtacaaaaa gttagataac atattatcaa ccaaaatgtg nnnnnnnnnn nnnnnnnnnn 180
ttgggcacgc tgtgttttggg taaactagtt aagggtggtg ggtgttttgg tccaggtaata 240
aatctataag aaaaaaaaaa cctctgaatg gaaaactcat ggtgaggtga aaacttcacc 300
ccttctagtc acttatgttt aactggtcta ctggattttt ttaaaaggta agaaaaacaa 360
aactgcctca tccaaaaata cacttttggt ccacggatgc cactggaaga catctgaatt 420
ttagacctcc agagaagaaga tctgggtggc tagctccaga gtggaggcat gcttgctttt 480
tctttadact tgtgaagagg aatggatcgg gacatctgca atctgggtag aggaaggcag 540
gcagcaagct tagcactcgg gccaggcttc tccgctctta ctctagacat gtgatccttc 600
ctccactgta tatacttcac aactttctta cggctactca aggcattcca agttaaaagg 660
aaggctcagat gtgattctca ctttattatg ataaaaaaa ttactattta aatactataa 720
ataaatatta taataaatat taagctagaa ccatcagaat acatcacttc tgtatccagt 780
tttcaagta tctttggtgt ttgtcaggaa taatlaaaag taatcatttt attcttatta 840
aattatctct ggcactagtg gctagtactt ttgtacttat tagtacaacc ttaaaaagtc 900
ttaaaaagat ttcttttggg ttcaagaacat aa 932

```

```

<210> 62
<211> 554
<212> DNA
<213> Homo sapiens

```

WO 02/48370

PCT/US01/51341

25

```

<400> 62
ctggcagatc cggacgggca ggactgggtg tgtccatga gagcacctcc ttccctggcct 60
ttccctgtgga ctttgtccca caccacctgc ctgggttccct tcccttagtc acttccagct 120
ccaggccacag cagttgggtga ctccctgggtg ggagccgtgt cccacccggt cctgatactg 180
cogctcttctc ttccacagtc ctccaggttt gggccagcct tggggggaagc agagcttctg 240
gggtgagtggt cagatcctg tgccttgaga gggtagtca gggagagggc tggctggggc 300
agggtctgcc gggcaggaca caggatgggg ccggccaggg tggggccaaag gtgttcagac 360
ctggactttg ggtcgtgctt ttcttcacgg ttgcgccttg ctgcctgtcc ctggagctcc 420
tcatttgggt ttgctttttt tgtttgtttg ttttcaccta atttttgcca gacttaagct 480
agttttgctg ccttttgaaa ctagtggaa aatcatttta tttccctggg ataatttggg 540
ggcttttgaa tcca

```

```

<210> 63
<211> 786
<212> DNA
<213> Homo sapiens

```

```

<400> 63
ccagtgccct gbtccctagc aaatgagagc caccctgaaa aataaaatcc tgtctcccca 60
aagccagccc tggcaaggca cccagaactc tccggaatgc tgaagggaag ggcctggcct 120
ttccattgggg tccagggtctg tgggttccct ggccgtactg tgggcttcca gagcggggca 180
tgtgggttga agaccgtctc cccaccatgg tgggaaggga caaagggtgg ccttgacaga 240
tccggacggg caggactggg tgtgtcccat gagagcaact ccttctggc ctttccctgtg 300
gactttgtcc caacccaccc gcttgggttc ctcccttttg tcaacttcag ctccagggaac 360
agcagttggt gactccttgg tgggagcctg gtcccacccg gtccctgatac tgcgtcttcc 420
tctttcacag tccctccaggc tgggacagc cttgggggga gtagagcttc tggggtgagt 480
tccgaatcc tgtgtccgga gagcgctagt cagggagagg gctgttcggg gcaaggctgc 540
ccgggcaagg caacggatgc agccggccag gctggggcca agtgcttcag acctggaactt 600
tgggctcgtg ctttcttcac ggttgcgctc tgcctgctgt ccttggagat cttcatttgg 660
ttttgtcttt tttgtttgtt tgttttcacc taatttttgc cagacttaag ctgattttgc 720
tgccttttga aactagtggg agaaccattt tatttccctg ggataatttg ggggcttttg 780
aatcca
786

```

```

<210> 64
<211> 575
<212> DNA
<213> Homo sapiens

```

```

<220>
<231> unsure
<222> (411)
<223> a, c, g or t

```

```

<400> 64
ggcacagcta gttggtgact ccttggtggg agccgtgtcc caccgggtcc tgatactgcc 60
gtcttctctt tcacagtcct ccaggcttgg gccagccttg ggggcagcag agctttctgg 120
gctgacatgg ggtcatttgc tctttctctc aagccctctg agggacatca aaagcgtggg 180
acgcacccac ttttccacca tcttggcttg cccactgttt cctccatcc tggagggtcc 240
tccctaagca catgtgtggg ggtggggaag caccctggtt gatagctgtg gatggggcgg 300
tgacatcctt caeccctgcc cccatgggaat gcattgatcca ttaggaggga ccgtctgcac 360
aaaggtctct tgcctgttgc aagcttctct caagacttga cttgcaaaaag ntcagcctg 420
tatgtctgga gttcccatg cctgccaatc tctgttcgac tgcgagtcag ctccgatact 480
tcaccagatt cagccacctg ggggagctgg aagtgaatct cctcgtagct gagccttctg 540
atcagactgc agcccggtt gacacctgga ttgca
575

```

```

<210> 65

```

WO 02/48370

PCT/US01/51341

26

```

<211> 834
<212> DNA
<213> Homo sapiens

<400> 65
cagcagttgg tgactccttg gtgggagccg tgtccaccc ggtcctgata ctgcctctt 60
ctctttcaaa gtctctccagg cttgggcccag ccttgggggc agcagagctt ctgggctgac 120
atgggctcat tgcctcttct ccaagccctc tgaggacatc aaaaagctgg acgcatcaat 180
ttccacacac ttgctgcccac ctgtccctcc atcctgagggc ctccaaagca catgtgtggg 240
gtggcagggca cactgctgat agctgtggat ggggcctgga catccttcaac cctgtccccc 300
atggcagtgca tgatccattta gggaggaccg tctgcacaaa ggtctcttgc cctgtgcagc 360
ttcctgcaga ctggacttgc aaagtccagc ctgtatggct ggagttccca tgcctgccaa 420
tctcctgtcg aotggagtc agctccgata cttaacacga ttccagccac tgggggagct 480
ggagtggaat ctctctgtag ctgagccttc tgatgagact gcagcccccg ctgacacctg 540
gatgtcagca ctcatgaaa agcctgagca gcaggacag ttggcagag cccgaattcc 600
tgacccacag gaactgggag aaaaaactct gtggttttaa tcttctcatt tttagatgct 660
caggttccat gtggtgtgaa cagcttccat tcaactctgg cccttggggg agatgtctag 720
tggttcccg gctgtccca ctccacacca tggcagtgaa gagctgtga agtacctgct 780
tcattagtcct ttggtgtccc totatgagta cagttcctgt ttgtggagta gcaa 834

<210> 66
<211> 437
<212> DNA
<213> Homo sapiens

<400> 66
cgagaagaaa aaggtatagc ttaaagtggc ttttgagcag gcattgagtt atggaaacaa 60
ggattccctgt gaagacattt tcttttgata aaagaattt gataagaata ttataccaaa 120
ttgaacaaaa gtggccacag tatgaaggat toagtacatg gccaaataac ttattttcaa 180
atagtttaga gttatatccc ttgaagacgg aggttggatg gggattaatt tttgtaaga 240
cgccaatggc tgttaaaaaa aagagctgag atggtatgtc tcttgatta aaaataaaa 300
tatttttaatt atactattac atcataaaca ttctatgtct ctacttttcc atctagaagg 360
aagaattctt taqtacttcc cagacatcta ctgtgtgagc tatcttgtgt tatgacaaat 420
tgcttatatt tatttacc 437

<210> 67
<211> 80
<212> DNA
<213> Homo sapiens

<400> 67
acaaaaacat atgcttcaac acctcaggtt gaccttttgg ggggagctgt tatgggtgtt 60
ttaagatggc ggggtatgcc 80

<210> 68
<211> 663
<212> DNA
<213> Homo sapiens

<400> 68
gtgtagagca tgggaagcag gagaocagtt aggagctctat tgtaatagtc ctggtgagag 60
accacagcgg ctggactaa gatggcaact aagataatga tggttgcagg gccctcttcc 120
aatggaggca ttggcagcct tctggccatg aaggagaag tgatttcaac taaccacaga 180
aactcttacc tctaaatgga gatacttctt gataacagaa gaactctggg atctaaccca 240
gaaataccag ctgagtagga gaagagaana ggcacccgac agtccaggtt tcagaggct 300
gccaacagtc ttgttaagcc acottgggag tagatgagaa cggcaatcaa tcaacatggt 360

```

WO 02/48370

PCT/US01/51341

27

```

ttggtgaaca aaccatata tacaaagtgc ttctgtgaag ttctgcatct cacaactaat 420
gggtgagaca ttctctattg ttctgtctca cccaggaaata ccatgctgtg ccagctcttg 480
ccatttatta acaactgat aatgggtgcag tgcgttagtc atggaagcta ttctaaaagg 540
taaggaagt ctactggaat cctgggttctt ctagtgtcca tttagactta tttttaagt 600
ctcattgaaa tgtaatgcat gttatggaaa gtcaggatga aataaaattg agattttttt 660
ttt                                     663

```

```

<210> 69
<211> 695
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (309)..(482)
<223> a, c, g or t

```

```

<400> 69
gaaacacaga aagaggggag aaacaggaga ggggaagag agaggagaga gaaacaaagg 60
aatgtgaca tataataatt ttttaagaa tatttttcca tttttttatt gaggtatasa 120
ctacatgtag taagggtatg caataactca aactttatgt gatittttta tgcacatgta 180
taoctgtgta oaoctgtgta acaactaact aagtoaatg agagaacttt ttaatacatc 240
taaaagattt cctgtgtctc ttccacacaa taactgtgta tgagccact ctccttacag 300
ctatcgcnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnccttcact ttaatgaaca ttggaattgt ttctatgttc ttgttatgaa tcaaatggt 540
tatgaatagt ttggttatga agagtttttc acatgttttt agtctatttt gtttctctta 600
aatatatact tagtcacggg attactgtgc atatagtata ggcaggcaga tgttaagctt 660
taattgacac aaccaactgt ttttgaagg ggttg                                     695

```

```

<210> 70
<211> 739
<212> DNA
<213> Homo sapiens

```

```

<400> 70
gggtttctct catggacatt gtttgcatct acatgtgaca cttagggaatg atctgtttag 60
tttcaatcac tcaactcctg atctgcctgt ctctctctga gataacaaag gctttaatgt 120
ttagccacct gcatcagagt tggtaggtgt gtttgaacaa attcatccta atataaaag 180
aacagctttt gtaagggggc actgagtgte tcaaacagcc gcatgggag gaagagtgt 240
cagttccagtt ttggttgaa cttgtctgtt gccctaaggg ctctatgaa agactgacag 300
gcttggaactg aatcttgtga ttggtgaccc aagggtcacc tgtgggcca gagctagctc 360
tgaagaatgg ggtagtctt ttgagaacct ccacagcaaa agtttggtcc tctgttccaa 420
atgcatgtcc cactttacca gctacatccc ccagtacctg cccatggctc atgactcatg 480
aaatataaaa ctccagtaggc aggcataact ggttcagacc tggcagggtc atgtgggaac 540
tatcattggt acaaaaactc taagtgtgga gaagactgtg gttagacaaga ggggacatgt 600
ctgtttctaaa cgcacatcag aaactcccaa tgactatggc caagttagat aagggtgtac 660
agaactcttc aggaatgca gacctatgtg tcaactataa ctgaattcca aataaatatt 720
ttgtggattt cctgtggac                                     739

```

```

<210> 71
<211> 9883
<212> DNA
<213> Homo sapiens

```

WO 02/48370

PCT/US01/51341

28

<220>
 <221> unanure
 <222> (7153)
 <223> a, c, g or t

<400> 71
 ataagaataaa aaattacccc aaatttccaa atoaagaagt aatcatgggt cagggtttggg 60
 cagatgtttct ttctaggcat gaacacacgt tatctcattg tttacttaac aocgggttat 120
 aaacattttac ccatagcatt tgaagaggtag ctatagatag aaaagaatca gagaagttct 180
 aaacacagctc ttgcggtttg tttcaaatte ttgcagagaa agatggagtc ttacgctttt 240
 tttttagctg gacggcaccg ttgcagcagt ggtgaacagg gcatggattt gactcaggaa 300
 acccagctgt gaccttgggc aagccacttg cctcttttga gcttccctcc tgcataaggca 360
 agggggcgcta ttccgtaacct gtctgcccac ctcaacaggct ctggtagaag ccttgatttg 420
 aacgctttta gctcccaagg ttgtggtttg gagatagggt aggtcacatg aacatgaaga 480
 ctgaaggaga aacgtggaag cactgtgtgc tgttgcctct tttccaaact aaatgtcttg 540
 gtgatctcct gaagactcca gctcctcttc tgggaaagca ggaatccacg aacctttac 600
 tgcgggtcat gggcagtcct agatggctcc ctcccccac agaggggggtg cagttagacc 660
 tccggaagtt actgctcttg ttacctcaaa agggatttbo agatcagaca gccctctac 720
 tccaaaggaa gtgtgtgagc ctgtgtacot ttattttatc cctgtctcaa cccctgtgga 780
 ctgctgtcac ccagaatggg gctctctcgt ctggcaagtg gctgagaacc tccactccac 840
 tcagcagggc tgttccccat ttaccgaaaa gctccgagag aaataacta accccatggc 900
 ggcctgttag ctactggcag agctcctcgt tcccccacct tagcgcctgt gggttttgtt 960
 tccatcgagag tgagcagatga atctgggac ccatcagcag tcaagttggg tgcctcgag 1020
 gcaaatgat agatgttgtt gaagggtatg tgtgaggata ttaatttaata ttaacatgct 1080
 agtttatatta atattctcat tggaaatttg ggggcttttg aagttatatt tcaaatatat 1140
 ggtttttatgt aatcctttta actgccttga aacctgaaa ttattgcaac tattttatag 1200
 atggagagac tgaggtctag aggggtgaat tgcttaagat cctgggggtg gaagacacca 1260
 gcttttctctg ttttgagtc ttggcccttc ctgctgagta gctaccccac acacagacct 1320
 gcccttggag agcttgcagc cactctggga aggcagatgt attggattgc tgccttagac 1380
 tggaaagcac gtgaataaag ctccaggtta aaaccatggg ggttccagga ggcagcagtc 1440
 ggcctctgctt gggggtgagc tgagcagcgt gtgctctctg gaaccaaggt agttgggctg 1500
 aggtctcagtg gacagtgagc gttggcaggt gaagtcagag aggtcttttg agggagtggt 1560
 accaccttga gacacacag aggaatgaga caggcaggtt actcaaggag cagaggtctc 1620
 gtgacacatt cccagagcat gtgggttctt agctcatctt ccaggaggag aaagtgcac 1680
 tatcacaga tttgtcaatg gagtttaaat aggatgtggg aaatctaga ttttcaaaa 1740
 cagtaacata ttgttttag aagaaaggtg gatgcaggat gcataggtta gataatttta 1800
 atagcagtaa cctcagagca tgtaagtatg atttgattta ctggagtgcc tggcctgtct 1860
 agtcagtggt agcacggctt gggctggagc atgaggttga caagggtct tctctbdaa 1920
 tgccttctctt ggtttgctaa gaggtatctc ctactggccc gggggcagaa gacttttctc 1980
 tcttttccca gttttcagta gttggccagc atttgtgaa gtgggagaaa ggcctgctct 2040
 gctttacatc agatgtgct gtctgactt gatctctggt gtgcttcca gagaccgccc 2100
 tccatctcct caactccctg gcttgatgct taggtgttga tggctgttgg gcaacaggag 2160
 tacataacag atctgtgatg gaccaggag cagagccaga tgagtgaatg tcatggagt 2220
 ggagtgtgtc tgcattggct tgggtctccc tgcagcttct gcaaggatag tggcaaggag 2280
 tgctcaccac tcatctggaa tggctagact ggaagcactt ggcctcttg ggcctctgac 2340
 cccacccccc tccacactgg cctgctgctg catcttctag ggcctctgg gagaccaatt 2400
 atggctgctt gtcatagttg ctccaggtcac cgttcacact tctctggacc aggcacatag 2460
 agccttgaga aggttcaagg ggcdaagtgg gcttagcctt ttactgacag ctgggaatag 2520
 caagcgtgtg gaccagagca ccaagtgagt tggggccgtg gtgggttccg caccgtgtcc 2580
 ctaccagag ctccatttgt tgaataacagc cttctctac cgtttcttca cttggcaaac 2640
 tttaaactat gtattggctg gtccgggtgg ctccagcctg taatccacag actttgggag 2700
 gccaggtgtg gccaggttaact tgaggtcagg aggtcagag cagcctggcc aacatgttga 2760
 aacctcatct ctactaaaaa tacaataatt agccacaggt ggtgacacgc aactgtaatc 2820
 ccagctcatc gggaggctga ggcagagaa tgccttgaac ttgggagga aagattgcag 2880
 tgagctgaga ttgcacacac gcacttcagc ctgggagaca gagcgagact gcatctcaaa 2940
 aaacaaacag aaacotacat atttctcata ttcccccaaa catlgagget cactttcttg 3000
 tgaacaacat taatgtact gtgctctctt ggcataatbt tccaaaatta cagatgttct 3060
 tataatttca ccggcagctc tgcctccag aatttatct acggatgggt taacacgtgt 3120
 gcaaaatgat ttatttgcaa ggttgtctat tgttgctta tttttaatag caaagattg 3180

WO 02/48370

PCT/US01/51341

29

```

gaggcagctt aaatgttcat tgcagggggc caatgaacaa accatggccc gtctaaacat 3240
gggataccgc gtggccataa tacataagat ggacgctcaa cgcactgtgc cggatttggc 3300
agcaagggtgg attgccgagg gaagaagcag gtctggggcg tctgtctcgg agctgccatc 3360
agtgtaaaag ggagagagaat caaatgtgtc ttctgtgtgc tatgccaggg gggctctcgg 3420
gcagacacgc caagtgggtg attgtgatgc ctctggaggg ggtgctggtc atgggagatt 3480
gottgtttgc tggagatccc atgtaccttt tgattgcgtg agcagggtgaa tgcagcctt 3540
tocaagaat taaaatgggc caggtgggtt ggctcagccc tglatccca gdagttggg 3600
aggctgatgt tggaggatca cttgaggtga ggggttcgag accagctcgg ccaacatgat 3660
gaaccccatc ctccactaaa aatcacaaaa ttacccagac atgggtgtac atgctgttaa 3720
tcccagctgc tctagaggct gaggcaggag aatcatgtga accctaggag ctgagtttac 3780
agtggcccaa gatcatgcaa ttgtactcca gctgagctca cagagcgaga ctctgtctca 3840
aataataaaa taaaataaat taaaacataa aggactgtaa ccttgctccc tggccaggtt 3900
aggaagggtca aggtcttggc tactttctca gtacaggagc ctccctcagg ccccgagcca 3960
ctaatacaaa aatatgtgct tggttctcac aaaggggcgt agtgtgaggg cttgggtgtt 4020
gctttgnaaa tccgaccccc ggctccgggc ttggaagat ggagcctct ctggagcccc 4080
tggacacatc gctgttggtc gactttgtca ttctcaaccc ttgctccgat tgggtcagct 4140
catgatttct gaacaccttg gggccttccc caotgaocaa aagataaact ttaactaagg 4200
ccaggccctc ttactgtggc ctctcttga gccgaacttg gctgtcac 4260
ctctctctct tctgctctct taactcccca cctccctgcc ttgtctcata tagtctctt 4320
tgcctgctct tccgtccaga gcagtctcca cgtgccagg tctgtctga ctttcaagg 4380
ccagcttagt ttccactctc gactgctct ctgacctccc tggctctgt gttaactgcc 4440
cagataagc caacaatgg tctctgccc caaggagct cctggggccc cctcctggc 4500
cactcgtctc tgcgggttag tccacactca cactctggca ctttgcctg gtgctgccc 4560
ctgctgttlt gggcctccca cacacagagt gtacagaaag gactctctg tgtctggctg 4620
cctcccccga gcatctgtag atcatccagg ttgctctgag tggcctcttg gttttttctt 4680
ctgctgctga ggaagttaac aaatatacaa ctatttatc attctctgt ggacaggcat 4740
ttgtctatgt ccaacctctt cgttgaaatc attctctgt ttggggccc gtgtggctc 4800
ttgtctgggt atcacccdag ggtgggttga tggcttacct gactcagaat gtgtttgcat 4860
gaatgaatc caggttggta tgaagaatc aggtgtctct ggttgagcc aggtctctt 4920
attacaggga cagagcaggt acaggatcc tggtttagac agctgtctc catgggggtg 4980
tagcattggt ggggtgcagg atgctgaatc tgcaggggac ctatccgctc agtgcccatg 5040
gggattttag ctggctggaa aggtgtgctc atgtagagg gctcaacat ccagctaaag 5100
aggctgaggg ttggtccatt gttctcaatt tgagagaaa ctgagatcat caaatlagg 5160
actggtatgt actaaaggaa agaactaat tacaaggctg aattgagtaa gccctgctg 5220
agggactctg gatttcttct ttgttctctc ttatttctc acccccaccc aagtgcagaa 5280
tatgtacatg attggtatgt ttgtcttctc tggtttagag attctggga acttggccca 5340
ggagaagggg gagaaatgt gagccctag agtggctccc gottgtttgt gtgtgattga 5400
ggggagagcg aaggagagct gtggacccct gaocctctgt gagggcatgt gatccttttc 5460
aaaaggctca ccaggcagaa gtgcttgccc aggggcccgt cttctctctc aatccctctc 5520
ggagaagggc caggctgtgg gttgctgacc tgcctctgat tggatcagcc tcccccaata 5580
atgcagctgc ccagaagctc agagagccca ggcacccccc aaaggcagga gggccggctg 5640
tcattccgtt tgtcattccc aggcggctgg agtgggagca gagcgctcag ttcagatgaa 5700
cagtgtctga gtctgacccc aaccagcgag ttatggttag atggaaagt ctccatctat 5760
attaaataag agaacaaaag cctccccagg ctgcatgaat attccaggga tatatatgt 5820
aacgggttgc cagtttagct tggctgtgtg ttggcagcgt cctgagtag caottgtgtg 5880
ctgagctctc aaagggtttg gatctgaaac taatgaatga aaatatgacc tcagaagatt 5940
taagagagag aataccccag caacagacc tgggtcccag agactgttgg gagatgaaa 6000
tcccaggctg gccgaaggag gaagtgggag agcaatggca gctganatca catggtgcaa 6060
gacctctca gtgctttctg tgttcaacta ttattcctgc cctctctctc agaggcaggt 6120
atggctgtct ccccatctta tagatgagga agctaaggca aggaagaggt gtgttaactg 6180
ctcacagaca caaagcttagc cagtggcaaa gctggaggtc aggtctaggt ggtcaggctc 6240
cagagttctg cggatttccac agcacggcag tggcagtcgg agaaacatt tgtcaggtga 6300
ttgtggggaa atgacgtcag ccttcaaac ctctgttttg catctgcaag ctgctgtgtg 6360
ctgcaacaaa ttacaaagaa cttagtgact taaaacacaa attaggtcgg gtgggtggc 6420
tcacatctgt aatcccgaca ctttggagg ctgaggttgg tggatcactt gaggtcagga 6480
gttcagagac agcctggcca acatgatgaa accctgtctc taacaaaaat ataaaaaatt 6540
agccaggcat ttggccgggt ctgggtgata accctgttaa tccagaactc ttggggagga 6600
aaggtgggct gaacacaagg ccaggagttc aagacacgcc tgaccatatt ggtgaaagcc 6660
tgtctctctc aagaatacaa aattagcagg acgtgttggc accggcctgt agtccagtt 6720

```

WO 02/48370

PCT/US01/51341

30

