

(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2005-500007

(P2005-500007A)

(43) 公表日 平成17年1月6日(2005.1.6)

(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	4 B O 2 4
A O 1 K 67/027	A O 1 K 67/027	4 B O 6 3
A 6 1 K 45/00	A 6 1 K 45/00	4 B O 6 4
A 6 1 P 35/00	A 6 1 P 35/00	4 B O 6 5
A 6 1 P 43