```

actgggaggc ggagggttgoa gtgagccaa atcagccac cctgggtgac 6780
agagtgagac tccatctcaa aaaaaaaa aaaaagtaca aaagagcaaa aaaaaacaaa 6840
agttatgaaa atgaaaacct gagccatcct tbatcttatt tcccaaatc cactaattat 6900
taacagaaag taagaacctat gaaaatgaa tgaaagtacg tgaatttcc ttgaagtgtg 6960
ttagaacctg ctttagtgtt cagctatggg tccctcatg aaggtcagct gagccatgac 7020
catgaacctc tggagcttgg actctagatt gaacctctg agatgccaaa gatgtccacg 7080
tctaatcccc atgtggggaga cagaataatg gccctgcaga ccttcccagc tggccatgac 7140
ccctcatctg acnagctctt ccttctctc tgaccagcac catgcttctc ctggtgacaa 7200
gcctctgctc ctgtgagtta ccaacccag catctcctc gatccacag aatctggatc 7260
tgogaacagt ggcaccagcc tctagtctca atgtgaggtt tgaacctagg aagatgaatt 7320
taagctggga ctgccagaaa aacacacac tccgaagtg ttcttaact gacaagaaga 7380
acagagtgct ggaacccagg ctacgaaca acgaatgttc gtgcacatt cgtgaattt 7440
gtctgcatga aggagtcaca tttagagttc acgtgaatac tagtcaaga ggatctcnac 7500
agaaaactgc ttatccaaat tcaggagggg aggttaccgc tgotcagaat ttctctgtt 7560
tcactacaaa tgcggattta atgaactgta cctggggcag ggtccgacg gcccccgtg 7620
agttccagta tttttgttac atacaaact caaaggagag gaggggatc cgggtgctt 7680
attacataca agaactcagg acccatgttg gatgtcaact gataaactg tcaagattaa 7740
cgtctcgaaa ttactttctg gtttaacggaa ccagccaga aattggcatc caattctttg 7800
attcaacttt ggaacacaa gaaatagaa gattcaacc tccagcaat gtcacgtac 7860
gttgcaaac gacgaactgc ctgtacggg ggaacagcc caggacatc cagaagctgt 7920
cgtacctgga ctttcagtao cagctggag tccacagaa gaataccag cctggacgg 7980
aaaactact gattaatgth cctgtgatt tggaaaatg atcaaatct ccaagctctg 8040
agccacagc aaaaacagb gtgaagatca gagctgcaga cgtccgcac ttgaattgga 8100
gtctcctgag tgaagccatt gaatttggt ctgacagcg gaacctcgc tctgtgtaca 8160
tttatgtct cctaactgtt ggaacccctg tctgtgcat cgtctcgcg tctctcttta 8220
aaaggtctg taggatacag ccgctgttcc ccgcagttcc acagatcaaa gacaaactga 8280
atgataacca tgaagtgaaa gacgagatca tctggaggag attcaccaca gaggaaagga 8340
aagctaccgc cgaagaggtc ttgaacctga aggaattac ctgaagacca gaggtgtag 8400
gaatggcatg gaactctccg cctcccgcac acgggggaa tgttttctg atgatctgt 8460
gaacctttat atcattttct atgtttttat ttaaaaacat gacatttggg gccaggcgcg 8520
ctggctccgc cctgtaatcc cagcaacttg ggaagccaa gcaagccgat cacttgaggt 8580
caggagttcg agaccagcct gcccaacatg gtgaacccc atctctacta aaatacaaaa 8640
aaaatagccc gggcgtgtgt gtggcgccct atagctccag ctacttgagg ggttgaggca 8700
ggagaattgc ttgaacctg ggaagtggag gttgcagtc gccagattt gtgcactgc 8760
actccacgac tggcgacag agccagatcc catctgtctc aaacaaacag aaaaaacaaa 8820
acaaaataaa ataggccag tatgatggct catgctata atccagcac tttgggaggg 8880
aagcgaggtg gatcaattga ggtccggagt tcgagacaa cctggccaat acagtgaac 8940
cctgtctcta ctaaaaatad aaaaattagc tggcgatggt ggtgcattgc tgaaccca 9000
gctaactcgg aggcctgagg aggaactca cttgaaccc ggagatggg gttgcagtga 9060
gctgagattt gccactgcac tccagcctgg gogacacgt gagactccat cttaastaga 9120
agaaaaggtt tctctcatg gacattgttt gcatctacat gtgacatca ggaatgatct 9180
gttttagtctc aatcaactcac tcttgatct gctgtctct ctctgagata acaaggcct 9240
taattgttag caaactgcac cagagtgtgt gaggtgttt gaaacaaatc atcctaata 9300
aaaaagaaca gcttttgtaa gggggcactg agtgtctcaa acagccgcat gggcaggag 9360
agtgtcagct ccagtttttg ttgaattgt cttgtgtccc taaggctccc tatgaagac 9420
tgacaggtct ggaactgaatc ttgtgatctg gacaccaaag gtccactgtg ggcacagagc 9480
tagctctgaa gaatggggtc gttctcttga gaacctccac agcaaaagt ttgctctctg 9540
ttcccaatgc atgtcccaat ttaccagcta catcccccag taactgacca tggctcatga 9600
ctcatgaatc ataaaaactda gttagcagggc ataactggtt cagactgcc agggctatgt 9660
gggaactatc attggtacaa aaactctaag tgtggagaag actgtggtag acaagagggg 9720
acatgtctgt tctaaacgca catcagaaac tatggcagac tatggcagac tgagataagg 9780
gtgtacagaa cttctcaggga catgcagacc tatgtgtcac tcataactga aattcaata 9840
aatatbtgtt ggaattccaa aaaaaaaa aaaaaggcg gcc 9882

```

<210> 72
 <211> 93
 <212> DNA
 <213> Homo sapiens

WO 02/48370

PCT/US01/51341

31

<400> 72
 gttatattaa aacaatagaa acattaatct gtctgtcttt tctccattct atccattcgt 60
 tctttaatgt ggtcactttt gaatgctgta tac 93

<210> 73
 <211> 299
 <212> DNA
 <213> Homo sapiens

<400> 73
 ctcgagcgt cacaattac cacctctgta aatcctttt baacttatto agggtagaag 60
 aattctgtgt ttctgtgccc othhaactc tgttatataa gtctccttc ccaaccaccc 120
 ccacacttac cacatcacgt tagcaagaat gagagcaatt tgaaggcagt ggctttgtat 180
 cttaattata gccctggcac ctaaacagtt tgtaaaaagt taactcgtg cagggtagca 240
 taacacataa gagtctgttt cttttgagat atttggcaga ggttgtggtg tgcggagat 299

<210> 74
 <211> 94
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (85)
 <223> a, c, g or t

<400> 74
 gctgtgttta tgcgtctggt tgaactggga ggaatatggt cctttgtctc tgaccagga 60
 gtttcaatgct ttctgccaag atacattaca tggatagata cattaggtag gtatatacat 94

<210> 75
 <211> 433
 <212> DNA
 <213> Homo sapiens

<400> 75
 gctgtgttta tgcgtctggt tgaactggga ggaatatggt cctttgtctc tgaccagga 60
 gtttcaatgct ttctgccaag atacattaca tggatagata cattaggtag gtatatacat 120
 tagatataga tagatacatt agatatagat agatacatta gatatagata gatacattag 180
 atatagatac attagatata gatggatata cagatagata oacagataga tagatagata 240
 gatagataga tagatagata gatagataga tagattcatt tattttattga gacagagctc 300
 tgcctctgta cgaagctgg agggtagtgg ctgtttcttg gctcactgca acctccacct 360
 cctgggttca ggtgattctc ctgcctcagc ctccacagca gctgggatta catgcccacc 420
 tattttgtac ttc 433

<210> 76
 <211> 334
 <212> DNA
 <213> Homo sapiens

<400> 76
 gctcagaggtt aatggaccat togggtata tggttcatal tttttgctca tttttatgbc 60
 atgggtgttta tcttttctgt gctgatttgt aaaagctatt ttaaaacctc tcatctgcca 120
 tatatgttac atttcttctc tgccttctgc caacttccaa ttgtttacca actttctctc 180

WO 02/48370

PCT/US01/51341

32

```

ccaaccttgg gccactggca tatacactca ttttaaatat cagaacttgt agtgcctttt 240
ggaatgcaga cagactatgg ttcaattctgc aactgcatat tagttaacag gcaaaaatac 300
cttagtaaga gaaagtgtct tttccttcta atgt 334

```

<210> 77
 <211> 547
 <212> DNA
 <213> Homo sapiens

```

<400> 77
ggcttatatg tggagaactg acgtctgaac ccagatctga ttcccaagtg taatactttc 50
caataggcag ccttatatct ctgtacotca aaagagaagg ctatatattt taaaagatta 120
ggaattgtcc tatatgggtt taaaatacac ttgctatagc acataataaa gtgggttagt 180
ggtagctgtc actccgtgtg gtttggttta aaacacagcc agtttgtacc ctggtggtda 240
tgataaaagc ataccacctc tactttgaga attttaacca tagagacaaa tatgtgtcaa 300
acagctbaaa aaagtartct ttctagttgc attttgagcg aatttgaat tgcttagact 360
ctttgaccaa aagtaaaaao tgcgtgttaa ctggtgacaa aatctgtttt catggacgtc 420
aggctaacta agctttattt tctctctaag caktctctgc ctttgttaag cactcttagc 480
gcagtakttg cttagettct aattttgggt ttgcttttgt gttttctctc tttctcttgg 540
ttgttcc 547

```

<210> 78
 <211> 263
 <212> DNA
 <213> Homo sapiens

```

<400> 78
tcgaggggtg aaatgagtgt cattagccaa gtgacattta agtgccttgg tttgtctgct 60
tgcttttctg tggattgaaa aaaactgacc actgttaata tgattgtaca gtgacactgg 120
aaattatgag atgtgtgtct ggttagtctc gcttgtattt cagttgagat gcataccaag 180
totgataatg cagagctttt ccaatttcacg tgtctgttta ccattttcat gatcttaagc 240
aataacattt tcttgacaa agc 263

```

<210> 79
 <211> 765
 <212> DNA
 <213> Homo sapiens

```

<400> 79
gcggaagag caccgagccc tgcgagtact atttcgcgt gtaccactcg ctgtgcocaa 60
tcagctgggt gagtccgcag agggggcgcc gggccaggcg tgtgcagggc tcggccgagg 120
ctgagccggc gtcccgtccc ctgcctttct gcttcccagg tggagagctg gaaagagcag 180
atcaagaagc ggaatttttgc cggaaaaatc tgaatgcacc agcgcggctt cctctgaaga 240
tgcaagtatc ctgcattctt ttgtctcgcg gagccccggc tctcggttat ccacccctac 300
ctcccagtg tcaagccacg aataatgcga ccagccttcg agttccttgt tccctttgtc 360
ctggtctcca cgtgtatgat ggggttctca ggcocaggct tcgaccagag gacctcttgc 420
caccacgctt tcttctctgc cttgagctac ctgggtgaac tcatgaaccc agggccctgc 480
tcacccagga tgtcccccag gtccctgccg ctgggaagt ccagcatgaa cgcctccaac 540
ttcgtggag ccagggtccc ctgcagctga gggacgocaa gcagacacac ctgcccctcc 600
cagccagctc ctgtctgtat gggcagatg actgagagcg cccacgtccc taaggctgtc 660
ctgacccctc atgctgcgac aaggacaagg aatggcgggt caatatgggc ctggtgtctc 720
ccctccccc caccocgggt ctgcccagct caagccagaa gtgac 765

```

<210> 80
 <211> 162

WO 02/48370

PCT/US01/51341

33

<212> DNA
 <213> Homo sapiens

<400> 80
 cgctgctcca agaccaggac ccgcgcgcgg aagagcagc agccctgoga gtactatttc 60
 cgctgtgtao actcgtgtg odcacacagc tgggtggaga gctgggaacga gcagatcgaa 120
 gaacgggatt ttctgctgtg gcaaacatct tgacttgccc ca 162

<210> 81
 <211> 986
 <212> DNA
 <213> Homo sapiens

<400> 81
 agcggggcgg gcaagcaggc tccattggc tggggctcgg gcgtcctagc caatccggcc 60
 gcgggggtgg tttctcctga cccgggtggg accgcacccc ggggaotcag aagcgagcgg 120
 caccocggga ccatccacac gcagatccag tggccgcaca cgtcaggctg gacttgccct 180
 ctctgtggat gttggatgtg gaagcccaag agcccccaca ggggaatgg tgaacgcgcg 240
 ootogacccc gcgctccccc agccagaacc agatccgtaa ctgctaccag aacttccctg 300
 actaccaccc ctgctcaaac accagagccc gccgcgggaa gagcacgcag ccttgccagt 360
 actatctcct gcgtglaaca ctgctgtgc cccatcagct ggggtggagag ctggaaacgag 420
 cagatcaaga acgggatttt gcgcgcacaa atctgactgc cccagcgagg ctctctctga 480
 agatgcagtg atctgcaccc ttcttgcttc gcggagcccc gggctcgggt tatctctga 540
 tacctcccaa tgtctaaagt acgataatg ccacagccc tggagtctct tttttgccc 600
 tgcctgtggt oacagttgt atgatgggt tctcagccc aggtctgac cagagagccc 660
 ctctggcccac aacagttctc tctgtgact tgagctacct tggggaactc atgacccag 720
 gccctgtgct ccatcaggat gtcccaccg gtctgcag ctgggaagt ccagcatgaa 780
 cgcctccaac ttctggagc ccaggtccc tgcagctgag gaagcccaag cagacacac 840
 tgcctccccc agacagctcc tctctgtatg gagagatga ctgagagcgc ccagctccc 900
 aaggctgtcc tgacctccat gctgcacaaa gacagggaa tggctgggtc ctatgggccc 960
 ggtgtctccc ctccccatc aaccgg 986

<210> 82
 <211> 369
 <212> DNA
 <213> Homo sapiens

<400> 82
 aacccaagat gactcgtctt ttggtgggag aattcactct gttcatgttt catttaacaa 60
 ttgatctact gtacttaatt acccttggct tctttacat ttattggttt atcttgtgtt 120
 ttctctccct ctgatctggt tatcgatttc cttttctctc cctgttgca ctttccattt 180
 cattattggc agctgtccct tctctgggt tctaatcaa acacatattc ttttagacat 240
 gcctcgatgg ggattctttt cgcagcacc tcatctggag ctacacagac ctgtcactct 300
 gtagggtctg gtcttttttt agcttaggaa catctatttg ttgcttgatt tgattattgt 360
 tagtttgtt 369

<210> 83
 <211> 923
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (354)..(565)
 <223> a, c, g or t

WO 02/48370

PCT/US01/51341

34

```

<400> 83
aacccaagat gactogtctt ttggtgggag aatcoactct gtccatgttt cattaacaa 60
ttgatctact gtaactaatt accttggct tattttacat ttattggttt atcttggtgt 120
ttcttccct ctgatctggt tatcgatttc cttttctctc cctgttgca ctttccattt 180
cattattggc agctgtccct tctctgggt tctaataca acacatatto tttagcaact 240
gcctcgatgg ggaatttttt cgcagcacc tcatctggag ctacagaaac ctgtcactct 300
gtaggttctg gtcttttttc agcttaggaa catctatttg ttgcttgatt tgaannnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
tgccatgggg taggggtgag ggtataagta gatcagagtg ggaagacctc agccttgggt 660
ggcttctctc tgccttcttg cagggtggag ggcctgloca cacttgatc cccgtaccac 720
agtgccagcc atgccttcc ctgggttacc atgtcccttc tccctaccca gttggttagag 780
gagtcaggag gtggagagcc gtgggcttg gttttataat gtaaccactg tgggggtggg 840
ggcggatggt gaaccatgta ttccagtga atatttaata tatttaata tcaataaaat 900
caaacctctt gtaaaaaag cag 923

```

```

<210> 84
<211> 338
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (12)
<223> a, c, g or t

```

```

<400> 84
ataatttttt tntttttaaa ggaatgaac gtggaggact ggggtgaagg gccagcctgg 60
gtagtttaat ctttttggga agacatgact ttaaggagat tccctgcttt gtgacaggtt 120
gtccatgct gtcttgggga caaggcctg tactgccttc aaatctgggc tcacccacaa 180
ttttggtgag ggaagatag ggtgggggga taaggaggag aaaagactct agcttttttt 240
ttcatgcat gatatactgt gtgggtttat caagagtgtg gacacagttg ctgttctcaa 300
atsataggcc aaataaaatg cgattctttt ttctcttg 328

```

```

<210> 85
<211> 436
<212> DNA
<213> Homo sapiens

```

```

<400> 85
ataatttttt totttttaaa ggaatgaac gtggaggact ggggtgaagg gccagcctgg 60
gtagtttaat ctttttggga agacatgact ttaaggagat tccctgcttt gtgacaggtt 120
gtccatgct gtcttgggga caaggcctg tactgccttc aaatctgggc tcacccacaa 180
ttttggtgag ggaagatag ggtgggggga taaggaggag aaaagactct agcttttttt 240
ttcatgcat gatatactgt gtgggtttat caagagtgtg gacacagttg ctgttctcaa 300
ataataggcc aaataaaatg cgattctttt ttctcttgaa acacacagaa cagccagctc 360
ataaacagg caactgaggga agaaccaaac gcataccgg caagactcta gcatgtcaag 420
gtcaaaagact ctccag 436

```

```

<210> 86
<211> 462
<212> DNA
<213> Homo sapiens

```

WO 02/48370

PCT/US01/51341

35

```
<400> 86
agggaacggtt ggatgtatgc acactgctgt tgggtttact tagacctca tttttccacc 60
agactstagt gttcaaaatt ctttttagta agagaaacct tttttctga accttttaoa 120
accatctcca aattataaaa cataagactt ttttttagta aaaatatatt tttttacaag 180
cacagtggct tgcacctgg aggggagagg aggtgttttg tcottggagc tgcctggcctg 240
agagaacctt gtcactgtgg gagctgggoc attctacac agtggctctgg caatgacccg 300
gtggtgggtgg aggcctgtga gtgggcactg gtaatgggaa cagctgtaaa acctggagg 360
ccagccccag gagagtgaac ttaccacagg aagttctggg aaacaaacca caggggaggt 420
ttacaggaaat ttttggttgt gcccaaggc aaggcacatg ag
462

<210> 87
<211> 1435
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (1012)..(1119)
<223> a, c, g or t

<400> 87
ttagaggtag aagaactgac tataagcaga agtgtttgag gaggtgcat ggagaacaa 60
gggcataatc ttggacattg gcaaggttggc aggatttgac ttgggtgaaga gaacgagaaa 120
ggggacttta actggaggga ctactctggc ttgattttct ccatctggg gagatttgctc 180
tttggaaagt tttagcttcca gagaccttcg atgttttgta acatgtocaa gctctacatt 240
tattgattgt tggttctggt catggctatg ttcaaatctt tgtacctttt tgcctccac 300
agtttcttgt ctcatccctg tcttccaact ctgtcccccg ctctgtctg gtctaakta 360
cttctctgtg tggagcagct tccctctttg ggtaaactca gacatgaccg cagcaaaaga 420
gcttggaatc ttctgttttg tcaagtgttc cccagcttc cccgacagata cagctgaatt 480
ggagcccttg aagcaaaacc agagaagtgc tgcactctgg gggcgaggag gctttgcttt 540
gcccagggtc ggcctcctga atgaattttg tgcagcctt aacggccggag ttgtgtctgt 600
gaaggtgcat tgcctctgtg ccagggcaatt catggagggg agaggaggtg ttttgtcctt 660
ggagctgtcg gcttgadaga accttgtcat gctgggagct gggccattcc taacagtg 720
tctggcaatg acccggtggt ggtggaggcc tctgagtggt caotggtaat gggaaacag 780
gtaaaaacct ggaggccagc cccaggaggag tgacctaac aggaagtto tgggaaacaa 840
accacaggga ggccttacag gaatttttgg ttgtgccac aggcgaaggca catgaggaaa 900
agaaatgtaa ttatagtttg taagtcatg aaaaaggcca atgagtgaca tgaatatag 960
gctctaatgt tcttctctct gtcggacagg aagaatggg gttttatgca tnnnnnnnn 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1140
taagcagata actgtataaa tgcataatta cacagcatgg tgagtgtct gaaggataag 1200
tgtggggagc ctcatctaga ttggaggatt gtgaaagtca agagacagga gagtcaagg 1260
gagggcaagg gactaaggc tatccaggca aagactgctt ggtagggggg tgcctcagca 1320
acgggaacaa acctggaaaa aatatgacac ctacggggaa ctaaaaggag ctgtatgtg 1380
ctgatgcaca gacagggaag ggcagggaag gtgctgaaag aaggcaggag gagaa 1435

<210> 88
<211> 459
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (437)
<223> a, c, g or t

<400> 88
```

WO 02/48370

PCT/US01/51341

36

```

gtctggtttg agtctaggat gaaggtaact tctccagga aggcctctgt gttcctctcg 60
ccagactcct gagggtctcg ccagttcaag cccacttgaa gccagctcg tttgggggta 120
cttgaaccat ctgggggatt ccaactagta tcttttagtc ctgacatgag ctgttctact 180
gtggggtcag ccttggtctg agactgtatc cctatagggt ccggtctctc tgttgacccc 240
tcacottctg tgggctctgg catggaactc tgatccttcc atctgaagaa gtgtcaaaa 300
aaaagtccat gcttcctggga atcaggaagt cgcctcaagg caaaagtagc tgagtgttcc 360
tatatctgtt ttgttttctt ttctaaattc tttttttgtt gggtaattct tcaccatctt 420
gttgatctct taagtcntag catacacacac attttaaaa 459

```

```

<210> 89
<211> 1263
<212> DNA
<213> Homo sapiens

```

```

<400> 89
gtctggtttg agtctaggat gaaggtaact tctccagga aggcctctgt gttcctctcg 60
ccagactcct gagggtctcg ccagttcaag cccacttgaa gccagctcg tttgggggta 120
cttgaaccat ctgggggatt ccaactagta tcttttagtc ctgacatgag ctgttctact 180
gtggggtcag ccttggtctg agactgtatc cctatagggt ccggtctctc tgttgacccc 240
tcacottctg tgggctctgg catggaactc tgatccttcc atctgaagaa gtgtcaaaa 300
ataaaagtcc atgcttcctgg gaactcaggaa gtgcctcaca gccaaaagta gctgagtgtt 360
tttatatctg ttgttttctt ttctctatct tctcttttgc gtgggttaatt ctccacatct 420
ttgtttgattc ttttaagtctt agcataacac acatttttaa aatccagttg ttttagttgc 480
tttctgtctc catagaaggt ccccatggtt ctacgctctg tgggactctg agcctgttac 540
catgacccag gacagggagt cctcatgctg ttttaagcag tgggtgatcta agttttattt 600
cttaggttgag taaaggtctg aaagcttga gacccctgct ctaggggctg taccctgtcc 660
tttctcctct ttctcctgtc tggactaggg ttccgaaggc ctggtgggac atgtggagac 720
caagtagctg acaatcccca ggaactgttg gctcagacac agggcctctg acctctcagc 780
ccttcctgct tgaactcagc acctcctctg cctggccctc ctttctgca tgagctctct 840
gcctctgcca ggaggaacct ctgtcctgtt tctagatgag ccatatctcc tccacatctc 900
tgctcttctc tccagtgttg tgcctctgta cctctctctc cctcgaaggc taaatcaaac 960
cctacctctc tacaagagag gaagttaatt ctgggttgat gttatgcatc gccagattca 1020
tgctgagcca acaggttagg ggcctggagaa acagtgtatg gcttaaccag gccctgccag 1080
cctgccaccc ccaggtcttg tgagggttag aaaaaacata aagtgaattc gataaataat 1140
ataatctatc catatccata tttttatttt ttattatttt gggacgaagt cttgctctgt 1200
cactccagcc tgagctacag agtgagaccc tgtctcaaaa aaaaataaga aaaaaaaa 1260
aaa 1263

```

```

<210> 90
<211> 554
<212> DNA
<213> Homo sapiens

```

```

<400> 90
gtctgagctg tttttctcag gtgagtagaa caatggcatt ttaaactcaa gaggaacctc 60
gtaaatacat ttatttcaat tcttttctca cataggggaa gaaacagagg ctgcaaaaaga 120
tttagtttag tcaagaaaaa acagtataat ttggagtttt tgactttgtg agttttgtta 180
cggcgtgac attcattctt ttgtgcgttc agtgatttca aatcttcaaa tctagagcac 240
attgtatgct gggcagaagg cacagbactt gaggattcag tggcagctga taccagaaaag 300
gtctgctgctc ttgggcaactg atgagcctcg ggctactaca agtaagcagg cagtggcagt 360
aggctgaatg agggctgcag gtccctggcat catggatacc aatttgggct tagaatggaa 420
gcccaggctt ccttgaaagaa cagcggctca agctgagact tgtaggaaata gtggttaatta 480
acaaacagac aggaagaaga gotttccagg aagacagcaa aacataggga aaggctctga 540
gaggagagag agca 554

```

```

<210> 91

```

WO 02/48370

PCT/US01/51341

37

```

<211> 435
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (406)
<223> a, c, g or t

<220>
<221> unsure
<222> (411)..(412)
<223> a, c, g or t

<220>
<221> unsure
<222> (421)
<223> a, c, g or t

<400> 91
tattagtcaca taaggcttat ttctagtatt aaacaatgct taagaatagc ttggatccat 60
gaaaactttt gagaaggagg acaaaagcaga cggaaacctaa tctctgaaca atttcaatta 120
catcttttac aagtggctgt tggctagtaa ttaaaaatga gccattcaaa cttgtggaca 180
ccttttttgc catgcagaact tgaattgcaa agccttttatt atccctgggt asgaacagca 240
cagctaataa aaacgaatca tatggcttta aactacttgc atccaacagg gacatcctaa 300
aastgggtcg gatagtgaat toatgaacct ttaggctgca agtgccatag ttactaatga 360
gaacagatat ttccaaatgg cggcaataga ttatggaaaa tggagaaagg nnagagaagta 420
ntttactttc agcta                                     435

<210> 92
<211> 580
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (551)
<223> a, c, g or t

<220>
<221> unsure
<222> (556)..(557)
<223> a, c, g or t

<220>
<221> unsure
<222> (566)
<223> a, c, g or t

<400> 92
aaaaaaactg tttagaaaaa ctccatattt actctccctt tcaaaactatt ggccctgatt 60
tttacagata atcaaaagtc aggtcgccaa acttatttct tttgaatttg gaatatcttt 120
taaaatttgc cttttttctt ttattattta gcccaataag gctatttcta gtattaaaaa 180
atgcttaaga atagcttgga tccatgaaaa cttttgagaa ggaggacaaa gcagacggaa 240
cctaactctct gaacaatttc aattacatct tttaacaagt gctgttggct agtcattaaa 300
aatgagccat tcaaaccttg ggacaccttt tttagcaatg agacttgact tgcnaagcct 360
ttattatccc tggtaagaaa cagcacagct aataaaaaag aatcattatg ctttzaacta 420
cttgcatcca acaggggacat cctaaaaatg gtcaggatag tgacttcatg accatttagg 480

```

WO 02/48370

PCT/US01/51341

38

```
ctgcaagtcg catagttact aatgagaaca gatatttcca aatggcggea atagattatg 540
gaaatggag naaggnnaga gagtanttta etttcagcta 580
```

```
<210> 93
<211> 724
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> unsure
<222> (297)..(502)
<223> a, c, g or t
```

```
<400> 93
tactgatgtg cttttgattt gtctggaggg tgactactac ctctttgagg tgctctctgg 60
gacctcctaa atattaactt ttatactctg tgtagcctgt actttaagcc agaacattca 120
agttacactg aagaaatgtg ttgaaaatct atgcaacctt ttgcgatta tgtactagca 180
ataaacaact ctttaatttc tggaaatttc catcttctct agtgatattg ttgattgatt 240
tgtatttttc ttctcttgct aggtttcagt atcagggtcg taccaatttt tttcttman 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nntgtgcctt ctttatgaag tgaattatga agctttccaa tcttttttat tttgtagaac 660
agtttaataa cacaacaata tactaagttc ttagattgaa gctgttttta aatcacaaag 720
acag 724
```

```
<210> 94
<211> 586
<212> DNA
<213> Homo sapiens
```

```
<400> 94
ctaagacagt ggccaatctg actgtgaaaa taagggcagg ctacactgga gagcagggat 60
agggacaccc ggggggcaga gatgtgggtc accttagggg aggaacaccc caggaggroog 120
gcccatgatg gcccatgaag gctggggagca cgggtgctca ggatcagctc atcagggaac 180
ttgaccaaatt tttagagcaag gccctttgat agtgataga gatgtttgtt ctaagcagca 240
atagaaagct tctggaatct gttccattaa gaggtgatag aaacaaaaa ttagtctgttt 300
tggagttgtt ttcagcagag tcacaatgat agcaccatta tagatatttt acagacataa 360
tctgatcttt ttgggtggat gaccagaatg tctagtgtgt tcactgagcc ctggttttga 420
cccaatatgg taattcgtga actcttagga ggccagaat atcctaacc tgtgcaagge 480
agggacccct ggactgtaac tgtctgtct gcttttggtc gtgaaggaga ctcagaggcc 540
caaacaagaa tttaggaaaa agagcaatag gatgtgtgtt aaaaaa 586
```

```
<210> 95
<211> 491
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> unsure
<222> (480)
<223> a, c, g or t
```

```
<400> 95
```

WO 02/48370

PCT/US01/51341

39

```

aataaattta acctaggaaa agaaaaagaa aattgaaaat tggagctaaa ataatttgat 60
ttttccctca acaggggttat tggctgtctt ttaagtgaat aaagagcgt atctttatgt 120
gaatttttagg catggtcata tgattaatat aaggataaag caaccaaatg ctctcagtat 180
ttattcccggt gctatttggc tgttttttag ttcattggagt attgtattgt acttggtaat 240
ttgatgcttt tgagatgtcc tttagacaga tttttaacta caggacttcc tctgtagaat 300
cgacaalgtg tttaactctc tgtggcaattg acaatgtttt tgaatgctta attgttcagt 360
agaaactccgt ggttattatt acaactttgt acattattat aaatatttta tattagttgt 420
atattccact gcagatagca accagaaaac taaaatacag aaatattaca tattagaggn 480
gattataatg g                                     491

```

```

<210> 96
<211> 631
<212> DNA
<213> Homo sapiens

```

```

<400> 96
aataaattta acctaggaaa agaaaaagaa aattgaaaat tggagctaaa ataatttgat 60
ttttccctca acaggggttat tggctgtctt ttaagtgaat aaagagcgt atctttatgt 120
gaatttttagg catggtcata tgattaatat aaggataaag caaccaaatg ctctcagtat 180
ttattcccggt gctatttggc tgttttttag ttcattggagt attgtattgt acttggtaat 240
ttgatgcttt tgagatgtcc tttagacaga tttttaacta caggacttcc tctgtagaat 300
cgacaalgtg tttaactctc tgtggcaattg acaatgtttt tgaatgctta attgttcagt 360
agaaactccgt ggttattatt acaactttgt acattattat aaatatttta tattagttgt 420
atattccact gcagatagca accagaaaac taaaatacag aaatattaca atagagagaa 480
tataatgtac aaaaaaaatc ttgggagatg agtgctttgg gtttaattct atttttactg 540
aaacacagaa ataataggat tcaaatctac ctaatttttc tatttttctg attttccact 600
ctgtatgctc tcttttgaat tttttccttg gtca                                     634

```

```

<210> 97
<211> 397
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (326)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (331)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (337)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (371)
<223> a, c, g or t

```

```

<400> 97
aataatgac caagttgttg tgccttgagt tttttgagtc tgggtttta tatctgtcaa 60
caatttttga aaattatcag ccattttatt tgaagtcttt ctctgttccc atatttcctt 120
tctttataca attagaaattg catttatatt agggagtttg atattatccc acagatccgt 180

```


WO 02/48370

PCT/US01/51341

40

```
gatgatatat ttcattttct tctttttctt ttctctagtg ttccagtttg gacgsgtttt 240
acggagcatat ctttaaggtc actaatgatt ttctcagctg tctcaagtct cctgataagg 300
ccagataaga gactatctct attatngtgt nttaanttc tagcatttcc atttttattct 360
tagagtttaa nctctctaac gaaattacco atcttat 397
```

<210> 98
<211> 342
<212> DNA
<213> Homo sapiens

```
<400> 98
ataaagatgg ggtgaggga gaaagatga caaaaggaga ggaccaggca tgagaagagg 60
aagaggagaa tgcggaggag gctgcttgcc tgetgtggga tggatggcag gggcacttcc 120
ccagactcac tttctcaga tgaataactg accagccttg tgccacagat gtgaagatag 180
ccccatagaa ctttaagagc agaccataac ttcccatgaa tgagagctac taacatttac 240
atctgaaaaa caatttggat acttacccaa gtctccaa caaagatca cactgaagct 300
ggagagacac ctcataacac ccggaataac attttttttt aa 342
```

<210> 99
<211> 873
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> {338}...(528)
<223> a, c, g or t

```
<400> 99
ataaagatgg ggtgaggga gaaagatga caaaaggaga ggaccaggca tgagaagagg 60
aagaggagaa tgcggaggag gctgcttgcc tgetgtggga tggatggcag gggcacttcc 120
ccagactcac tttctcaga tgaataactg accagccttg tgccacagat gtgaagatag 180
ccccatagaa ctttaagagc agaccataac ttcccatgaa tgagagctac taacatttac 240
atctgaaaaa caatttggat acttacccaa gtctccaa caaagatca cactgaagct 300
ggagagacac ctcataacac ccggaataac attttttttt nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
aatccctctg gtgacccattg gacaggcccc agagacaaat cttcttacct gggcaattca 600
gaaggaggcc aagaccacct ggtgacccatc aaacaggcca tctggaggca aaactcctta 660
tctggggaat tttagagtaa tcaacttcc ctagtatctg aagacggcat ctgactctga 720
taccaggact agaaagaaat catttaggca gttagtggag gtgagggaag agagaggccc 780
tctcatattg tttatttagg ccatttagtg gggtagggga agagagagac cctctcatat 840
tgttccatat tgtttttatac tcagtacctg att 873
```

<210> 100
<211> 297
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> {48}
<223> a, c, g or t

<400> 100

WO 02/48370

PCT/US01/51341

41

```

ggaaaaggcc ccccttaact tectctctag gccactcag caaatgtngc cactttgtgg 60
ccactttgtg taaggcattc cagagatctg gtgaggcacc tatctacaaa tatttataga 120
cacacattca tatatgggtt cagtcacaaa atggggctcat tctctccctt gacctatcat 180
ttagggcatt ggaacatggc tgcattgtggc tctgtttgtg agggctccag ggatggacag 240
ggaggctctg cattratttg cttttacaaa cattgcagca tgaacgtttt ttttaact 297

```

```

<210> 101
<211> 258
<212> DNA
<213> Homo sapiens

```

```

<400> 101
aatataaata cgcctttaat agtaacacct aattacctaa caccatcaaa aatgggggtgc 60
tccatgaaga agcacataat tcaaatattt gaagttttat ctttctaatt accacataga 120
tttctcttgc cccattaaaa aattagataa tcagttattc taggatagtt gttttcttcc 180
aaccaattaa ggcataatct atgtacgaga acattcagag gatgatgcct ggtcaacatt 240
tgaataaaca atcactgt
258

```

```

<210> 102
<211> 712
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (603)
<223> a, c, g or t

```

```

<400> 102
aatataaata cgcctttaat agtaacacct aattacctaa caccatcaaa aatgggggtgc 60
tccatgaaga agcacataat tcaaatattt gaagttttat ctttctaatt accacataga 120
tttctcttgc cccattaaaa aattagataa tcagttattc taggatagtt gttttcttcc 180
aaccaattaa ggcataatct atgtacgaga acattcagag gatgatgcct ggtcaacatt 240
tgaataaaca atcactgtga tgttacctct atttaagatg actccaataa aacttctatg 300
gtttgcatta ttgattgacg agactttaag catttatctt tgatagggtc aaggaaacctg 360
tottaactcc ccatctctga ccaaatataa ctgtttttot ataagotata aagcaagata 420
gccaatttta tgagaattgt cccataacta tatccatgtg agcgataggt gccctggcatg 480
aagatgcata aaggaggcag taatatacaa caactgaagc ataacctctg gagccagctot 540
tcttcagaca aatcccaatt ccattactca ctggccacct aaaaagcta cttaattcat 600
ctnccctcag tttcttcaac tgtttaatgg gtatgatcaa caaaccaact tcagtggtgt 660
atcataaata ttaataaatg agagaatgca tgtgaacaaa agctataagc aa 712

```

```

<210> 103
<211> 173
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (95)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (140)
<223> a, c, g or t

```

WO 02/48370

PCT/US01/51341

42

```

<400> 103
gaatgtggct ggtgagtagg cacttggtgt ggcagtggtg ctagtgggta agaacatggc 60
tgggtattag gcattgtggtg tggcagtgta gctgggaggg acgagcatgg ctggtaggta 120
agaacgtggc tgggagtaga agcatggcgg gtgggtggga atgtggctag tga 173

```

```

<210> 104
<211> 668
<212> DNA
<213> Homo sapiens

```

```

<400> 104
tctgaatggt ttggggaata aatctgttct tcagcaaccc taactgtctc tccaaactgc 60
ctaaaagagat ccaagtactga tgaactgtgt ctccatctct tactccctgg aaactaacca 120
cggtgtctctc ttctctctad caacaacacg gagctcagag atctaagctg ctttccatct 180
ttctcccaag cccacaggaca ctgactctgt acagagtggt ggcgtctctc tgcctctctc 240
tcattctaat cccctctctc cagctgatac acctggggag taactagtggt tctctagact 300
ccgttatgga taagaagatc aagatgttct tcaacagctc agatcacagt cctctctcta 360
taagcaagaa gctctctgtg cttagtgtta aaagccaagg cagacccgtc tctgtccctg 420
ctgggatggc tgtcactggc tgtgttgttg gctatggctg tggttcctgg gatgttcagg 480
tggaaacacac ctgcactggc cagtgacagt tgggtggact gaccactgcc cgtctgtgcc 540
acctgacctg acagggaggga ggtgagaaac tcaagtttgt gacctgaca gtaagtgaac 600
cagggtccca acaagaaat ctactcaaa cgtcccaact cattgttcc attctcgatt 660
cttgggtaat aaagacaaac ttgttaa 688

```

```

<210> 105
<211> 977
<212> DNA
<213> Homo sapiens

```

```

<400> 105
ggcttggaga ggggtcacaga ggctagtagc tgtgtggaact tgcaggcagc cccaaatgct 60
caccatgtgt cagagtcagc atgtctctgc tccccggta atgtggtcgc ctgcatctct 120
gtggccagcg ctctcgttca tcattcagtc tgatggcttg agtgactcta tgtttgtctc 180
atgtcgagac cgtattctag tgccttatct tggaggtact ggggtgacct acagatttaa 240
gaatgcaaat ctggaggtac acccagtgga ttcaagtag tctcatagaa caaaggagct 300
tatatagtga cctttgtctc atccactagt atacaccatc tgaggtctct tgaactgaaa 360
atgaatgttg aagcaaggga acagtggtgt gttcagctct cagatctcac atggcatctg 420
atttggcttg aggtgcctcc cctctctctc gtccctggc tctgggctca tggattgcca 480
gagcccaagt abggcttcgg ttttacttgc tataatatcc agaggcaatg tactagtcta 540
cctagaaaat tgtgtctcag gcattccctt gtccacattaa taagcattat ggacactacg 600
acattttatt aagtattttg ttctggtatc tacttgatta tagtaaatca tcaaaatcct 660
tatttagctc atggactctc attaaagcat gttctggaaa ccttggccat aggttaggag 720
cctgtaaagt ttgattcatt gcaagatata agtgattagc agttggtagt agtgacattg 780
atgggcccca ttaaaaggct tattggatgt ggtggtggca tagcgatagg ttggagttgg 840
aggtcagcat ggaatgtctc gatttagaac caagcttacc ttgcataac ctatagtgac 900
actctcttca tctcccaacg ccttagcat gtctcctga ggttcatact gtttgaatt 960
tcacaggctc atttata 977

```

```

<210> 106
<211> 500
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure

```

WO 02/48370

PCT/US01/51341

43

<222> (357)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (367)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (391)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (410)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (430)
 <223> a, c, g or t

<400> 106
 cagagcaggc attgaacctag atgtcttccc ctgccttcat tgggagggtg ctgagccacg 60
 gggttccacct ctgccaagg cadacctagg agactctcca tgtccagctg agaagagggg 120
 gacacctctct gtctgagact gcagctcaca ctgctgcatg ctctctggac accatctctc 180
 tgaacttggc cgcattctgc tagctctcag ctacgttctc tgacctccag ctcttctctc 240
 ttctccctctc ggttaatacca aagctctcag aacacagccc tcacttctag acagaaaggc 300
 ctccaccaggc cccacctgtg tggccctagg gtgacctcat gtacaaaac atctccnaaa 360
 atcaccctct cgtcactatg gacctagta ntatcatga gtttaacnctn atttctgtgt 420
 taatcggggn tgcagcacat ttgtgtgcag attcattgtg gotttggggg gccatttggg 480
 atctctcccc atgcacaatg 500

<210> 107
 <211> 476
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (466)
 <223> a, c, g or t

<400> 107
 gccatcttcc cactcattcc ttctcaaaag gaatgtagta ccatatagta gttagaata 60
 tagacactgg agcogatctt cttaggttcc aatagtggct ctctacttt ttaaatctca 120
 ttttctctca tctttaatt gaagatagta acaatctcat ggggttctga taactaagg 180
 ggtaatgcct gtaaatgtct tagaaaaatgc ctggacatag gaagctctaa gtttctgtct 240
 actactgtta ttatggttac tattattaat cattgcaagg aaaatgtatc aacagatgaa 300
 tttggttcaa tactgccttc tagttttgtg acottagaat ttatagggaac aaaaaagatt 360
 tgagggagg ttgggtctga tcatagagag ccttgattcc atgttttagg atgtatacac 420
 agtgagaagt ccttcaggtt ttgtctctgg gaagagtgtg gaatongaaa gttaac 476

<210> 108
 <211> 824
 <212> DNA

WO 02/48370

PCT/US01/51341

44

<213> Homo sapiens

<220>

<221> unsure

<222> (824)

<223> a, c, g or t

<400> 108

```
ataagtatgc atgcttcata tacttcattt attctttctt ccttgaagcc tctctctctt 60
attaggcaact attcatttgt ctacttggta cctgtatttt tttaatgtca ctattttgac 120
agtaaccaata aaggttaagc caotcaatta cgcaggcgcto tctctttatg ctttgggttag 180
gtgcaactgtg gcaactgagg ggaagggtcag tgttatcaag gttacctgtt attacaagta 240
gaagaaccca caaagatcag gagagagctc attttctctc attagttaga ggtaggacta 300
tacattcaaca aacaagaaac ttaaaataga tcaaaaaata gtgtcatata tgtaccocag 360
catctttcca ctcattctctt ctcaaaaagg atgtagtacc atatagtatg taagaatata 420
gacactggag ccgattctctt tgagttccaa tagtggctct tctacttttt aaatctcatt 480
ttccttcabc tttaaattga agatagtaac aatctcatgg ggttgtgata actaaggggg 540
taagtcatgt aaagtgttta gaaaatgccc ggcaatagga agctctaatg ttgtgtctac 600
taactgtatg atggttacta ttattatcca ttgcaaggaa aatglatcaa cagatgaatt 660
tggttcaata ctgccttcta gttttgtgac cttagaattt ataggacaaa aaagatttg 720
aagggagggtt gggctggabo atagagagcc ttgattccat gttttaggat gtataccag 780
tgagaagtcc ttcagggtttt ggtcctggga agagtgtgta atengaaagt taac 834
```

<210> 109

<211> 498

<212> DNA

<213> Homo sapiens

<400> 109

```
tttaaatgtg gagttaagga tgagcaactt taactgtatta aaaaatactc accagttaaa 60
aaaaaatact cttttccctt tctcgggaca cctaatacta agagaacaac tccatataaa 120
aaatgatata aaatcatac attttggag tatgtttcta actgttctga gaggtgtcat 180
ggtaaaagctg aagtgaaga tgtattttta atctgtatat atgagcaagt ataatattgat 240
gattgaagct aggtgtctgc taatacatg gccagactt tgaggaatta tagtgtaatg 300
gctgggaata caggtttgga gtcacacagt agagctgaaa gottggcttt tattttagctg 360
tgggtccttg ggcaggatc gtaactctgc tgtgcttgaa ataccaccca caccatcct 420
ttaatggggg gataataagc ctgcctatct catggggcta ttaagaattt tcagttaact 480
tttacttatg aagtgcata 498
```

<210> 110

<211> 259

<212> DNA

<213> Homo sapiens

<400> 110

```
tttaaatgtg tttagtthta gtcacttaga ttgtcttttt atggagtgac tggagtttg 60
ggagggggagc agggagggtt ttcttttttt ctttataaca ctggctaagt attttaatta 120
ctgtcataga aggaagaagc taaaagtatt gcattcacaa atattgcata gattatacaa 180
acacagaaat atatgcatat gcattgttaa aatatatgcc acatatcaac accatgtatc 240
caacttgast aaggtcatt 259
```

<210> 111

<211> 414

<212> DNA

<213> Homo sapiens

WO 02/48370

PCT/US01/51341

45

```

<400> 111
atgaaggga tgaagggaac tcaaaagtac aatgtctac ttggagcagt aagttcagta 60
gacatatcac ttgcctcatt aacatcagc atcccaaac ccagtcctgg tcagttttgc 120
ccagagtggt gtttgtatga cagggttct cctgggtatc tatacctagc cagaatcag 180
ttgcaaaagc caggccatag caaattgtcc tgcagccag atagcagaga atctgacggc 240
agcaggcaga aggaacccgt ccattgcagt aagccaagt cggccactt gctcattac 300
atcagcctc ccaaaaccca gtctgggtca gttttgccca gagtgaagtt tctagaacac 360
gggttctcct gggatctata cctagccag aatcagttcc aaaagtccaa aaga 414

```

```

<210> 112
<211> 589
<212> DNA
<213> Homo sapiens

```

```

<400> 112
ctgggcaaca ttggggagac tctgtctcta aagaaaaaa ggagagctgg tgggtgaagt 60
gtgaaggacc caggaagtac agacactggt ggtcaaaaga caagggtagg agtgcacaa 120
aatgatagtg ttggcagcat gggagctgtg ggtagagagt gagataccct aatttatgat 180
ttctgggtgg cagtaacttc taggggtgtg ctgtgggagt gggcctctga atgggttga 240
ggagaaatc attaaagatt agaaaatctt gggatttga ggttaggttg tgggattggg 300
gatacaggtt agtgttgcct ttgccagggt taacgccaa agttggcaga gaaaaataa 360
ctgacctaga cttaataaaa ggaatttggg atgacagaga agcaacagta aaaaataagg 420
ataattagat gtttgggtgt ttgcctggc tgtgtctgtc ctgtgtctgg ccaattatta 480
caatgtatct acactgtaaa taccagtaat caataataa gttttataa tagcaaaatg 540
tagtttaata aaaaaccatc ttagtctctc taccgaatat ttagttaac 589

```

```

<210> 113
<211> 471
<212> DNA
<213> Homo sapiens

```

```

<400> 113
cccaggtcgg gggtcagggt aggaaggagc tgggatccag caagcctagt gaacccagg 60
ggacagctga ctgggtcaca tccaggatgg tgatcaacag ctgcactcct ccgttctctt 120
ctcaagctac aattccagag ccttggccac ccggtgctgt tatctttcgt attccagacc 180
cctgggggtc cagcccccct ctgccttcc tttcctctca ccccttgact catcttctc 240
gctacttgct aottgagata cctaagatga tgtgtgttat ggagaggtta gagcaacagg 300
ttcagaacaa cctgttgact ttggcctagt caactgacat ttctagaatt tgggtctctt 360
attcataaag gcagtgctga ctgcttgctg atgttatcgt gaacctgaat tctctcttag 420
agtttctaag tgccttctgg ggattaccc tttaaatcct tgcagtagcc c 471

```

```

<210> 114
<211> 1032
<212> DNA
<213> Homo sapiens

```

```

<400> 114
aatgaggag ctcttgagct cctttgatga gcaccacaca gggccctctg ggaagcagta 60
agaaccctac cdaagggtca ataagaacct aaccagcctt gggatggccc ttccctttct 120
gccagggtcc ttcccatgcc aaacctcagg ccttatcttt ggtatctgtc accaccacc 180
acccccacca caccacaca gtcatgcaag ttgtaagaca gtgacagaag atttgaagaa 240
gaccaccaga gcaagggtga gcagaacatg cagacttagg gggaaagcag gogttcctac 300
caaagaatta gacctgttgg gtacccaggc tgggggtcag gtgaggagg agctgggata 360
cagcaagcct agtgaacccc aggggacagt ggaactggct acatccagga tgggtgctaa 420
cagctgcctc atcccgttc cttctcaagg gacaattcca gagccttggc caccagggtc 480
ttgtatcttt gttattoaga cccctgggg ttccagccc ttactgctt cactttctct 540

```

WO 02/48370

PCT/US01/51341

46

```

tcaccccttg aetcatcttt cctgctactt gtcacttgag atacctaga tgaatgtgtg 600
tatggagagg ttagagcacc agcttcagaa ccacctgtg actttggcct agtcacctga 650
catttctaga ctttggtgtc ttcatcaca aaggcagtg ggaactgtg ctgatgttat 720
cgtgaacctg aattcctctt tagagtcttc aagtgccttc tggggattaa ccttttaaat 780
ccttgacgta gcccaataag gtaggatatt ttgttatccc cattttacag gtaaggaaac 840
tgaggcacag agagtaattt gcacaaggtt tatggctttt tagtgagaga gccaaagatc 900
aaattaaagag tggttgagtc aggcattggt gccctgctt atagtcocag ctacttgaaa 960
gagtgaggtg ggaggatcgc ttgagccag gagttcaagt ctacagagca agacctcaac 1020
tctttaaaaa aa 1032

```

<210> 115
 <211> 440
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (428)
 <223> a, c, g or t

```

<400> 115
ggactacatc catgttccac cacaccagge tccaattaca ttttgacttc tccacttggg 60
tgtttaaaat gcttctcaaa tttaacatat cctaaagata atttctgttc tcccacaaa 120
acttgctctt tttgcatcca ttgctgtctt agttaatggc accaccatcc atactgttac 180
tttagccaga aacctttgaa acctccaat tggctttctt gattttctct gtttcacaa 240
ttattctcca cagacaggat actccaaaat gtacccaaag ccattgtctc ttatactttt 300
caactctataa aatatacata cataagagta tataaaatat attataaagt aaatatccat 360
gtatcccaac acccagggtt agnactggga acacaatatg caaaagaata atattgggac 420
ccccctancc tcatgtcata 440

```

<210> 116
 <211> 249
 <212> DNA
 <213> Homo sapiens

```

<400> 116
aaaaaaagtt ctgacaattt gtttgctttt acattttcaa atttctgaaa tgtagagata 60
attttgtttt caaatctttg taattccctg aagcaaatat ttccaagcca gttgcnaaat 120
gctgcttttag aaataattca tataaacatg ctctctctat taatcacaaq gggagatgtg 180
gagaatggat gttttatttt ttcatgtagt ttgtctctat aaaaatatta aattgctatt 240
atgattact 249

```

<210> 117
 <211> 1017
 <212> DNA
 <213> Homo sapiens

```

<400> 117
gccctttttt ggtgtgcccg ctgaatgagc actccaggct gtggagttcg ggacatgctt 60
tggtttgttg ggacatgctt gctgtcctgt cagacccaag catcgatact gtgtgtctac 120
ctgatgaaag tgctcagtat gtgtctgcat gacttgggga cactaagaaa acbbaagggg 180
ttagcaacaa agagagcttg tcaactttgt gcggaaccag ctggcatctc acagggacaa 240
cctacacact gagctgtggt gctccactta aactcgggac cctagggaac cctgtttact 300
cctgtctctc tggagcttat taggggcttg gctacccaaq ggaagaaggg gaaaatagac 360
cggagagctt atgctagagc catkttttt gtctccgtg atgcagacag agataaaact 420
gcaattttaa tgaaccttta acaatdagta caatgtttt ccttaagaa cttgttaata 480

```

WO 02/48370

PCT/US01/51341

47

```

gcatttatct ttcagaggtt cttctctctt tttgtgtatt attttataaa cttaaaggaa 540
aaagagaaaa agtcagtggt tccagcattt gctttagttt gtgacttaaa tggattataa 600
ctcttgacag ctgacattta ccaagataaa tcagtggtca tagatgtgga gcttgatgtc 660
ctctcggctc tgggaccaat ccccttggac aaaagtttct ctgtgttctt agtattctga 720
ctctggctaca gcaactttta ggaataaaa ggttacaata aaagtctctg caatttcttt 780
gctttttacat tttaaaattt gtgaaatgta gagataattt tgttttcaaa totttgtaat 840
tccctgagagc aattactttc aagccagttg caaaatgtgt ctttagaaat aattcatata 900
aacatgcttc tctatttaat cacaaggaga gatgtggaga atggatgttt tattttttca 960
gtagtttttg ctctataaaa atabtaaat gctattctga ttactaaaga taaaaaa 1017

```

```

<210> 118
<211> 332
<212> DNA
<213> Homo sapiens

```

```

<400> 118
ctgcttcac gtggattaac acattttctc cctcatctca caaggcagtt cctgtttcta 60
ttccctcttc acacaaaata acttctgtat ttgttagtaa gcaggagaa cagcttttga 120
actcaggact gtttaagac caaggtctgt gcosctgaaa taaaacatct gcaactggca 180
gattaatgaa aggcctctga aggaacaaa aaaccacaga gactgtggc agttagtagct 240
gagtttttagg gggaagaagt gttttagttt tccctgtata ctttctgtgt tagtttttaa 300
aatctacagt atttaccatt tcaaaacaaa at
332

```

```

<210> 119
<211> 344
<212> DNA
<213> Homo sapiens

```

```

<400> 119
ggcaggggga aattataggt ggctgtggtt gtaattccaa agttctgtca cgtcttcatt 60
gttaggagga aaagaattca ataactctat cagttctgct gtaaaacaaa ttagctatga 120
aattctgtgt aacactgatt ttatgtctcc attcttgagg acactgttag ttgttttca 180
tctgtatgcc ttgattagag caataaacct taaatctcct taaggaaact tagatataca 240
tcatttccag tttttatcaa atgtgaattt tttttgtcat actgcccacc taacatggga 300
tgthttctca gaattatggt cacttatgtg tttagtttcc ttaa
344

```

```

<210> 120
<211> 718
<212> DNA
<213> Homo sapiens

```

```

<400> 120
aaaaaatcat aatagtttat gatcttgaag ggtttaaaa tatttgatga agatgtcttt 60
tgaattkatt tgtaggctct ctgtgtattt taaaagctaa gttatcttgt aatcattttt 120
ttctatacct ttgtcagtaa cctcttagtg atgaataaaa aaagattagg taatcatcca 180
gcaatgggga agaagttaag gaacaaagag ctacagattaa actagttttt agaatcctaa 240
cattttctga tgaatttgaa tcatggaaaa caaatgttag cactccaaca ttgtatgcaa 300
aactaaaagt ggaatactgc ttgtatattt gaatgaattg aaaaataatt aacatccttg 360
gaactgtatg taagaaggga cttcccaagt attatagata cccccaacct cagccctttt 420
cccctgtatc totttgatca catcctaac tcatagataa cccatgtgct gaagactttc 480
agttctgtat ctteattcta gatctcctga actcaagata agaatacttt totgaacttc 540
gactgtgtat ttctggatgt tatacagaaa cctcagctca aactcagtat tccctaaacc 600
attgtttttg aaactttatg ttggatgtga aactcgtatt gtagaataac attaaaaaaa 660
gaagaatatg tatgcataat atcagagtgc attgtatgta gaagagtag gtattttc 718

```


WO 02/48370

PCT/US01/51341

48

<210> 121
 <211> 2617
 <212> DNA
 <213> Homo sapiens

<400> 121
 atgtggaatc aacctacctg tccaggaaca gatgaagaga taagaaaatg cagtgtgtat 60
 acacagtgga atgctcttca cctaaaaaaa ttcaagggaat catgttoattg cagcaacatg 120
 gtggacaatg taagaaaagc tcccgggaga agctgtacag aagctgcttc ctcagcagtc 180
 agggccaggt acgggagctg tttttaccca aggcacagggc cggccccaag tcatccaga 240
 gctgccaatg caccctctca gtgggtctct gaggaaatctt acacaagcta cttatatcag 300
 tgatcactag gataatccat agaacctttg ggaagagagt ttaagacctt totcccacca 360
 tttcagcagg ataaattcca actggattag aaaaagaaat gtttaataatg caataaagta 420
 catatttata tctgtatata aaatacagtt gatatttgcc tgggtgttag gtgtctaaag 480
 gactttctaa gcataaaggc aaaaaaaagt cataaaaatg ctatagcagt ttgagactct 540
 atgcaggaaa ggcacccatc acglgcattg atgaatctgt atctaatttt aaacaatttc 600
 caatggctgc tgtttccttt tctttgaaaa tctctggaga aatagttcct cttgtctgtg 660
 ctttctttag gcaagaattt ttaataatg atgtgtagtc tgaatcctgg ctaagkataa 720
 aocctttatt tttataacct gttcttagtg aaaaagaaac tgbgactttt tttttaattc 780
 ctttctgttg toaaaaacta caattaactc ttctgagttt cttctctggc tgaacaaaca 840
 atggtcccat tgcctttcca gggaaatcca ggcgcctcca aaaaacctca tgtttcattt 900
 ctttccagag ctcccacaaa gaatagcttg ctcttgactg tgcacatgtt agtggaaatg 960
 taaggactac tttgaaaga tgaataattt gttgttctag tgatttgaaa atagaactct 1020
 ttagataacta ttagataattg ggaagaaggc tgaagaggt aggtataacc gaagcactt 1080
 acaaatctg aataatctg tacttgattg catattatgt tatctagga acagttgggt 1140
 ttctctgagt gttaaattat ttattoactt attcacttc aagccagcta aatgatttgt 1200
 tccctgatgg caaagctctc agattgattg cacagtttat ttggttggat tgtttatgct 1260
 ctttttatta tttattctta tttaaccaat gaaaatatca ctaagttctt tggtttgttg 1320
 accgtgattg accctacttt acaaatcact gctttctctg acccagtttt ctcatatagt 1380
 ggcagtgata accctgcata cttacagata taataacatg aaagttaaag tattgggtta 1440
 tacttctctc ctatcttttt tttattttga aaaaagataa aaatggcat aatgtattag 1500
 ttaagatgga ataactatct gttgatctcc agccatttct tctctcaaat gataggaga 1560
 tttttatgtg aaactacttg tgagagatct taacaatttg tagttagaga aagcactatt 1620
 atactctttg gaaatgcaag aaacaagtta cctttggggc aacagagccc cttgtcattt 1680
 tctcaaaaga aggaagcctc agcattttga tgaatgtgtt gagattgtag aaatgatgaa 1740
 ggtgaaaag ttattcttag tlatgtttag caaaatgaaa tgaacccaaa taataaaaca 1800
 gttacaacat tgaatctctt tgggagaaaa aaaaagata gaatgctaatt gtctctcaga 1860
 acttctttaa ccagacccct aaaaaaaga gaagctttta aaaaatcata atagtttatg 1920
 atcttgagg gttaaaagt atttgatgaa gatgtctttt gaatttattt gtaggctctc 1980
 ttgtgtattt aaaaagtaag tbatcttgta atcatctttt tctatacctt tgcagtaac 2040
 ctcttagtga tgaataaaaa aagattaggt aatcatccag caatggggaa gaagttaagg 2100
 acaaaagagc tcaagattaa ctagttttta gaatctaaag atttctgcat gaatttgat 2160
 catggaaaac aaaaatgagc actccaacat ttgatgcata actaaaagt gaatttgat 2220
 ttgatatttg aatgaattga aaaaataatta acatctctgg aactgtatgt aaagaaggac 2280
 ttcaacagta ttatagatca ccccaacctc agcccttttc ccatgtatct ctttgatcac 2340
 atocctadct catagatcac caatgtgtgt aagactttca gttctgtatc ttcattctag 2400
 atctctctgaa ctcaagatca gaatactctt ctgactcttg actgtgtatt tctggatgtt 2460
 atacaagaac ctcaagctdaa actcagatct cctaaacca ttgtttttga aaactttatg 2520
 tggatgtgaa atctgtattg tagaataaca ttaaaaaaag aaagaatagt atgcacaaata 2580
 tcagagtgca ctgtatgtag caagagtagg tattttc 2617

<210> 122
 <211> 373
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure

WO 02/48370

PCT/US01/51341

49

<222> (74) .. (294)
 <223> R, C, G or T

<400> 122
 gtattataat aatggcetta atgaataaca ttctctatat tcacacttat ttgcaatata 60
 atactggcat tctnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
 tctcaaaagag ggcctgatct tcaagaatta ataaccctct caagtctcta caatctaagt 360
 caattacatt ggg 373

<210> 123
 <211> 308
 <212> DNA
 <213> Homo sapiens

<400> 123
 gctgaagcc cagagcagag ctgtctctat ggggaaggac cctgtctctc caateactct 60
 aggggttcat tgaggatgag gactgtcttc ctccatcaga ccgagagttc ccaagggaac 120
 ggggtgtctc tccctggcca gaaagggaga tccocgagag aagaggtctc gtctctctca 180
 tcagactggt agcccccaca accacaaga tatgtctact ttcatcaga ggagctcctc 240
 aagtgaggaa ggyttctccc tattttcccc ttccaggtgg gaaattctcg gccagggctc 300
 cctgtctc 308

<210> 124
 <211> 774
 <212> DNA
 <213> Homo sapiens

<400> 124
 gccacaacca aagggggcac gggagaagga caggaggggt ggtttccctc agcaagctct 60
 cagtcccaact gacactggcc caagagggct gactgtactg ggcaactcag caggagatt 120
 gtbtccgaag gccctggga aagttggtga atgcaacag caggcagcca gaggacctgc 180
 tgcaagagag accagagag atgcccagga agggcagaga agtgtgcaaa agactcagca 240
 gtgggaagga gcttggctcg tgagtgtgag gagataaccc gggccttagg ccttctctgc 300
 cccaacttc caccactgg cccagccctc tgcagcgggt aggccttagc tctctctgct 360
 ggytttgtga gagcccgagc tgcctccagt agggtaacag agtactctcc ccaggcagga 420
 aggggtggcg gctccctcc aggtaccaca gaggaaatgt tagcagctga aagcccaga 480
 gcagagctgt tctcatggg agggacctg tcttcccat catcctaggg gtctattgag 540
 gatgaggaat gtctctctcc atcagaccga gacttccaa gggcaagggc tgtctctccc 600
 tgggtcagaca gggagctccc caggggcaga ggtcctgtct cctccatcag actggtagcc 660
 cccacaacca caaagctatg tctactttca tcagaaggag ctccctaagt ggggaagggt 720
 tctccctatt ttcccttcc aggtgggaaa ttcttgcca ggttccctg tctc 774

<210> 125
 <211> 271
 <212> DNA
 <213> Homo sapiens

<400> 125
 aagtcgtacg catgggttaa aaaaaaagaa aagaaaaatc aaaaatgtac tgaaggatg 60
 cagtacacag gaagctccg cccaaatcca cctccagct tccctcttg gaggtatctg 120
 ctgtatggg ctctcaaga tactcttagc catgtctgt ttgtgcatgc ttatccctgc 180
 acagacagca gaagctgtct tggccaccaa gaccaggag catgtgtatt tgcagggtta 240
 ttgaaaaatt catttaaggt ggagaaacat a 271

WO 02/48370

PCT/US01/51341

50

<210> 126
 <211> 1950
 <212> DNA
 <213> Homo sapiens

<400> 126
 atgatgccac aggatgagcg cacttcaaa ctaggaaggaa gcttggtgag ggagcagggc 60
 agaattctct cctggactgt gagggtacat aoggtggatg tgtatggott cattgaagat 120
 gccagtcott gcattggcat ctgcagattt gaagaagtat gccccctctc tagtcttcac 180
 ggactggatt tggcaagaaa agtcctttcat cagtcagcca ttacagaact ctgggaagcc 240
 tatctggtaa cgtccatggg agggcaaaat ttgocattca gctacaagaa gtggcgttgg 300
 agacagactt tcaacttcag catcttcaga gtctgccttg actttcaagc tggggccatg 360
 gactcttcag gagctcctag ccaatggctg agaacaacgt gcttaacaca tgttctcttt 420
 ctctttgatg gccaaagcat ggcctggcca tgggatgctt ctctctccaa agggagcagg 480
 agagctggag ataccctcct tgcacaacgc agcttgaggc tccagcgctt ggtgcaagcc 540
 ccaacgggac cccaagaaag tgcaccaacc ctggagacta tggagctcta cagctgtaga 600
 gccacccagg aagtggactg caaggccctg gctctccat tcagattctg caaagagatc 660
 ctgactgggtt gggccaatgg gtccaggcat cagtcagctc tggctaaggg agctgcctgg 720
 tgcacggagc agcctaaccg ggaaccacag tctccacaga agggggcagg cgttctgaga 780
 gccacaagat cctggctgac agtctctcct ggtctgatcc taacccctg ctctgggggt 840
 gacagctctg ccgtgagcgc tgcctgggct ccgaagggca tccagagagt gatcgggaga 900
 oagcgtgacg ccggaggtgt ctggggaccc cgaacagctg gattggcggg caccataggc 960
 cgaagacccc gggaaagaga ccttgagagc ctgctgaatg agcaaaagca ctgcaaaaca 1020
 ttcataggcc atggtcctgt ttcttacagt gtcaaaaagt ctattcagcc ctgtgtcaat 1080
 ggtatctgac agatgggttg atcagagcac ctctctgtga tgtcacaaat cggggccttt 1140
 ctacgctctc taaccttgga ggttctgctc agcagctgct actggcgtct cgtctctttg 1200
 gctctgggtc tggggcaact gaaggtaaac tccctgctga gttggagcca gcagcattga 1260
 gtgggtggct gttttccage caggatttac ccaggctttt atggtctgca aagccttctc 1320
 ccaggcgctt tctcaggcat ttaattatca caaaaatgtg gccaggtaca aaattattat 1380
 tatggggaaa ctgaggccag actgtaaagt ccacaggtca ggttctttgt ggtcactctt 1440
 tgtatccctg ggcccttttg actgatttgc acatggcaga tctcaagaa cattttccag 1500
 tggatagagg ttccagaggg ccattgcagct tggccagagg gccacagcc agagaggcag 1560
 ggattctgtt ctgtctctgc caagtcacca cctcttttat ggagccagcc tgtctgtgtt 1620
 ctttgaagag agcctctgac cttaagaag ggtctcacc tttttccttt ctgttaatta 1680
 agtcgtacgc atggttaaaa aaaaagaaa agaaaatcca aaatagtact gaaggtatgc 1740
 agtcacaggg aagcctccgc ccacctccac ctccagctt cccctttgg aggtatctgc 1800
 tgtatgggac tctcaagat actctagcc atgctctgtt tgtgcatgct tatccctgca 1860
 cagacagcag aagctgtctt ggccaacaag accaggaagc attgggtatt gcaggttaat 1920
 tgaaaaatcc atttaagggt gagaaccata 1950

<210> 127
 <211> 209
 <212> DNA
 <213> Homo sapiens

<400> 127
 gttgggtgtg gtgggggtgt ttgtgttaa tgttgtttt gccagctctg gttgataaga 60
 tttattattg agaattatgc ttgtctctg agtactcctg acttagaaaa ggagcatagc 120
 cctactaaag gggacttcaa agtagaatac gtcaataacc ttttacttgc taccggttag 180
 ggctcacaac tgatgttttt aaagatctt 209

<210> 128
 <211> 495
 <212> DNA
 <213> Homo sapiens

WO 02/48370

PCT/US01/51341

51

```
<400> 338
gcttcctggg gatggaatga agcaaggatg ggggctgctt gcagagctgt gtcactcact 60
tgtattcagc ttctctgctt ctggctctct gtcttttacc nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
ctaagaaaac aatgatcacc atacatgctc tgetttccaa ctatacttcc acatccaaag 480
taaccccgaga ttcata                                     496
```

```
<210> 129
<211> 252
<212> DNA
<213> Homo sapiens
```

```
<400> 129
catttctaac atttattgac ctccagtcac aagaagtaac ccattgtcat gtcactcta 60
tgataggcta gaactatagg gtgtctctat attgatcagg tttttaaga taaaaatgaa 120
aaaaaaatcc tatccagaca aaataaatca gttgtttata tttttggagc atcagaactt 180
actttaagac ctactggta attcttttagc ctctcacatg tgataaagac attgtgctta 240
catttttita aa                                     252
```

```
<210> 130
<211> 149
<212> DNA
<213> Homo sapiens
```

```
<400> 130
atcagaatcc tgggaagggt ttgttaaac actactagcc agggtaggt aacctaaagag 60
cttttggagg ccaggtgag agggatcact tgcggccagc agagttcaag agcagccag 120
gcacacacag gagacctctt ctctacaaa                                     149
```

```
<210> 131
<211> 390
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> unsure
<222> (217)..(273)
<223> a, c, g or t
```

```
<400> 131
agcaagtagc cagcattggg aatgaaccaa actcgtagga ggcacagcc actcagtggt 60
cgggcccggg cgagctgcag gctgaacac caccacacct cttagatgtg totgtgggcc 120
atagaaatta ctagggttgt cttgggtgtg gcctcaacct gttcaacac aggtgtgctg 180
tttccattct ggaacccagt cctctgtctt ccagaannnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
gcttttggag gcccaaggtga gagggatcac ttgaggccag cagagtccaa gagcagccca 360
ggcaacacag ggagacctct tctctacaaa                                     390
```

```
<210> 132
<211> 1079
```

WO 02/48370

PCT/US01/51341

52

<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (874)
<223> a, c, g or t

<220>
<221> unsure
<222> (879)
<223> a, c, g or t

<220>
<221> unsure
<222> (885)
<223> a, c, g or t

<220>
<221> unsure
<222> (887)
<223> a, c, g or t

<220>
<221> unsure
<222> (890)
<223> a, c, g or t

<220>
<221> unsure
<222> (894)
<223> a, c, g or t

<220>
<221> unsure
<222> (896)
<223> a, c, g or t

<220>
<221> unsure
<222> (899)
<223> a, c, g or t

<220>
<221> unsure
<222> (921)
<223> a, c, g or t

<220>
<221> unsure
<222> (924)
<223> a, c, g or t

<220>
<221> unsure
<222> (926)
<223> a, c, g or t

<220>

WO 02/48370

PCT/US01/51341

53

<221> unsure
<222> (931)
<223> a, c, g or t

<220>
<221> unsure
<222> (933)
<223> a, c, g or t

<220>
<221> unsure
<222> (944)
<223> a, c, g or t

<220>
<221> unsure
<222> (950)
<223> a, c, g or t

<220>
<221> unsure
<222> (975)
<223> a, c, g or t

<220>
<221> unsure
<222> (977)
<223> a, c, g or t

<220>
<221> unsure
<222> (988)
<223> a, c, g or t

<220>
<221> unsure
<222> (993)
<223> a, c, g or t

<220>
<221> unsure
<222> (995)
<223> a, c, g or t

<220>
<221> unsure
<222> (1007)
<223> a, c, g or t

<220>
<221> unsure
<222> (1013)
<223> a, c, g or t

<220>
<221> unsure
<222> (1030)
<223> a, c, g or t

WO 02/48370

PCT/US01/51341

54

```

<220>
<221> unsure
<222> (1037)
<223> a, c, g or t

<220>
<221> unsure
<222> (1050)
<223> a, c, g or t

<220>
<221> unsure
<222> (1061)
<223> a, c, g or t

<400> 132
gggatgaana cttccittaa aagaatcctg ttgtatttta atattgttcc gggatttctt 60
gcatacttat atgctctata tgaacaatac tgaatgaac atocataatc atgacctctc 120
tctgaactcc aggcacagat atgcaactcc ctatttgaca ggtctgtctg aaacttctgt 180
gggaatccca gaggaacat ggaatcaatg gaaggtttga tttgtctc caagccagtt 240
cttcccttga cttctacat ttcaacaaat gatacccaa ccactcaatt attctagccc 300
aagatctcgg agttattctt aggttttctt ttacccctc cactggatc catcagcagg 360
ttctgttctt tttctctccc aaatataatc caagtccatg ctctctctgc tgcctctact 420
gccactatcc aagctctgag gccatcattt acatggacaa ctataaaata catgtcttaa 480
tgaatattta gcaatagagt tgcataagta aaagatttgt gtgttttatt ttgataagct 540
ttgtacacat attctcaag aggtctttct agtgtttatc gcttattata tgagaatttc 600
tgtttcttga ctctgtcacc accactgaat atcagggtca ctcttagccc atagcctctg 660
gagaattaga agtcaactcc tctgggtgag gcagctagct ccacagcaca gacttaacaa 720
gtggaacttt agcatgtatt taattccacc tcattctctt acctatgtgt ccttctgcag 780
tcacactct acacaactgt acatgaccac aatgctgtgc ataaataatt ttttagactc 840
tttgtaaatc tatatgtaaa aaatggcacc ttantttgna taagnanggn ggangncant 900
taaaattcct tttccttggg ntgnonaatt nanagacttt cotnattttn aggttctcta 960
acaaatttgg aaatnonggg gtttaacnaa ggnonatcat atatttnacc atnaaaattt 1020
tttctctggn accttaagtt tgttaaaagn acctttttat ngaaaccttt aaattttta 1079

<210> 133
<211> 303
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (295)
<223> a, c, g or t

<400> 133
ttaagtattc aattttctgt ttaaatgcca agaggtagaa attaaaggta ggcattggtg 60
tcacagtcca ctaaaaaact agtattccaa ctctatttcc ctggcacact actaaatagg 120
caaccaggga tttaaaaaat ggtttctggt gtccaggtaa gtttgcaata aaccaaaata 180
aaactgttta atactgggac cactacatta atctatgggt ctaaacagtg ctgtgaaccg 240
tggggtaagg ggcctgggga taaagttgca accatttttt ggggggttgg gggangagga 300
ggg 303

<210> 134
<211> 546
<212> DNA
<213> Homo sapiens

```

WO 02/48370

PCT/US01/51341

55

```

<400> 134
cgggcaaaatt taacaaaaa aaaaaagtaa tatgaccata attaatatca gtcaaaatat 60
tctttaaagg aaaaaaatat taataagaga actctatataa aataaagaat ataataaaaa 120
gagatcacat ttgcaaatTT acattgttta atatcatagc ctcaaaataa attgcatata 180
aattttaaaa cctatggaga aattgacaaa tccaccaaca ctgtgggaaa tttttaatat 240
atatctotta gctattaatg cataaagtag gtaaggaaaa ccaataggat gcaataaatt 300
tgaacaataa aatcaacaac ttgtatttag ttgatataca tatacageca ctgtgattta 360
gtaattggaa aatatacatt attttccaac acacacaaaa aaacacttgc aaaaatgggc 420
tgtgtottaa atttttcaaa gaactgatat catacagaa acatgttatg accataatgt 480
agttacatta gaaaatgtgg cagggaattct gattctcctt tctgtgctag ggcatacagt 540
taaatc 546

```

```

<210> 135
<211> 590
<212> DNA
<213> Homo sapiens

```

```

<400> 135
aaaaaagtaa tatgaccata attaatatca gtcaaaatat tctttaaagg aaaaaaatat 60
taataagaga actctatataa aataaagaat ataataaaaa gagctacat ttgcaaatTT 120
acattgttta atatcatagc ctcaaaataa attgcatata aattttaaaa cctatggaga 180
aattgacaaa tccaccaaca ctgtgggaaa tttttaatat atatctotta gctattaatg 240
cataaagtag gtaaggaaaa ccaataggat gcaataaatt tgaacaataa aatcaacaac 300
ttgtatttag ttgatataca tatacagaca ottgatttta gtaattggaa aatatacatt 360
attttccaac acacacaaaa aaacacttgc aaaaatgggc tgtgtottaa atttttcaaa 420
gaactgatat catacagaa acatgttatg accataatgt agttacatta gaaaatgtgg 480
cagggaattct gattctcctt tctgtgctag ggcatacagt taataacat tttcaccttc 540
cttgtattta tgagaattag ctctgtcctt atgaatgtgg gcagaagtga 590

```

```

<210> 136
<211> 165
<212> DNA
<213> Homo sapiens

```

```

<400> 136
gctcgaggcc tggcatctga gtctctctgt tcaggagaaa cactttcagc aggcattga 60
gagggtcata ggaggtagc ctggagccc ttaggaggg aggggtgttt gcagctctgg 120
gcctggcagg ctcccccct gggccagtt tcaattctgc atgca 165

```

```

<210> 137
<211> 172
<212> DNA
<213> Homo sapiens

```

```

<400> 137
tagttacagt ccttaaatat atgtcttggg tgcoctgtgg ctgtgatttt ttaaggaaaa 60
tbaacttatt ttaataaaaa taaacttaat ttaataaaa attttgttat cttaaggcaa 120
atagaaaaaa ttccacattt tttcttacag tgctcattca tcagaacctt tt 172

```

```

<210> 138
<211> 809
<212> DNA
<213> Homo sapiens

```


WO 02/48370

PCT/US01/51341

56

```

<400> 138
agtaagtaga gbatcaaaac gtctccctcc tttctctgt gatttggtct ttctccttag 60
agaatgtctt cctccaact ccaaaagaca tgcctctgt gtatagttac agtccctaaa 120
tatatgtctt ggggtccctg tgggtgtgat tttttaagg aaattaaact attttaata 180
aaataaactt aattaaaaat aaaattttgt tatctaaagc caaatagaaa aaattccaca 240
ttttttctta cagtgcctat tcatcagaa ctttttttt tctcttatt ttttttttt 300
ttggggagaa tgggtccctc ctttggtgag oatoaggggg aataagaggt acaaaadagc 360
ggtgattata cgtccacttg ggggtttgga aaactccggg gcatcattgg gattccatt 420
ttgtctcaca gctccgggag tagctaggac atactgggtt tgcacacaaa ggcggggata 480
aatttcacaa tttttctcac gagacaaagt ttgggattct tggcccaagg attgggacgg 540
ggtatctcac aaaagaaact atttcagggg cgttagaga ggtccaatg acactactt 600
atcaggggtt tccagtggag agaatgtac cctacccctt ctacatttta agtgggtgct 660
ctccctccac ctttaacctt taacattac ggaactggcg ctatcatttt aaagtcaact 720
aaactggact ttggaattct ttaaaccttc agctccggga tccaaactaa aattctaggg 780
aaggcctaab ggaaggtaga agctcagc 809

```

```

<210> 139
<211> 294
<212> DNA
<213> Homo sapiens

```

```

<400> 139
gtctttttta ttcatagtaa ccttgcaaaa caaacatata gaacagagac attatggaga 60
ctgagggatt gattttatgt attgattatg tatgtaagtc cggataacat cctcgggtca 120
ggaaattgca gaaaaaagat tgggaatcag aacagcagaa aggtattttt ggaagggtaa 180
tttactgatt ttctgtttta aattgttgac attgccttcg cgggtggaaa tgaattactt 240
atggaatctt ggcaggaaca caatttttaa aattagaaaa ttagtctctc ttat 294

```

```

<210> 140
<211> 1056
<212> DNA
<213> Homo sapiens

```

```

<400> 140
aactcaaac attttaatta ttttttgcct atttttggag aaacctctcc ctttgacac 60
tattatgaac attctaaaac ttaaatattg gaaaacaaaa cctcggggaga tagattgtaa 120
ttttattcca tggggaagggt gttaaaccag ctttgagtt tgaattttat tcttaaggc 180
ctgtcagttc ttaoctggat gtcaaatga tttttaattc caactgctgt agacctcact 240
ctgtgggaac tagaataaat gtccaactgc cgtccagttc ggcgacattc cagcogttcc 300
cccacccac gataacggcc tgaactcttc tcaattcatg acagcccaat ctacacataa 360
cctttctctt ctggcacagg tctctccagc agagagggat cctgcccctc ccttccact 420
ctccagcata cagaccagca ggaagccaca agagggaaaa acaaaagcct tctgtataag 480
gcctatgaaa ggaccatggg ccagccctag aatctgctgc cctacacaaa cagtatttct 540
caaatgatag ttccacattt acttaataag gaggaactat tttctaat taaaaattgt 600
gttctgcaca gattcacata agtaattcat ttccacggc gaaggcaatg tcaacaattt 660
aaaaagaaaa atcagtaaat tacccttcca aaaaacact tctgtgttcc tgattccca 720
tctttttctt gcaatttctt gaaccagaga ttttatcggg acttccatac ataataata 780
cataaaatca atctccaagt ctccataatg tctctgttct atatgtttgt ttbgcagggt 840
tactatgaat gaaaaagaca atttcatgaa tgcagaaat ctgggggatcg tgtttgggc 900
caactctgat aggcctccctg aggaacagcac cctgaccacc ctgcatgata tgcggtacca 960
aagctgatt gtgcagattt taatagaaaa cgaagacgtt ttatttcaat ccactcaggga 1020
aatgagctga atggccctag caccatccaa gttgac 1056

```

```

<210> 141
<211> 968
<212> DNA

```

WO 02/48370

PCT/US01/51341

57

<213> Homo sapiens

<220>

<221> unsure

<222> (319)

<223> a, c, g or t

<220>

<221> unsure

<222> (497)

<223> a, c, g or t

<400> 141

```
acgagatgto ccagtaacct aaaattatcc agtcggctctt cttactttac aactaagaaa 60
aatlaaggctt agaaagaggg atcgccagaa actttggcag ctggattgoc tctgcttggt 120
cctctaagcc atacctaaat tctgcagtaa atacttaact ttttaatact gaaattgctt 180
cagataaact tgaccagtgga taccgtaaaa taattagact attggactaa tgggttaaca 240
caagtggtctt taaaagctct gcttaaaaaa caatttttat tttagaaaaa tggaaaaata 300
aaaacatctct caaaatttng gagcctgaag gggctgtttg ttctatatac ggataatctt 360
tgaagaaggca agtcctgtat gtatttttca ttgttgaaa gaagattggt tatcaatagg 420
cttgcataca taatttgctt taaagttctt tcaagttttt attcaataaa acctattgat 480
ttggaaacttt aaaaaaaaaa acaacaaaaa aatcttttca ggggttttga atttcaagtg 540
gttttttaag gggagcaata gtttgcaatt taccaaaggo ttctccagat aattttctaa 600
atgtttctac ttaaaaataa agctatttaa taataagctg tcatgggac catttgaaag 660
cagggaaaat agaaaatttt tatgttaag ggaagaactt atcttttaa ttttatgga 720
taacagagtc tgcaggtctt aactcatttc agcctgtcaa atgtgcaatt aaaaatgaat 780
ttctaatatg tattcaaatg aggcctctata gtgaatagac aatcaactct ctaagtcttt 840
tcccagttaa ttgtttttaa agtggtgtac tctcttgcaa gaacgtttaa aagttaaagc 900
ttgttaactgt taacatctaa tgtatttaata taagccattt gttttttacc atttttttaa 960
gcccgtat
```

<210> 142

<211> 1466

<212> DNA

<213> Homo sapiens

<400> 142

```
gaaaatttga gtatcttttt gaaattttta attgaaattt ggatagagat ggttatggag 60
agaaatcaaa caactggaa atgtgtttga tatcacttaa aagtataaa attttaagtt 120
gaatctggct agtttgcatt ggcctatttg taagaatat caagactctc tgagaaaaat 180
gaaaagtga taataaatg cttaaaatct ggtactctcg agttaaggtt ttgctctttg 240
agcttaatcc aatttgggat gatttttcat cctagggtct ttgttttcc btttttattt 300
ttattttttc tttttttagg ggaaggggac ttgctttctt ttccaaaaag gtgaatcctt 360
cttgtaggac ataggtaaaa aaacaaagc tgaatatata gttttgaata tagatagcta 420
attccctggg atataatata ctttcaattt tttttttttt ttgggcccag tctgcttttg 480
gatgtttcaa aagctcgaac gagatgtccc agtaacctaa aattatccag tcggtctctc 540
tactttacaa ctaagaaaaa taaggcttag aaagagggat tgcagaaac tttggcaagc 600
ggatggcctg tgcctgttcc tctaagccat acctaaactc tgcagtaaat acttaacttt 660
ttaataggga aattgcttca agataaactg accagtata cggtaaaaaa atttttattt 720
tggactaatg gtttaacaca agtggcttta aaragctgc ttaaaaaaa atttttattt 780
agaaaaata gaaaaataa aacatcttca aaatttagga gactgaagg gctgtttggt 840
tcataatgg ataatctttg aaaaggaag tctgttatgt atttttcat ttgtgaaga 900
agattgggta ccagtaggct tgcacacata atttgctttt aagttcttcc aagtttttat 960
gcaataaaac ctattgattt ggaactttba aaaaaaaa cacaacaaaa tactttcagg 1020
gttttgtaat ttcaagtgtt ttttcaagg gagaatagt ttgcaattta ccaagggtct 1080
ctccagataa ttttttaaat gttctacttt aaaaaaaa gctattata ataaagcttg 1140
atgggatcca ttgaagaca gggaaatag aaattttt ttgtaaagg aagaacttat 1200
ccttttaatt ttatggacta acagagctcg caggctctaa ctcatctcag cctgtcaaat 1260
```

WO 02/48370

PCT/US01/51341

58

```

gtgcaattaa aaatgaattt tataattgta ttcaaatgag ggtctatagt gaatacagaa 1320
tcactctttt aagttttttt ccagtttaatt tgtttaaaag tgttgtaact tcttgaaga 1380
acgtttaaaa gtttaagtctt gtaaatgtta acatctaatt tattaatata agccatttgt 1440
tttttaacct ttttttaagg cegtat 1466

```

```

<210> 143
<211> 306
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (289)
<223> a, c, g or t

```

```

<400> 143
gacacagcct atctcaaga gagatgagaa gagccaggcc cctctcttct tctctccatg 60
ctgttagctc accagggcag atcttgacct caaagaatgc cgtcttccct tctggagctg 120
gtctctgtgat gtgaacctgg ctatcttcaa ttcaacaggat agggagtaag acatttcatt 180
ttggccttag gtccaagcca tcttcttcaa tgtagctact actagagagc ccaaatgaa 240
gcaataaatt ggctcccat ttggcaattt gtgtcttttt cagaagagag aagggttagt 300
aatcac 306

```

```

<210> 144
<211> 494
<212> DNA
<213> Homo sapiens

```

```

<400> 144
gacacagcct atctcaaga gagatgagaa gagccaggcc cctctcttct tctctccatg 60
ctgttagctc accagggcag atcttgacct caaagaatgc cgtcttccct tctggagctg 120
gtctctgtgat gtgaacctgg ctatcttcaa ttcaacaggat agggagtaag acatttcatt 180
ttggccttag gtccaagcca tcttcttcaa tgtagctact actagagagc ccaaatgaa 240
gcaataaatt ggctcccat ttggcaattt gtgtcttttt cagaagagag aagggttagt 300
taatacagcct ttttaagtac cagcatgcag cattacaacg ttctcaaggc ctgcaagcca 360
tagggtttct gtcttccctg tattggcctt gtaatctctg accatgatta gggttaagat 420
taagagactc ccaggacagg aaacggaaaa catcagattg tgtatggaat gaacctctct 480
ggctggatgt ggty 494

```

```

<210> 145
<211> 174
<212> DNA
<213> Homo sapiens

```

```

<400> 145
gtggaaacaac tctatgccat aaaatttctt atttcacagt taaatgaaca tatttgbgtt 60
atgtcacttt ctttttagctt gcatttcctt tataggaagg ccattttagg agtctcgggg 120
cattttgact caacttctta aatcatttat tctattcaaa aaaggtttat tgaa 174

```

```

<210> 146
<211> 445
<212> DNA
<213> Homo sapiens

```

```

<220>

```

WO 02/48370

PCT/US01/51341

59

```

<221> unsure
<222> (371)
<223> a, c, g or t

<220>
<221> unsure
<222> (391)
<223> a, c, g or t

<220>
<221> unsure
<222> (406)
<223> a, c, g or t

<220>
<221> unsure
<222> (427)
<223> a, c, g or t

<400> 145
tgatttttaa caattgtgtg tgtgcaccca gtaaacatc tctacaatcg atctagaaaca 60
ttttcatcac ttcaagtgtt ctogtatatt ccttccacgc taacccatga tccccaccc 120
tggccatagg aaccgcgtga tccatcttct atcactttag attgaatttg tctttctctac 180
tgttttatat aaagaattta cctcctttta gtctatcaa attctggtgc acccttaaaa 240
aacaattttt aggtattacc ataaaacott ccatgacatt ctctgcttta tcttctctgt 300
gctactttgt ccattccatt ttgcattgta atgtatttct gtacatgta tatcaactaaa 360
ctgtctctct cttgaaggga gggacatgtg ntcacatctc atttcaagg cttatcacaga 420
aactganaca tagtagatgc ttact 445

<210> 147
<211> 734
<212> DNA
<213> Homo sapiens

<400> 147
tgatttttaa caattgtgtg tgtgcaccca gtaaacatc tctacaatcg atctagaaaca 60
ttttcatcac ttcaagtgtt ctogtatatt ccttccacgc taacccatga tccccaccc 120
tggccatagg aaccgcgtga tccatcttct atcactttag attgaatttg tctttctctac 180
tgttttatat aaagaattta cctcctttta gtctatcaa attctggtgc acccttaaaa 240
aacaattttt aggtattacc ataaaacott ccatgacatt ctctgcttta tcttctctgt 300
gctactttgt ccattccatt ttgcattgta atgtatttct gtacatgta tatcaactaaa 360
ctgtctctct cttgaaggga gggacatgtg ttcactctc tatbttcaag gcttatcaaa 420
gaaactgaaa catagtagat gcttacttgg gaattattat tctcaaaata gaaaaacac 480
cagcaaatcg catcttatat tagtctttag aattagatc aaagcctaact tattatgaca 540
cttgaacatc taaataaact agaaaacaaa gaattaaaag ttttatgata aagccagaaa 600
ctttttatca tgaccatttt taaatactga catttcagat taattggggg cagatgatat 660
atgaaattat agtttatact gtgaatttct aatacttcag ttgtgttaga taaactgata 720
gttcgtcaca tttt 734

<210> 148
<211> 29
<212> PRT
<213> Homo sapiens

<400> 148
Met Leu Lys Ile Ile Asp Lys Leu Tyr Phe Ser Tyr Leu His Ser Ala
1 5 10 15

```

WO 02/48370

PCT/US01/51341

60

Asp Ile Leu Cys Asn Thr Glu Ser Tyr Thr Leu Ser Met
 20 25

<210> 149
 <211> 87
 <212> PRT
 <213> Homo sapiens

<400> 149
 Met Gly Trp His Glu Ile Gln Ile Pro Val Leu Ile Phe Leu Leu Ala
 1 5 10 15

Val Tyr His Arg Thr Ser His Phe Thr Ser Leu Pro Leu Gly Pro Gln
 20 25 30

Phe Ser Val Phe Leu Ile Tyr Lys Tyr Ser His Pro Ala Phe Arg Gln
 35 40 45

Val Leu Arg Leu Asn Lys Glu Phe Asn Leu Leu Trp Leu His Ile Lys
 50 55 60

His Ile Leu Val Ser Val Cys Leu Val Ile Ser Asn Ala Asn Ile Leu
 65 70 75 80

Ser Ala Pro Cys Pro Glu Cys
 85

<210> 150
 <211> 45
 <212> PRT
 <213> Homo sapiens

<400> 150
 Ser Ser Val Ala Leu Ala Leu Gly Ala Leu Thr Val Trp His Ala Val
 1 5 10 15

Leu Ile Ser Arg Gly Glu Thr Ser Ile Glu Arg His Ile Asn Lys Lys
 20 25 30

Glu Arg Arg Arg Leu Gln Ala Lys Gly Arg Val Ser Arg
 35 40 45

<210> 151
 <211> 152
 <212> PRT
 <213> Homo sapiens

<400> 151
 Met Val Pro Glu Val Leu Ile Leu Cys His Gly Leu Ala Val Trp Lys
 1 5 10 15

Trp Phe Pro Gly Leu Ala Val Leu Arg Ile Pro Gly Cys Val Thr Gly
 20 25 30

Asn Lys Pro Phe Asn Leu Pro Gly Thr Val Phe Phe Cys Lys Met Arg

WO 02/48370

PCT/US01/51341

61

35	40	45
Gly Leu Gly Ala Ser Phe Leu Arg Pro Trp Gly Leu Val Ala Glu Phe		
50	55	60
Ile Ser Pro Thr Pro Cys Pro Ser Ser Tyr Gly Ser Thr His Lys Ala		
65	70	75
Phe His Ser His Lys Glu Lys Ala His Lys Val Pro Gln Pro Pro His		
85	90	95
Thr Gln Glu Pro His Leu His Pro Ser Leu Lys Ala Arg Leu Pro Leu		
100	105	110
Pro Gln His Thr Gln Val Leu Leu Gly Leu Pro Ala Leu Phe Ser Ser		
115	120	125
Ser Pro Glu Trp Asn Gly Pro Ala Met Ala Ser Gln Arg Thr Ala Ser		
130	135	140
Trp Gln Ser Trp Glu Trp Val Glu		
145	150	

<210> 152
 <211> 29
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (14)

<220>
 <221> UNSURE
 <222> (21)

<400> 152
 Met Gly Leu Arg Val Leu Leu Leu Gly Leu Ser Leu Xaa Met Ser
 1 5 10 15
 Gln Lys Pro Leu Xaa Gln Arg Pro Thr Ala Leu Gly Pro
 20 25

<210> 153
 <211> 46
 <212> PRT
 <213> Homo sapiens

<400> 153
 Met Phe Leu Val Glu His Lys Val Cys Ser Gly Asn Thr Gln Val Ser
 1 5 10 15
 Ile Lys Cys Leu Pro Val Val Ser Glu Lys Phe Val Met Lys Tyr Phe
 20 25 30
 Gly Asn Arg Cys Ile Val Ser Val Gly Gly Ala Asp Gln Phe
 35 40 45

WO 02/48370

PCT/US01/51341

62

<210> 154
 <211> 34
 <212> PRT
 <213> Homo sapiens

<400> 154
 Met Thr His Ser Glu Leu Leu Leu Val Ile Thr Ile Asn His Lys Met
 1 5 10 15

Pro Gln Gly Pro Arg Val Thr Asn Trp Glu Pro Pro Pro Leu Thr Arg
 20 25 30

Ile Thr

<210> 155
 <211> 99
 <212> PRT
 <213> Homo sapiens

<400> 155
 Met Asp Ser Phe Leu Leu Leu Arg Gln Arg Glu Gly Gly Lys Arg Asn
 1 5 10 15

Phe Lys Arg Asn Leu Gln Thr Cys Cys Ala Val Gly Pro Thr Gly Ile
 20 25 30

His Gly Gly Glu Thr Asn Ser Ile Met Leu Leu Gln Ile Leu Leu Lys
 35 40 45

Lys Gly Phe Asn Cys Leu Thr Lys Tyr Ser Ser Phe Phe His Leu Leu
 50 55 60

Thr Leu Gln Pro Asn Gln Val Pro His Thr Thr Gly Arg Cys Arg Glu
 65 70 75 80

Ile Pro Gln Pro Glu Lys Ile Ile His Ala Gly Gln Arg Gln Lys Phe
 85 90 95

Thr Pro Gly

<210> 156
 <211> 55
 <212> PRT
 <213> Homo sapiens

<400> 156
 Met Gln Phe Leu Leu Cys Leu Ser Leu Leu Asp Phe Phe Ser Ser Thr
 1 5 10 15

Tyr Lys His Ala Val Met Ser Pro Asn Gln Lys Lys Cys Lys Asn Pro
 20 25 30

Phe Ser Pro Met Leu Thr His His Pro Ala Val Val Leu Phe Leu Pro

WO 02/48370

PCT/US01/51341

63

35 40 45

Phe Thr Leu Leu Tyr Tyr Ser
50 55

<210> 157
<211> 59
<212> PRT
<213> Homo sapiens

<400> 157
Met Leu Gln Val Asp Val Cys Thr Leu Met Val Arg Thr Trp Ser Ser
1 5 10 15

Trp Pro Cys Trp Val Phe Ala Lys Glu Thr Val Leu Cys Ser Trp Gly
20 25 30

Arg Phe His His Leu Ile Arg Ala Val Val Pro Thr Trp Cys Ser Leu
35 40 45

Asp His Leu Tyr Lys Met Phe Ile Gly Gln Gly
50 55

<210> 158
<211> 62
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (41)

<220>
<221> UNSURE
<222> (57)

<400> 158
Met Thr Lys Arg Met Glu Lys Cys Leu Asn Ile Tyr Lys Arg Leu Asp
1 5 10 15

Val Tyr Arg Gln Ile Val Ser Lys Gly His Arg Ile Val Arg Asn Ser
20 25 30

Val Ile Leu Phe Cys Val Ile Asn Xaa Pro Phe Leu Tyr Pro Phe Thr
35 40 45

Leu Ile Ile Asp Ile His His Phe Xaa Val Ile Ile Gln Leu
50 55 60

<210> 159
<211> 47
<212> PRT
<213> Homo sapiens

<400> 159
His Leu Asn Arg Phe Ala Asn Ser Val Lys Val Phe Thr Arg Arg His

WO 02/48370

PCT/US01/51341

54

1	5	10	15
Ala Phe Val Lys Lys Phe Phe Arg Gly Ser Ala Cys Asn Cys Ala Glu			
	20	25	30
Ser Ala Leu Leu Ser Ser Gln Leu Ala His Cys Val Gly Arg Trp			
	35	40	45

<210> 160
 <211> 43
 <212> PRT
 <213> Homo sapiens

<400> 160
 Met Gln Glu Ala Glu Gly Arg Leu Asn Lys Pro Gln Gly Gly Arg Val
 1 5 10 15
 Gly Ala Glu Arg Val Gly Asn Ile Phe Phe Leu Leu Leu Asn Ser Arg
 20 25 30
 Lys Ala Lys Thr Gln Ser Lys Leu Phe Leu Ser
 35 40

<210> 161
 <211> 62
 <212> PRT
 <213> Homo sapiens

<400> 161
 Met Phe Gly Ile Leu Glu Lys Ser Ser Lys Tyr Val His Leu Glu Gly
 1 5 10 15
 Ser Leu Lys His Pro Val Ile Lys Leu Val Ser Ile Ser Val Val Lys
 20 25 30
 Asp Glu Tyr Ser Leu Ile Asn Lys Arg Asn Lys Tyr Leu Asn Ser Leu
 35 40 45
 Thr Ser Ile Leu Asn Arg Phe Cys Gly Gln Met Arg Leu Pro
 50 55 60

<210> 162
 <211> 78
 <212> PRT
 <213> Homo sapiens

<400> 162
 Met Thr Pro Ala Leu Ala Ala Trp His Val Leu Ile His Pro Asn Val
 1 5 10 15
 Cys Phe Leu Ala Pro Ala Asp Ser Leu Glu Gly Ser Ile Lys Glu Asp
 20 25 30
 Trp Val Asn Met Asp Leu Glu Asn Ala His Leu Gln Arg Glu Asn Gly
 35 40 45

WO 02/48370

PCT/US01/51341

65

Gly Trp Ala Ala Phe Pro Ser Pro Ala Pro Val Pro Gly Ile Trp Pro
 50 55 60

Arg Ser Ala Ser Val Cys Phe Gly Ala Lys Leu Gln Ala Pro
 65 70 75

<210> 163
 <211> 51
 <212> PRT
 <213> Homo sapiens

<400> 163
 Met Ser Ser Trp Ile Pro Phe Ile Ile Thr Pro Leu Phe Ser Gly Ile
 1 5 10 15

Arg Leu Glu Ala Trp Cys Gln Phe Tyr Ser Ser Leu Tyr Pro Phe Ile
 20 25 30

His Phe Leu Ser Ile Leu Phe Pro Lys Tyr Phe Phe Ser Ala Pro Ser
 35 40 45

Pro Ala Ala
 50

<210> 164
 <211> 27
 <212> PRT
 <213> Homo sapiens

<400> 164
 Met Gly Ile Ile Pro Lys Cys Met Phe Leu Leu Gln Ser Arg Leu Met
 1 5 10 15

Gly Val Ile Thr Asn Thr Ser Leu Leu Leu His
 20 25

<210> 165
 <211> 52
 <212> PRT
 <213> Homo sapiens

<400> 165
 Met Lys Val Leu Lys Tyr His Asn Glu Ala Cys Gly Phe Tyr Ser Val
 1 5 10 15

Val Trp Met Leu Ser Ser Ser Ile Pro Trp Met Pro Thr Gly Met His
 20 25 30

Cys Leu Ile Leu Glu Phe Lys Arg Trp Pro Gln Thr Val Arg Leu Ser
 35 40 45

Met Trp Pro His
 50

<210> 166

WO 02/48370

PCT/US01/51341

66

<211> 47
 <212> PRT
 <213> Homo sapiens

<400> 166
 Met Gly Arg Lys Ser Thr Asn Lys Thr Ala Cys Thr His Ile Asn Thr
 1 5 10 15
 Tyr Val Ser Thr Asn Asp Lys Leu Tyr Leu Tyr Arg Ala Trp Glu Gly
 20 25 30
 Ser Tyr Ile Thr Leu His Val Ser His Pro Pro His Thr Ser Arg
 35 40 45

<210> 167
 <211> 42
 <212> PRT
 <213> Homo sapiens

<400> 167
 Met Cys Trp Gly Tyr Phe Ser Ile Ser Lys Lys Phe Pro Asn Leu Thr
 1 5 10 15
 Ser Val Leu Met Asn Leu Gly Thr Asp Leu Ala Val Arg Pro Thr Ser
 20 25 30
 Ile Phe Pro Thr Asp Ser Ile Leu Leu Glu
 35 40

<210> 168
 <211> 55
 <212> PRT
 <213> Homo sapiens

<400> 168
 Met Asn Lys Ile Lys Gly Lys Ser Val Leu Phe Tyr Met Pro Glu Thr
 1 5 10 15
 Ser Arg Ile Phe Arg Lys Val Gln Phe Lys Glu Asn Gln Ala Ala Leu
 20 25 30
 Asp Ser Thr Asn Lys Asn Val Ser Leu Ser Glu Glu Leu Val Asn Gln
 35 40 45
 Gly Thr Gln Ser Ala Phe Ser
 50 55

<210> 169
 <211> 24
 <212> PRT
 <213> Homo sapiens

<400> 169
 Met Met His Met Gln Leu Ile Ser Glu Phe Ser Cys Leu Cys Cys Phe
 1 5 10 15

WO 02/48370

PCT/US01/51341

67

Phe Phe Leu Gly Ile Tyr Ile Lys
20

<210> 170
<211> 68
<212> PRT
<213> Homo sapiens

<400> 170
Met Ile His Leu Ser Glu Val Ser Gly His Leu Lys Glu Arg Lys Gly
1 5 10 15
Lys Ala Ser Cys Gln Lys Gln Lys His Val Leu Tyr Lys Arg Phe Lys
20 25 30
Asn Gln Asn Gly Ile Arg Leu Ser Asn Cys Lys Arg Gln Ser Ser Ala
35 40 45
Phe Lys Ile Leu Arg Lys Asn Asn Val Tyr Ile Lys Ile Phe Ile Ile
50 55 60
Ile Phe Asn Phe
65

<210> 171
<211> 100
<212> PRT
<213> Homo sapiens

<400> 171
Ser Phe Ala Phe Phe Ser Leu Arg Gln Ser Leu Thr Leu Ser Pro
1 5 10 15
Arg Leu Glu Cys Ser Gly Thr Ile Ser Ala His Cys Asn Leu Cys Leu
20 25 30
Leu Gly Ser Ser Asn Ser Ser Ala Ser Ala Ser Gln Val Ala Gly Ile
35 40 45
Thr Gly Thr His His His Ala Gln Val Ile Phe Ile Phe Phe Ile Glu
50 55 60
Met Gly Phe Arg His Ile Gly Gln Ala Gly Leu Lys Leu Leu Thr Ser
65 70 75 80
Gly Asp Pro Pro Ala Ser Ala Ser Glu Ser Ala Gly Ile Thr Gly Val
85 90 95
Arg His His Thr
100

<210> 172
<211> 58
<212> PRT
<213> Homo sapiens

WO 02/48370

PCT/US01/51341

62

<400> 172
 Met Glu Cys Leu Ser Ile Asn Leu Thr Lys Asn Val Ser Tyr Leu Tyr
 1 5 10 15
 Thr Gly Pro Leu Asn Thr Ser Glu Thr Lys Leu Lys Ser Tyr Leu Ile
 20 25 30
 Gly Asn Gln Phe Pro Pro Arg Phe Ile Tyr Arg Val Ser Glu Ile Pro
 35 40 45
 Ile Lys Ile Ser Ala Arg Ser Leu Arg Asn
 50 55

<210> 173
 <211> 47
 <212> PRT
 <213> Homo sapiens

<400> 173
 Met Asp Lys Glu Glu Ser Ala Val Leu Val Gly Gly Ser Ile Leu Pro
 1 5 10 15
 Asp Lys Leu Phe Leu Val Gly Phe Thr Asp Thr Ser Pro Asp Leu Leu
 20 25 30
 Pro Ala Ala Thr Val Cys Phe Tyr Asp Ala Cys His His Asp Ile
 35 40 45

<210> 174
 <211> 106
 <212> PRT
 <213> Homo sapiens

<400> 174
 Met Thr His Val Gln Leu His Ala Leu Asp Leu Leu Leu Lys Asp Glu
 1 5 10 15
 His Lys Ser Glu Ile Ser Thr Pro Trp Gln Pro Tyr Tyr Gln Leu Leu
 20 25 30
 Ile Cys Ser Pro His Val Ser Thr Pro Phe Leu Ala Thr Ser Phe Cys
 35 40 45
 Pro Ser His Ile Asn Thr Cys Gly Gln Trp Leu Thr Met Leu Lys Leu
 50 55 60
 Lys Leu Tyr Pro Asp Glu Ile Leu Lys Arg Asn His Leu Cys Ser Ser
 65 70 75 80
 Val Leu Thr Gln Glu Ser Gln His Val Phe Leu Phe Gln Glu Thr Ile
 85 90 95
 Ile Ile Cys Thr Asn Ile Tyr Pro Asp Asn
 100 105

<210> 175

WO 02/48370

PCT/US01/51341

69

<211> 35
 <212> PRT
 <213> Homo sapiens

<400> 175
 Met Ser Met Leu Arg Lys Gly Leu Lys Ser Phe Phe Ser Val Cys Val
 1 5 10 15
 Leu Pro Ser Glu Pro Asn Ile Gly Ile Ser Ala Ser Lys Ile Pro Gln
 20 25 30
 Gly Gln Glu
 35

<210> 176
 <211> 54
 <212> PRT
 <213> Homo sapiens

<400> 176
 Met Ser Ser Ser Pro Leu Val Ser Ala Lys Phe Ser Phe Leu Phe His
 1 5 10 15
 Glu Gly Arg Ala Pro Ser Leu Phe His Pro Leu Met Thr Ser Gln Pro
 20 25 30
 Leu Glu Phe Cys Leu Met Met Asp Phe Ser Glu Ile Cys Leu Cys Asn
 35 40 45
 Glu Asp Lys Asp Ser Gly
 50

<210> 177
 <211> 20
 <212> PRT
 <213> Homo sapiens

<400> 177
 Met Arg Pro Leu Lys Met Ile Arg Thr Ala Lys Lys Leu Phe Val Tyr
 1 5 10 15
 Leu Gly Ser Tyr
 20

<210> 178
 <211> 66
 <212> PRT
 <213> Homo sapiens

<400> 178
 Met Met Tyr Tyr Pro Asp Asp Leu Trp Asn Leu Leu Arg Asn Arg Asp
 1 5 10 15
 Cys Val Ala Phe Leu Ile Met Gly Thr Gly Pro Ser Leu Leu Arg Leu
 20 25 30

WO 02/48370

PCT/US01/51341

70

Pro Met Cys Val Gly Thr Glu Leu Leu Trp His Ser Ser Ser Arg Leu
 35 40 45

Met Glu Leu Ser Ser Ser Glu Ala Ser Trp Val Val His Ala Asn Leu
 50 55 60

Val Leu
 65

<210> 179
 <211> 70
 <212> PRT
 <213> Homo sapiens

<400> 179
 Met Cys Val Ile Tyr Gln Arg Gly Ile Cys Asp Glu Lys Lys Asn Leu
 1 5 10 15

Lys Cys Pro Gln Met Phe Gln Leu Ser Glu Thr Glu Lys Thr Leu Thr
 20 25 30

Ser Val Phe Arg Ile Ile Val Ser Asn Ile Leu Lys Ile Asp Val Ser
 35 40 45

Ser Val Met Ile Phe Leu Arg Leu His Gln Arg Thr Ser Leu Asn Leu
 50 55 60

Ser Val Ile Gln Asn Gln
 65 70

<210> 180
 <211> 30
 <212> PRT
 <213> Homo sapiens

<400> 180
 Met Asn Pro Val Cys Trp Val Gly Phe Gly Glu Val Asn Ile Glu His
 1 5 10 15

Met Glu Phe Lys Tyr Ile Glu Met Asp Thr Val Ile Glu Met
 20 25 30

<210> 181
 <211> 55
 <212> PRT
 <213> Homo sapiens

<400> 181
 Met His Ala Cys Gly Ser Leu Arg Leu Asp Lys Asp Pro Thr Thr Leu
 1 5 10 15

Leu Cys Val Asn Thr Arg Cys Thr Arg Ser His Leu Pro Gly Ala Gly
 20 25 30

Gly Trp Trp Arg Lys Val Lys Ser Gln Gln Thr Val His Arg Thr Tyr
 35 40 45

WO 02/48370

PCT/US01/51341

71

Ser Ala Thr Gly Lys Lys Ser
50 55

<210> 182
<211> 16
<212> PRT
<213> Homo sapiens

<400> 182
Met Pro Ala Leu Arg Glu Ala Phe Pro Gln Ala Pro Leu Ala Leu Ala
1 5 10 15

<210> 183
<211> 48
<212> PRT
<213> Homo sapiens

<400> 183
Met Thr Phe Gln Lys Leu Met Ile Leu His Ile His Asp Gln Met Phe
1 5 10 15

Ser Leu Met Glu Ala Ser Asp Val Cys Ser His Gln Ile Arg Phe Lys
20 25 30

Met Ser Val Ser Ser Lys Ser Ser Lys Thr Ser Pro Ser His Gln Lys
35 40 45

<210> 184
<211> 55
<212> PRT
<213> Homo sapiens

<400> 184
Met Ser Val Leu Lys Arg Phe Leu Lys Pro Ser Leu Ser Ile Ala Lys
1 5 10 15

Thr Cys Tyr Val His Tyr Pro Pro Asn Ser Tyr Leu Lys Thr Thr Pro
20 25 30

Lys Met Leu Tyr Phe Val Phe Lys Val Arg Glu Glu Asn Arg Gly Glu
35 40 45

Val Phe Leu Cys Ser Phe Pro
50 55

<210> 185
<211> 14
<212> PRT
<213> Homo sapiens

<400> 185

WO 02/48370

PCT/US01/51341

72

Met Trp Leu Arg Asp Leu Asn Tyr Lys Ile Ala Arg Leu Asp
 1 5 10

<210> 186
 <211> 42
 <212> PRT
 <213> Homo sapiens

<400> 186
 Met Met Phe Phe Tyr Ile Phe Cys Ser Met Gly Leu Leu Ile Pro Phe
 1 5 10 15

Ser Thr Leu Lys Met Leu Leu Ile Val Phe Pro Leu Ser Leu Phe Pro
 20 25 30

Lys Arg Asn Leu Leu Ser Phe Leu Ser Leu
 35 40

<210> 187
 <211> 100
 <212> PRT
 <213> Homo sapiens

<400> 187
 Leu Phe Phe Phe Leu Arg Trp Ser Leu Ala Leu Val Thr Gln Ala Gly
 1 5 10 15

Val Gln Val Val Asp Ile Gly Ser Leu Gln Pro Leu Pro Pro Gly Phe
 20 25 30

Lys Gln Phe Ser Cys Pro Ser Leu Leu Ser Ser Trp Asp Tyr Arg His
 35 40 45

Gly Pro Pro Arg Pro Ala Asn Phe Phe Val Phe Leu Val Glu Met Gly
 50 55 60

Phe His His Val Gly Gln Ala Gly Pro Glu Leu Leu Thr Ser Ser Asp
 65 70 75 80

Pro Pro Ala Leu Ala Ser Gln Ser Ala Gly Ile Thr Gly Val Ser His
 85 90 95

Leu Thr Trp Pro
 100

<210> 188
 <211> 106
 <212> PRT
 <213> Homo sapiens

<400> 188
 Met Ser Cys Leu Trp Pro Ser Leu Asp Leu Pro Ser Leu Ser His Ser
 1 5 10 15

Lys Gln Ser Ser Ser Gln Ala Glu Gly Gln Val Thr Ser His Thr Arg
 20 25 30

WO 02/48370

PCT/US01/51341

73

Gln Arg Phe Pro Asp Gly Ala His Leu His Pro Ser Leu Thr Leu Val
 35 40 45
 Leu Ser Gln Asp Ala Pro Leu Arg Leu Ala Pro Pro Thr Leu Cys Leu
 50 55 60
 Leu Cys Tyr Trp Ala Ser Leu Pro Ser Pro Arg Thr Pro Glu Leu Leu
 65 70 75 80
 Asn Ala Gly Gln Lys Ser Ile Pro Asp Leu Gln Gln Arg His Phe Asp
 85 90 95
 Ile Lys Glu Met Ala Leu Asp Phe Cys Leu
 100 105

<210> 189
 <211> 46
 <212> PRT
 <213> Homo sapiens

<400> 189
 Met Val Ile Ser Arg Ile Ser Ile Leu Arg Lys Met Thr Lys Phe His
 1 5 10 15
 Lys Phe Cys Ser Gln Leu Thr Glu Pro Gly Arg Arg Thr Gln Pro Lys
 20 25 30
 Glu Asn Pro Trp Ser Leu Tyr Asp Thr Asp Trp Leu Glu Lys
 35 40 45

<210> 190
 <211> 46
 <212> PRT
 <213> Homo sapiens

<400> 190
 Met Ser Arg Val Arg Ala Glu Lys Pro Gly Arg Val Ala Lys Leu Ala
 1 5 10 15
 Ala Cys Arg Pro Leu Pro Arg Leu Gln Met Ser Gly Ser Ile Pro Leu
 20 25 30
 His Lys Cys Lys Glu Lys Ala Ser Met Pro Pro Leu Trp Ser
 35 40 45

<210> 191
 <211> 50
 <212> PRT
 <213> Homo sapiens

<400> 191
 Met Arg Pro Ala Arg Leu Gly Pro Arg Cys Ser Asp Leu Asp Phe Gly
 1 5 10 15
 Leu Val Leu Ser Ser Trp Leu Arg Leu Ala Arg Cys Pro Leu Glu Ser

WO 02/48370

PCT/US01/51341

20 74 30
 25 40 45
 Ser Phe Gly Phe Ala Phe Phe Val Cys Leu Phe Ser Pro Asn Phe Cys
 35 40 45
 Gln Thr
 50

<210> 192
 <211> 76
 <212> PRT
 <213> Homo sapiens

<400> 192
 Met Glu Gly Thr Val Gly Gln Ala Lys Met Val Glu Lys Trp Met Arg
 1 5 10 15
 Pro Thr Leu Leu Met Ser Leu Arg Gly Leu Gly Glu Arg Ser Asn Glu
 20 25 30
 Pro His Val Ser Pro Glu Ser Ser Ala Ala Pro Lys Ala Gly Pro Ser
 35 40 45
 Leu Glu Asp Cys Glu Arg Glu Asp Gly Ser Ile Arg Thr Gly Trp Asp
 50 55 60
 Thr Ala Pro Thr Lys Glu Ser Pro Thr Ser Cys Ala
 65 70 75

<210> 193
 <211> 54
 <212> PRT
 <213> Homo sapiens

<400> 193
 Arg Thr Val Cys Thr Lys Val Ser Cys Pro Val Gln Leu Pro Ala Asp
 1 5 10 15
 Trp Thr Cys Lys Val Gln Pro Val Trp Leu Glu Phe Pro Cys Leu Pro
 20 25 30
 Ile Ser Cys Arg Leu Arg Val Ser Ser Asp Thr Ser Pro Asp Ser Ala
 35 40 45
 Thr Trp Gly Ser Trp Lys
 50

<210> 194
 <211> 27
 <212> PRT
 <213> Homo sapiens

<400> 194
 Met Glu Pro Arg Ile Pro Val Lys Thr Phe Ser Phe Asp Lys Arg Ile
 1 5 10 15

WO 02/48370

PCT/US01/51341

75

Leu Ile Arg Ile Leu Tyr Gln Ile Glu Gln Lys
 20 25

<210> 195
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 195
 Met Leu Gln His Leu Arg Leu Thr Ile Trp Gly Glu Cys Val Trp Val
 1 5 10 15

Phe

<210> 196
 <211> 51
 <212> PRT
 <213> Homo sapiens

<400> 196
 Met Arg Asn Val Ser Leu Ile Ser Cys Glu Asp Ala Asp Phe Thr Glu
 1 5 10 15

Ala Leu Cys Asn Ile Trp Phe Val His Gln Thr Met Leu Ile Asp Cys
 20 25 30

Arg Ser His Leu Leu Pro Arg Trp Leu Thr Lys Thr Val Gly Ser Leu
 35 40 45

Leu Lys Pro
 50

<210> 197
 <211> 62
 <212> PRT
 <213> Homo sapiens

<400> 197
 Met Ser His Gly Gln Val Leu Gly Asp Val Ala Gly Lys Val Gly His
 1 5 10 15

Ala Leu Gly Thr Glu Asp Gln Thr Phe Ala Val Glu Val Leu Lys Glu
 20 25 30

Thr Thr Pro Phe Phe Arg Ala Ser Ser Gly Pro Thr Gly Asp Pro Trp
 35 40 45

Cys Pro Asp His Lys Ile Gln Ser Lys Pro Val Ser Leu Ser
 50 55 60

<210> 198
 <211> 400
 <212> PRT
 <213> Homo sapiens

WO 02/48370

PCT/US01/51341

76

<400> 196
 Met Leu Leu Leu Val Thr Ser Leu Leu Leu Cys Glu Leu Pro His Pro
 1 5 10 15
 Ala Phe Leu Leu Ile Pro Glu Lys Ser Asp Leu Arg Thr Val Ala Pro
 20 25 30
 Ala Ser Ser Leu Asn Val Arg Phe Asp Ser Arg Thr Met Asn Leu Ser
 35 40 45
 Trp Asp Cys Gln Glu Asn Thr Thr Phe Ser Lys Cys Phe Leu Thr Asp
 50 55 60
 Lys Lys Asn Arg Val Val Glu Pro Arg Leu Ser Asn Asn Glu Cys Ser
 65 70 75 80
 Cys Thr Phe Arg Glu Ile Cys Leu His Glu Gly Val Thr Phe Glu Val
 85 90 95
 His Val Asn Thr Ser Gln Arg Gly Phe Gln Gln Lys Leu Leu Tyr Pro
 100 105 110
 Asn Ser Gly Arg Glu Gly Thr Ala Ala Gln Asn Phe Ser Cys Phe Ile
 115 120 125
 Tyr Asn Ala Asp Leu Met Asn Cys Thr Trp Ala Arg Gly Pro Thr Ala
 130 135 140
 Pro Arg Asp Val Gln Tyr Phe Leu Tyr Ile Arg Asn Ser Lys Arg Arg
 145 150 155 160
 Arg Glu Ile Arg Cys Pro Tyr Tyr Ile Gln Asp Ser Gly Thr His Val
 165 170 175
 Gly Cys His Leu Asp Asn Leu Ser Gly Leu Thr Ser Arg Asn Tyr Phe
 180 185 190
 Leu Val Asn Gly Thr Ser Arg Glu Ile Gly Ile Gln Phe Phe Asp Ser
 195 200 205
 Leu Leu Asp Thr Lys Lys Ile Glu Arg Phe Asn Pro Pro Ser Asn Val
 210 215 220
 Thr Val Arg Cys Asn Thr Thr His Cys Leu Val Arg Trp Lys Gln Pro
 225 230 235 240
 Arg Thr Tyr Gln Lys Leu Ser Tyr Leu Asp Phe Gln Tyr Gln Leu Asp
 245 250 255
 Val His Arg Lys Asn Thr Gln Pro Gly Thr Glu Asn Leu Leu Ile Asn
 260 265 270
 Val Ser Gly Asp Leu Glu Asn Arg Tyr Asn Phe Pro Ser Ser Glu Pro
 275 280 285
 Arg Ala Lys His Ser Val Lys Ile Arg Ala Ala Asp Val Arg Ile Leu
 290 295 300

WO 02/48370

PCT/US01/51341

77

```

Asn Trp Ser Ser Trp Ser Glu Ala Ile Glu Phe Gly Ser Asp Asp Gly
305              310              315              320

Asn Leu Gly Ser Val Tyr Ile Tyr Val Leu Ile Val Gly Thr Leu
325              330              335

Val Cys Gly Ile Val Leu Gly Phe Leu Phe Lys Arg Phe Leu Arg Ile
340              345              350

Gln Arg Leu Phe Pro Pro Val Pro Gln Ile Lys Asp Lys Leu Asn Asp
355              360              365

Asn His Glu Val Glu Asp Glu Ile Ile Trp Glu Glu Phe Thr Pro Glu
370              375              380

Glu Gly Lys Gly Tyr Arg Glu Glu Val Leu Thr Val Lys Glu Ile Thr
385              390              395              400

```

<210> 199
 <211> 10
 <212> PRT
 <213> Homo sapiens

<400> 199
 Met Asp Arg Met Glu Lys Arg Gln Thr Asp
 1 5 10

<210> 200
 <211> 20
 <212> PRT
 <213> Homo sapiens

<400> 200
 Met Cys Tyr Ala Thr Leu His Gln Ile Asn Phe Leu Gln Thr Val Leu
 1 5 10 15

Val Pro Gly Leu
 20

<210> 201
 <211> 31
 <212> PRT
 <213> Homo sapiens

<400> 201
 Met Cys Leu Cys Cys Trp Leu Tyr Trp Glu Glu Tyr Gly Pro Leu Ser
 1 5 10 15

Leu Thr Gln Glu Phe His Val Phe Cys Gln Asp Thr Leu His Gly
 20 25 30

<210> 202

WO 02/48370

PCT/US01/51341

78

<211> 54
 <212> PRT
 <213> Homo sapiens

 <400> 202
 Met Asn His Ser Leu Ser Ala Phe Gln Arg Ala Leu Gln Val Leu Ile
 1 5 10 15
 Phe Lys Met Ser Val Tyr Ala Ser Gly Pro Arg Leu Glu Lys Lys Val
 20 25 30
 Gly Asn Lys Leu Glu Gly Gly Arg Lys Gln Glu Arg Asn Val Thr Tyr
 35 40 45
 Met Ala Asp Glu Gly Phe
 50

 <210> 203
 <211> 35
 <212> PRT
 <213> Homo sapiens

 <400> 203
 Met Ile Lys Ala Tyr His Pro Tyr Phe Glu Asn Phe Asn His Arg Ala
 1 5 10 15
 Gln Tyr Val Ser Asn Lys Leu Lys Lys Tyr Ser Phe Gln Leu His Phe
 20 25 30
 Asp Gly His
 35

 <210> 204
 <211> 76
 <212> PRT
 <213> Homo sapiens

 <400> 204
 Met Lys Met Val Asn Arg His Met Lys Trp Lys Ser Ser Ala Leu Ser
 1 5 10 15
 Asp Leu Val Cys Ile Ser Thr Glu Ile Gln Ala Gly Leu Thr Arg His
 20 25 30
 Thr Ser His Asn Phe Gln Cys His Cys Thr Ile Ile Leu Thr Val Val
 35 40 45
 Ser Phe Phe Gln Ser Thr Glu Lys Gln Ala Asp Lys Pro Arg His Leu
 50 55 60
 Asn Val Thr Trp Leu Met Thr Leu Ile Ser Thr Leu
 65 70 75

<210> 205
 <211> 94
 <212> PRT

WO 02/48370

PCT/US01/51341

79

<213> Homo sapiens

<400> 205

```

Met Glu Gly Gln Asp Ser Leu Arg Asp Val Gly Ala Leu Ser His Leu
 1           5           10          15

Ala His Thr Asp Arg Ser Trp Leu Gly Arg Ala Gly Val Ser Ala Trp
 20          25          30

Arg Pro Ser Ala Ala Gly Asp Pro Gly Phe His Glu Val Gly Gly Val
 35          40          45

His Ala Gly Thr Ser Gln Leu Ala Gly Pro Gly Gly His Pro Gly Gly
 50          55          60

Ala Gly Ala Trp Gly His Glu Phe Thr Lys Val Ala Gln Gly Gln Glu
 65          70          75          80

Glu Thr Val Val Ala Glu Gly Pro Leu Val Glu Ala Trp Ala
 85          90

```

<210> 206

<211> 53

<212> PRT

<213> Homo sapiens

<400> 206

```

Met Pro Gln Asp Gln Asp Pro Pro Arg Glu Glu His Ala Ala Leu Arg
 1           5           10          15

Val Leu Phe Pro Arg Val Pro Leu Ala Val Pro His Gln Leu Gly Gly
 20          25          30

Glu Leu Glu Arg Ala Asp Arg Arg Thr Gly Phe Ser Ala Cys Ala Asn
 35          40          45

Ile Leu Thr Cys Pro
 50

```

<210> 207

<211> 75

<212> PRT

<213> Homo sapiens

<400> 207

```

Trp Ser Thr Pro Pro Phe Asp Pro Arg Phe Pro Ser Gln Asn Gln Ile
 1           5           10          15

Arg Asn Cys Tyr Gln Asn Phe Leu Asp Tyr His Arg Cys Leu Lys Thr
 20          25          30

Arg Thr Arg Arg Gly Lys Ser Thr Gln Pro Cys Glu Tyr Tyr Ser Cys
 35          40          45

Val Tyr His Ser Leu Cys Pro Ile Ser Trp Val Glu Ser Trp Asn Glu
 50          55          60

```


WO 02/48370

PCT/US01/51341

80

Glu Ile Lys Asn Gly Ile Phe Ala Gly Lys Ile
 65 70 75

<210> 208
 <211> 44
 <212> PRT
 <213> Homo sapiens

<400> 208
 Met Arg Val Leu Arg Lys Glu Ser Pro Ser Arg His Val Leu Lys Asn
 1 5 10 15

Met Cys Leu Ile Arg Asn Pro Arg Glu Gly Thr Ala Ala Asn Asn Glu
 20 25 30

Met Glu Ser Ala Thr Gly Glu Glu Lys Gly Asn Arg
 35 40

<210> 209
 <211> 83
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> {80}
 <223> a, c, g or t

<400> 209
 Met His Arg Lys Lys Lys Leu Glu Ser Phe Leu Leu Leu Ile Pro Pro
 1 5 10 15

Pro Tyr Leu Pro Leu Thr Lys Met Trp Gly Glu Pro Arg Phe Glu Gly
 20 25 30

Ser Thr Gly Pro Cys Pro Gln Asp Ser Met Glu Gln Pro Val Thr Lys
 35 40 45

Gln Gly Ile Ser Leu Lys Ser Cys Leu Pro Lys Lys Ile Lys Leu Pro
 50 55 60

Arg Leu Ala Leu His Pro Ser Pro Pro Arg Ser Phe Pro Leu Lys Xaa
 65 70 75 80

Lys Lys Leu

<210> 210
 <211> 40
 <212> PRT
 <213> Homo sapiens

<400> 210
 Met Thr Arg Phe Ser Gln Ala Ser Ser Ser Lys Asp Lys Thr Pro Pro
 1 5 10 15

WO 02/48370

PCT/US01/51341

81

Leu Pro Ser Met Val Gln Ala Thr Val Leu Val Lys Lys Tyr Ile Phe
 20 25 30

Thr Lys Lys Lys Ser Tyr Val Leu
 35 40

<210> 211
 <211> 87
 <212> PRT
 <213> Homo sapiens

<400> 211
 Met Pro Arg Pro Thr Glu Gly Glu Gly Ser Thr Glu Asp Arg Asp Pro
 1 5 10 15

Ile Gly Ile Gln Ser Gln Thr Arg Ala Glu Pro Thr Val Glu Gln Leu
 20 25 30

Met Ser Gly Ala Lys Asp Thr Ser Trp Asn Pro Pro Asp Gly Ser Ser
 35 40 45

Asn Pro Lys Arg Ala Gly Leu Gln Val Gly Leu Asn Trp Arg Asp Pro
 50 55 60

Gln Glu Ser Gly Arg Arg Asn Thr Arg Ala Phe Leu Glu Glu Gly Thr
 65 70 75 80

Phe Ile Leu Asp Ser Asn Gln
 85

<210> 212
 <211> 38
 <212> PRT
 <213> Homo sapiens

<400> 212
 Met Met Pro Gly Pro Ala Ala Leu Ile Pro Pro Thr Ala Thr Ala Cys
 1 5 10 15

Leu Leu Val Val Ala Arg Gly Ser Ser Val Pro Lys Asp Ser Ser Leu
 20 25 30

Phe Cys Ile Thr Val His
 35

<210> 213
 <211> 88
 <212> PRT
 <213> Homo sapiens

<400> 213
 Met Ser Leu Leu Asp Ala Ser Ser Leu Lys Pro Tyr Asp Ser Phe Leu
 1 5 10 15

Leu Ala Val Leu Phe Leu Thr Arg Asp Asn Lys Gly Phe Ala Ser Gln
 20 25 30

WO 02/48370

PCT/US01/51341

82

Val Cys Met Ala Lys Lys Val Ser Thr Ser Val Asn Gly Ser Phe Leu
 35 40 45

Met Thr Ser Gln Gln Pro Leu Val Lys Asp Val Ile Glu Ile Val Gln
 50 55 60

Arg Leu Gly Ser Val Cys Phe Val Leu Leu Leu Lys Ser Phe His Gly
 65 70 75 80

Ser Lys Leu Phe Leu Ser Ile Val
 85

<210> 214

<211> 42

<212> PRT

<213> Homo sapiens

<400> 214

Met Val Ile Arg Glu Leu Leu Gly Gly Gln Lys Tyr Pro Asn Pro Val
 1 5 10 15

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

Glu Gly Asp Ser Glu Ala Gln Thr Arg Ile
 35 40

<210> 215

<211> 49

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> (46)

<400> 215

Met Pro Asn Cys Ser Val Glu Leu Arg Gly Tyr Tyr Tyr Asn Phe Val
 1 5 10 15

His Tyr Tyr Lys Tyr Phe Ile Leu Val Val Tyr Ser Thr Ala Asp Ser
 20 25 30

Asn Gln Lys Thr Lys Ile Gln Lys Tyr Tyr Ile Leu Glu Xaa Ile Ile
 35 40 45

Met

<210> 216

<211> 37

<212> PRT

<213> Homo sapiens

<220>

WO 02/48370

PCT/US01/51341

83

<221> UNSURE
 <222> (6)

<220>
 <221> UNSURE
 <222> (8)

<400> 216
 Met Glu Met Leu Glu Xaa Lys Xaa Thr Ile Ile Asp Ile Val Ser Leu
 1 5 10 15
 Leu Ala Leu Ser Gly Asp Leu Thr Gln Leu Arg Lys Ser Leu Val Thr
 20 25 30
 Leu Lys Ile Cys Arg
 35

<210> 217
 <211> 72
 <212> PRT
 <213> Homo sapiens

<400> 217
 Met Gly Ser Tyr Gly Leu Leu Phe Lys Phe Tyr Gly Ala Ile Phe Thr
 1 5 10 15
 Ser Val Ala Gln Gly Trp Ser Val Leu His Leu Arg Lys Val Ser Leu
 20 25 30
 Gly Lys Cys Pro Cys His Pro Ser His Ser Arg Gln Ala Ala Ser Ser
 35 40 45
 Ala Phe Ser Ser Ser Ser Ser His Ala Trp Ser Ser Pro Phe Val Ile
 50 55 60
 Phe Ser Ser Leu Thr Pro Ser Leu
 65 70

<210> 218
 <211> 49
 <212> PRT
 <213> Homo sapiens

<400> 218
 Met Gly Ser Phe Ser Pro Leu Thr Tyr His Leu Gly His Trp Asn Met
 1 5 10 15
 Ala Ala Cys Gly Ser Val Cys Glu Gly Pro Gly Asp Gly Gln Gly Gly
 20 25 30
 Ser Ala Leu Phe Cys Phe Tyr Gln His Cys Ser Met Asn Val Phe Leu
 35 40 45

Thr

WO 02/48370

PCT/US01/51341

84

<210> 219
 <211> 34
 <212> PRT
 <213> Homo sapiens

<400> 219
 Met Leu Thr Arg His His Pro Leu Asn Val Leu Leu His Arg Leu Cys
 1 5 10 15

Leu Asn Trp Leu Glu Glu Asn Asn Tyr Pro Arg Asn Thr Asp Tyr Leu
 20 25 30

Ile Phe

<210> 220
 <211> 34
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (17)

<400> 220
 Met Leu Leu Leu Pro Ala Thr Phe Leu Pro Thr Ser His Ala Arg Pro
 1 5 10 15

Asa Gln Pro His Cys His Thr Thr Cys Leu Ile Thr Ser His Val Leu
 20 25 30

Thr His

<210> 221
 <211> 111
 <212> PRT
 <213> Homo sapiens

<400> 221
 Met Gly Pro Ser Ser Cys Leu Leu Leu Ile Leu Ile Pro Leu Leu Gln
 1 5 10 15

Leu Ile Asn Leu Gly Ser Thr Gln Cys Ser Leu Asp Ser Val Met Asp
 20 25 30

Lys Lys Ile Lys Asp Val Leu Asn Ser Leu Glu Tyr Ser Pro Ser Pro
 35 40 45

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

Ser Ser Cys Pro Ala Gly Met Ala Val Thr Gly Cys Ala Cys Gly Tyr
 65 70 75 80

Gly Cys Gly Ser Trp Asp Val Gln Leu Glu Thr Thr Cys His Cys Gln
 85 90 95

WO 02/48370

PCT/US01/51341

85

Cys Ser Val Val Asp Trp Thr Thr Ala Arg Cys Cys His Leu Thr
 100 105 110

<210> 222
 <211> 111
 <212> PRT
 <213> Homo sapiens

<400> 222
 Met Gly Pro Ser Ser Cys Leu Leu Leu Ile Leu Ile Pro Leu Leu Gln
 1 5 10 15
 Leu Ile Asn Leu Gly Ser Thr Gln Cys Ser Leu Asp Ser Val Met Asp
 20 25 30
 Lys Lys Ile Lys Asp Val Leu Asn Ser Leu Glu Tyr Ser Pro Ser Pro
 35 40 45
 Ile Ser Lys Lys Leu Ser Cys Ala Ser Val Lys Ser Gln Gly Arg Pro
 50 55 60
 Ser Ser Cys Pro Ala Gly Met Ala Val Thr Gly Cys Ala Cys Gly Tyr
 65 70 75 80
 Gly Cys Gly Ser Trp Asp Val Gln Leu Glu Thr Thr Cys His Cys Gln
 85 90 95
 Cys Ser Val Val Asp Trp Thr Thr Ala Arg Cys Cys His Leu Thr
 100 105 110

<210> 223
 <211> 83
 <212> PRT
 <213> Homo sapiens

<400> 223
 Met Asn Val Glu Ala Arg Glu Gln Cys Asp Val Gln Leu Ser Asp Leu
 1 5 10 15
 Thr Trp His Leu Ile Trp Leu Glu Val Pro Pro Leu Leu Ser Val Pro
 20 25 30
 Trp Leu Trp Ala His Gly Leu Ala Glu Pro Ser Tyr Gly Phe Arg Phe
 35 40 45
 Thr Cys Tyr Asn Ile Gln Arg Gln Cys Thr Ser Leu Pro Arg Lys Leu
 50 55 60
 Cys Ser Arg His Pro Phe Val Thr Leu Ile Ser Ile Met Asp Thr Thr
 65 70 75 80
 Thr Phe Tyr

<210> 224

WO 02/48370

PCT/US01/51341

86

<211> 132
 <212> PRT
 <213> Homo sapiens

 <320>
 <221> UNSURE
 <222> (3)

 <220>
 <221> UNSURE
 <222> (11)

 <220>
 <221> UNSURE
 <222> (14)

 <400> 224
 Met Asp Xaa Thr Arg Val His Asp Asp Glu Xaa Val Ile Xaa Gly Asp
 1 5 10 15
 Val Phe Val His Glu Val Thr Pro Gly Pro His Arg Trp Val Leu Val
 20 25 30
 Arg Pro Phe Cys Leu Glu Val Arg Ala Val Phe Leu Arg Leu Trp Tyr
 35 40 45
 Tyr Arg Gly Glu Lys Glu Glu Glu Leu Glu Val Arg Glu Arg Ser Cys
 50 55 60
 Arg Leu Gly Arg Cys Asp Gln Gly Gln Arg Asp Gly Val Gln Glu Ala
 65 70 75 80
 Cys Ser Ser Val Ser Cys Ser Leu Arg Gln Glu Val Ser Pro Ser Ser
 85 90 95
 Gln Leu Asp Met Arg Ser Leu Leu Gly Val Pro Leu Ala Glu Val Glu
 100 105 110
 Pro Val Ala Gln His Pro Pro Asn Glu Gly Arg Gly Arg His Leu Gly
 115 120 125
 Gln Cys Leu Leu
 130

 <210> 225
 <211> 38
 <212> PRT
 <213> Homo sapiens

 <400> 325
 Met Ile Asn Asn Ser Asn His Asn Asn Ser Ser Ser Ser Lys Leu Arg
 1 5 10 15
 Ala Ser Tyr Val Gln Ala Phe Ser Lys His Phe Thr Cys Ile Thr Pro
 20 25 30
 Leu Val Ile Thr Thr Pro
 35

WO 02/48370

PCT/US01/51341

87

<210> 226
 <211> 58
 <212> PRT
 <213> Homo sapiens

<400> 226
 Met Ser Thr Phe Thr Val Leu Lys Asn Thr His Gln Leu Lys Lys Asn
 1 5 10 15
 Thr Leu Phe Pro Phe Leu Gly His Leu Asn Leu Arg Glu Gln Leu Leu
 20 25 30
 Tyr Lys Asn Asp Ile Lys Ile Ile His Phe Gly Ser Met Phe Leu Thr
 35 40 45
 Val Leu Arg Gly Cys Met Val Lys Leu Lys
 50 55

<210> 227
 <211> 26
 <212> PRT
 <213> Homo sapiens

<400> 227
 Met His Met His Ile Phe Leu Cys Leu Tyr Asn Leu Cys Asn Ile Cys
 1 5 10 15
 Glu Cys Asn Thr Phe Ser Phe Phe Leu Leu
 20 25

<210> 228
 <211> 47
 <212> PRT
 <213> Homo sapiens

<400> 228
 Met Leu Asp Val Met Arg Gln Val Ala Arg Ser Trp Leu Thr Ala Met
 1 5 10 15
 Glu Arg Leu Leu Leu Pro Ala Ala Val Arg Phe Ser Ala Ile Trp Leu
 20 25 30
 Ala Gly Gln Phe Ala Met Ala Trp Leu Leu Gln Leu Ile Leu Gly
 35 40 45

<210> 229
 <211> 53
 <212> PRT
 <213> Homo sapiens

<400> 229
 Met Gly Asn Ile Gly Glu Thr Leu Ser Leu Lys Lys Lys Arg Arg Ala
 1 5 10 15

WO 02/48370

PCT/US01/51341

88

Gly Gly Glu Ser Val Lys Asp Pro Gly Ser Thr Asp Thr Gly Gly Gln
 20 25 30

Arg Thr Arg Val Gly Val Ser Ser Asn Asp Ser Val Gly Ser Met Gly
 35 40 45

Ala Val Gly Arg Glu
 50

<210> 230
 <211> 80
 <212> PRT
 <213> Homo sapiens

<400> 230
 Met Val Ile Asn Ser Cys Ile Ile Pro Leu Pro Ser Gln Ala Thr Ile
 1 5 10 15

Pro Glu Pro Trp Pro His Gly Ala Cys Ile Phe Arg Ile Gln Thr Pro
 20 25 30

Trp Gly Ser Ser Pro Leu Leu Pro Ser Leu Ser Ser His Pro Leu Thr
 35 40 45

His Leu Ser Cys Tyr Leu Ser Leu Glu Ile Pro Lys Met Met Cys Val
 50 55 60

Met Glu Arg Leu Glu His Gln Leu Gln Asn His Pro Val Thr Leu Ala
 65 70 75 80

<210> 231
 <211> 40
 <212> PRT
 <213> Homo sapiens

<400> 231
 Met Phe Gln Arg Phe Leu Ala Lys Val Thr Val Trp Met Val Val Pro
 1 5 10 15

Leu Thr Lys Thr Ala Met Asn Ala Lys Arg Ala Ser Phe Val Gly Arg
 20 25 30

His Lys Ile Ile Phe Arg Ile Cys
 35 40

<210> 232
 <211> 24
 <212> PRT
 <213> Homo sapiens

<400> 232
 Met Leu Leu Tyr Leu Ile Thr Arg Gly Asp Val Glu Asn Gly Cys Phe
 1 5 10 15

WO 02/48370

PCT/US01/51341

89

Ile Phe Ser Val Val Phe Ala Leu
20

<210> 233
<211> 26
<212> PRT
<213> Homo sapiens

<400> 233
Met Pro Pro Arg Gly Leu Pro His Phe Ser Pro His Pro Thr Arg Gln
1 5 10 15
Phe Leu Phe Leu Phe Pro Leu His Thr Lys
20 25

<210> 234
<211> 37
<212> PRT
<213> Homo sapiens

<400> 234
Met Ser Tyr Glu Ile Leu Val Asn Thr Asp Phe Met Ser Pro Phe Leu
1 5 10 15
Arg Thr Leu Leu Val Cys Phe His Leu Tyr Ala Leu Ile Arg Ala Asn
20 25 30
Asn Leu Lys Tyr Pro
35

<210> 235
<211> 40
<212> PRT
<213> Homo sapiens

<400> 235
Met Gly Lys Gly Leu Arg Leu Gly Val Ser Ile Ile Leu Val Lys Ser
1 5 10 15
Phe Phe Thr Tyr Ser Ser Lys Asp Val Asn Tyr Phe Ser Ile His Ser
20 25 30
Asn Ile Lys Ala Val Phe His Phe
35 40

<210> 236
<211> 40
<212> PRT
<213> Homo sapiens

<400> 236
Met Glu Glu Thr Gly Pro Leu Pro Ser Gly Ser Ser Leu Ser Asp Gln
1 5 10 15

WO 02/48370

PCT/US01/51341

90

Gly Glu Thr Ala Leu Ala Leu Gly Asn Ser Arg Ser Asp Gly Gly Arg
 20 25 30

Gln Ser Ser Ser Ser Met Asn Ala
 35 40

<210> 237
 <211> 50
 <212> PRT
 <213> Homo sapiens

<400> 237
 Met His Lys Gln Ser Met Ala Arg Ser Ile Leu Arg Ser Pro Leu Gln
 1 5 10 15

Gln Ile Pro Pro Lys Gly Glu Ala Gly Arg Trp Arg Trp Ala Glu Ala
 20 25 30

Ser Cys Val Leu His Thr Phe Ser Thr Ile Leu Asp Phe Leu Phe Phe
 35 40 45

Phe Phe
 50

<210> 238
 <211> 49
 <212> PRT
 <213> Homo sapiens

<400> 238
 Ser Ser Trp Gly Asp Ser Phe Ala Val Ser Ala Ala Trp Ala Arg Lys
 1 5 10 15

Gly Ile Glu Glu Trp Ile Gly Arg Gln Arg Cys Pro Gly Gly Val Ser
 20 25 30

Gly Pro Arg Gln Leu Arg Leu Ala Gly Thr Ile Gly Arg Ser Thr Arg
 35 40 45

Glu

<210> 239
 <211> 54
 <212> PRT
 <213> Homo sapiens

<400> 239
 Met Leu Arg Pro Leu Thr Val Ala Ser Lys Arg Leu Leu Thr Ile Ser
 1 5 10 15

Thr Leu Lys Ser Pro Leu Val Gly Leu Cys Ser Phe Ser Lys Ser Gly
 20 25 30

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

WO 02/48370

PCT/US01/51341

91

Asp Trp Gln Lys Gln His
50

<210> 240
<211> 23
<212> PRT
<213> Homo sapiens

<400> 240
Met Lys Lys Lys Ser Tyr Pro Asp Lys Ile Asn Gln Cys Phe Ile Phe
1 5 10 15

Leu Glu His Gln Asn Leu Leu
20

<210> 241
<211> 59
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (6)..(7)

<220>
<221> UNSURE
<222> (9)

<220>
<221> UNSURE
<222> (13)

<220>
<221> UNSURE
<222> (23)

<220>
<221> UNSURE
<222> (27)..(31)

<220>
<221> UNSURE
<222> (38)..(40)

<220>
<221> UNSURE
<222> (43)

<220>
<221> UNSURE
<222> (45)

<220>
<221> UNSURE
<222> (47)

WO 02/48370

PCT/US01/51341

92

<400> 241
 Met Val Lys Tyr Met Xaa Xaa Leu Xaa Leu Thr Pro Xaa Phe Ser Asn
 1 5 10 15
 Leu Leu Gly Thr Leu Lys Xaa Arg Lys Val Xaa Xaa Xaa Xaa Xaa Pro
 20 25 30
 Arg Lys Arg Asn Phe Xaa Xaa Xaa Pro Pro Xaa Leu Xaa Lys Xaa Arg
 35 40 45
 Cys His Phe Leu His Ile Asp Leu Gln Arg Val
 50 55

<210> 242
 <211> 55
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (53)

<400> 242
 Met Val Ser Gly Val Gln Val Ser Leu His Lys Thr Lys Ile Lys Leu
 1 5 10 15
 Phe Asn Thr Gly Pro Thr Thr Leu Ile Tyr Gly Ala Asn Thr Cys Cys
 20 25 30
 Glu Pro Trp Gly Gln Gly Leu Gly Asp Lys Val Ala Thr Ile Phe Trp
 35 40 45
 Gly Val Gly Gly Xaa Gly Gly
 50 55

<210> 243
 <211> 75
 <212> PRT
 <213> Homo sapiens

<400> 243
 Met Val Ile Thr Cys Val Leu Tyr Asp Ile Ser Ser Leu Lys Asn Leu
 1 5 10 15
 Arg His Ser Pro Phe Leu Gln Val Phe Phe Cys Val Cys Trp Lys Ile
 20 25 30
 Met Tyr Ile Phe Gln Leu Leu Asn Ala Ser Val Cys Ile Cys Ile Ser
 35 40 45
 Thr Lys Ser Lys Leu Leu Ile Leu Leu Phe Lys Leu Phe Ala Ser Tyr
 50 55 60
 Trp Phe Ser Leu Pro Thr Leu Cys Ile Asn Ser
 65 70 75

WO 02/48370

PCT/US01/51341

93

<210> 244
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 244
 Met Ser Trp Val Pro Cys Gly Cys Asp Phe Leu Arg Glu Ile Asn Leu
 1 5 10 15

Phe

<210> 245
 <211> 30
 <212> PRT
 <213> Homo sapiens

<400> 245
 Met Tyr Val Ser Pro Asp Asn Ile Ser Gly Ser Gly Asn Cys Lys Lys
 1 5 10 15

Lys Ile Gly Asn Gln Asn Ser Arg Lys Val Phe Leu Glu Gly
 20 25 30

<210> 246
 <211> 57
 <212> PRT
 <213> Homo sapiens

<400> 246
 Arg Val Thr Met Asn Glu Lys Asp Asn Phe Met Asn Ala Glu Asn Leu
 1 5 10 15

Gly Ile Val Phe Gly Pro Thr Leu Met Arg Pro Pro Glu Asp Ser Thr
 20 25 30

Leu Thr Thr Leu His Asp Met Arg Tyr Gln Lys Leu Ile Val Gln Ile
 35 40 45

Leu Ile Glu Asn Glu Asp Val Leu Phe
 50 55

<210> 247
 <211> 70
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (38)

<400> 247
 Met Phe Ala Ser Leu Leu Ile Thr Asn Leu Leu Ser Thr Asn Glu Lys
 1 5 10 15

Tyr Ile Gln Asp Leu Pro Phe Gln Arg Leu Ser Ile Tyr Glu Thr Asn

WO 02/48370

PCT/US01/51341

94

20 25 30

Ser Pro Phe Arg Leu Xaa Asn Phe Glu Asp Val Phe Ile Phe Leu Phe
 35 40 45

Phe Leu Asn Lys Asn Cys Phe Leu Ser Arg Leu Phe Lys Ala Thr Cys
 50 55 60

Val Lys Pro Leu Val Gln
 65 70

<210> 248
 <211> 36
 <212> PRT
 <213> Homo sapiens

<400> 248
 Met Arg Arg Ala Arg Pro Pro Leu Phe Phe Leu His Ala Val Ser Ser
 1 5 10 15

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

Gly Pro Val Met
 35

<210> 249
 <211> 26
 <212> PRT
 <213> Homo sapiens

<400> 249
 Met Ser Leu Ser Phe Ser Leu His Ser Phe Tyr Arg Lys Ala Ile Leu
 1 5 10 15

Gly Val Leu Gly His Phe Asp Ser Thr Ser
 20 25

<210> 250
 <211> 43
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (6)

<400> 250
 Met Ser Leu Pro Ser Xaa Arg Arg Gln Phe Ser Asp Ile Thr Cys Thr
 1 5 10 15

Glu Ile His Tyr Asn Ala Thr Met Asn Gly Gln Ser Ser Thr Glu Lys
 20 25 30

Ile Lys Gln Arg Met Ser Trp Lys Val Leu Trp
 35 40

【国際公開パンフレット（コレクトバージョン）】

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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
20 June 2002 (20.06.2002)

PCT

(10) International Publication Number
WO 02/048370 A3(51) International Patent Classification: C12N 15/12;
C12Q 1/68, C12N 1/21, 5/10, C07K 14/47, 16/18, G01N
33/68, A61K 39/395, 38/17, 31/713141, Chuganua (CN/US); 1155 Randolph Way #34, San
Jose, CA 95117 (US)

(21) International Application Number: PCT/US01/51341

(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrrell
P.C., 65 E. Main Street, Marlton, NJ 08054 (US)

(22) International Filing Date: 30 October 2001 (30.10.2001)

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

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 60/244,717 31 October 2000 (31.10.2000) US

(84) Designated States (optional): ARIPO patent (GI, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BI, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application: 60/244,717 (CIP)
US
Filed on 31 October 2000 (31.10.2000)(71) Applicant (for all designated States except US): DI-
ADEXUS, INC., 1155 Randolph Way, Suite 34, San Jose, CA 95117 (US);
South San Francisco, CA 94080 (US).Published:
with international search report

(72) Inventors; and

(75) Inventors/Applicants (for US only): SUN, Yangming
(CN/US); 551 Shou Drive, Redwood City, CA 94065
(US); RECIPON, Herre (FR/FR); 85 Vertum Avenue,
San Francisco, CA 94118 (US); GUOSHI, Malavika, G.
(IN/US); 1735 Via Laguna, San Jose, CA 95120 (US).(88) Date of publication of the international search report:
12 September 2002For two-letter codes and other abbreviations, refer to the "Guide-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/048370 A3

(54) Title: COMPOSITIONS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic colon cells, including fragments, variants, and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating colon, tumor and non-cancerous disease states in colon tissue; identifying colon tissue; monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered colon tissue for treatment and research.

【国際調査報告】

INTERNATIONAL SEARCH REPORT					International Application No. PCT, ... 01/51341
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/12 C12Q1/68 C12N1/21 C12N5/10 C07K14/47 C07K16/18 G01N33/68 A61K39/395 A61K38/17 A61K31/713					
According to International Patent Classification (IPC) or to both national classification and IPC					
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) Documentation searched other than minimum documentation in the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, SEQUENCE SEARCH, WPI Data, BIOSIS, EMBASE, CAB Data					
C. DOCUMENTS CONSIDERED TO BE RELEVANT					
Category *	Citation of document, with indication, where appropriate, of the relevant passages				Relevant to claim No.
X	DATABASE EMBL [Online] 10 July 2000 (2000-07-19) WATERSTON R.H.: "The sequence of Homo sapiens clone" Database accession no. AC019146 XP002221235 nts:113028-113445				1,3-5,7,8
X	DATABASE EMBL [Online] 11 June 2000 (2000-06-11) DIAS NETO E., ET AL.: "CN1-BN0117-150400-188-F03 BN0117 Homo sapiens cDNA" Database accession no. BE005537 XP002221240 nts:49-172				1-5,7,8
☒ Further documents are listed in the continuation of box C. ☒ The cited documents are listed in annex.					
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "T" document which may throw doubts on priority claim(s) or which is cited to establish the prior art on date of another citation or other special reason (see specification) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "G" document number of the same parent family					
Date of the actual completion of the international search			Date of mailing of the international search report		
19 November 2002			17. 03. 2003		
Name and mailing address of the ISA European Patent Office, P.O. Box 5010 NL 2280 HV Rijswijk Tel. (+31 70) 511-2040, Tx. 51 551 650 nl Fax (+31 70) 540 5316			Authorized officer Blanco Urgeiti, B		

Form PCT/ISA/210 (revised 1 June 2001)

INTERNATIONAL SEARCH REPORT

INTERNATIONAL SEARCH REPORT		Internet PCT/	Application No. 01/51341
C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	
L	DATABASE EMBL [Online] 28 June 2002 (2002-06-28) ESCOBEDO J. ET AL: "Human cancer related polynucleotide SEQ ID NO 2944" Database accession no. ABN62977 XP062221241 L document cited to provide information on the relevant sequence disclosed in WO0214500-A2		
E	the whole document - & WO 02 14500 A (SCOTT ELIZABETH M ; SUDDUTH KLINGER JULIE (US); CHIRON CORP (US); H) 21 February 2002 (2002-02-21) the whole document	1-12	
A	--- WO 96 39419 A (HUMAN GENOME SCIENCES INC ; ROSEN CRAIG A (US); YU GUO LIANG (US)) 12 December 1996 (1996-12-12)		
A	--- WO 00 50508 A (INCYTE PHARMA INC ; KLINGLER TOD M (US); LAL PREETI (US); VOLKMUTH) 31 August 2000 (2000-08-31) -----		

Form PCT/ISAR/16 (continued on second sheet) (July 1999)

INTERNATIONAL SEARCH REPORT	Inventor application No. CT/US 01/51341
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)	
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:	
1. ☒ Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: see FURTHER INFORMATION sheet PCT/ISA/210	
2. ☐ Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:	
3. ☐ Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).	
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)	
This International Searching Authority found multiple inventions in this International application, as follows:	
see additional sheet	
1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.	
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.	
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:	
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-17 partially	
Remark on Protest	
☐ The additional search fees were accompanied by the applicant's protest.	
☐ No protest accompanied the payment of additional search fees.	

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01/51341

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claim 14 is directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claim 16 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01/51341

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 218

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: (1-17)-partially

The nucleic acid molecule of SEQ ID NO: 1, derived polypeptide of SEQ ID NO:148, vectors, hosts cells, antibodies, kits, methods to detect the polypeptides, methods to detect the nucleic acids, vaccines, therapeutic and diagnosis applications thereof

Invention 2: (1-17)-partially

Idem as invention 1 but limited to SEQ ID NO:2

Inventions 3-147: (1-17)-partially

Idem as invention 1 but limited to SEQ ID NOs: 3 to 147 (Invention 3 is limited to SEQ ID NO:3; Invention 4 is limited to SEQ ID NO:4;...Invention 147 is limited to SEQ ID NO:147)

INTERNATIONAL SEARCH REPORT

1 Information on patent family members

International Application No.

PCT/JP 01/51341

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 0214500	A	21-02-2002	AU 8504701 A WO 0214500 A2	25-02-2002 21-02-2002
WO 9639419	A	12-12-1996	CA 2221798 A1 US 5733748 A WO 9639419 A1 AU 2820595 A EP 0847398 A1 JP 11506342 T US 6337195 B1 US 2002086314 A1	12-12-1996 31-03-1998 12-12-1996 24-12-1996 17-06-1998 08-06-1999 08-01-2002 04-07-2002
WO 0050588	A	31-08-2000	AU 3479700 A CA 2361743 A1 EP 1155126 A2 JP 2002536995 A WO 0050588 A2	14-09-2000 31-08-2000 21-11-2001 05-11-2002 31-08-2000

フロントページの続き

(51)Int.Cl. ⁷	F I	テーマコード (参考)
C 0 7 K 14/47	A 6 1 P 35/00	4 C 0 8 5
C 0 7 K 16/18	C 0 7 K 14/47	4 H 0 4 5
C 1 2 N 1/15	C 0 7 K 16/18	
C 1 2 N 1/19	C 1 2 N 1/15	
C 1 2 N 1/21	C 1 2 N 1/19	
C 1 2 N 5/10	C 1 2 N 1/21	
C 1 2 P 21/02	C 1 2 P 21/02	C
C 1 2 Q 1/68	C 1 2 Q 1/68	
G 0 1 N 33/53	G 0 1 N 33/53	D
G 0 1 N 33/566	G 0 1 N 33/53	M
	G 0 1 N 33/566	
	C 1 2 N 5/00	A

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 ゴーシュ, マラヴィカ ジー.

アメリカ合衆国 カリフォルニア州 9 5 1 2 0、サン ホセ、ヴィア ルガノ 1 7 3 3

(72)発明者 リュー, チェングァ

アメリカ合衆国 カリフォルニア州 9 5 1 1 7、サン ホセ、ランチェロ ウェイ # 1 4 1
1 2 5

F ターム(参考) 4B024 AA01 AA12 BA31 BA44 BA80 CA03 CA04 DA06 EA04 GA11
HA12 HA15
4B063 QA18 QA19 QQ43 QR55 QR62 QS25 QS34
4B064 AG01 CA02 CA19 CC24 DA01 DA14
4B065 AA26X AA93Y AB01 AC14 BA02 CA24 CA45
4C084 AA13 NA10 ZB262
4C085 AA03 AA13 AA14 BB01 BB33 BB34 BB35 BB36 BB37 CC02
CC23 DD62 DD63 DD88 EE01 GG02 GG03 GG04 GG06 GG08
4H045 AA10 AA11 AA20 AA30 BA41 CA40 DA76 DA86 EA28 EA31
EA51 FA72 FA74

专利名称(译)	<无法获取翻译>		
公开(公告)号	JP2004528823A5	公开(公告)日	2005-12-22
申请号	JP2002550085	申请日	2001-10-30
申请(专利权)人(译)	Diadekusasu公司		
[标]发明人	スンヨンミン レシボンハーヴ ゴーシュマラヴィカジー リユーチェングア		
发明人	スン,ヨンミン レシボン,ハーヴ ゴーシュ,マラヴィカ ジー. リユー,チェングア		
IPC分类号	G01N33/53 A61K38/00 A61K39/00 A61K39/395 A61K48/00 A61P35/00 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/02 C12Q1/68 G01N33/566		
CPC分类号	C07K14/47 A61K38/00 A61K2039/505 A61K2039/53		
FI分类号	C12N15/00.ZNA.A A61K39/00.H A61K39/395.D A61K39/395.N A61K48/00 A61P35/00 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12Q1/68 G01N33/53.D G01N33/53.M G01N33/566 C12N5/00.A		
F-TERM分类号	4B024/AA01 4B024/AA12 4B024/BA31 4B024/BA44 4B024/BA80 4B024/CA03 4B024/CA04 4B024/DA06 4B024/EA04 4B024/GA11 4B024/HA12 4B024/HA15 4B063/QA18 4B063/QA19 4B063/QQ43 4B063/QR55 4B063/QR62 4B063/QS25 4B063/QS34 4B064/AG01 4B064/CA02 4B064/CA19 4B064/CC24 4B064/DA01 4B064/DA14 4B065/AA26X 4B065/AA93Y 4B065/AB01 4B065/AC14 4B065/BA02 4B065/CA24 4B065/CA45 4C084/AA13 4C084/NA10 4C084/ZB262 4C085/AA03 4C085/AA13 4C085/AA14 4C085/BB01 4C085/BB33 4C085/BB34 4C085/BB35 4C085/BB36 4C085/BB37 4C085/CC02 4C085/CC23 4C085/DD62 4C085/DD63 4C085/DD88 4C085/EE01 4C085/GG02 4C085/GG03 4C085/GG04 4C085/GG06 4C085/GG08 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA11 4H045/CA40 4H045/DA76 4H045/DA86 4H045/EA28 4H045/EA31 4H045/EA51 4H045/FA72 4H045/FA74		
优先权	60/244717 2000-10-31 US		
其他公开文献	JP2004528823A		

摘要(译)

本发明涉及新鉴定的存在于正常和癌性结肠细胞中的核酸和多肽，包括所述核酸和多肽的片段，变体和衍生物。本发明还涉及针对本发明多肽的抗体，以及本发明多肽的激动剂和拮抗剂。本发明还涉及本发明的核酸，多肽，抗体，突变体，衍生物，激动剂和拮抗剂，以及使用这些组合物的方法。这些用途包括大肠中结肠癌和非癌性疾病状态的鉴定，诊断，监测，分期，成像和治疗，结肠组织的鉴定，本发明多肽的激动剂和拮抗剂的监测和鉴定。和/或设计。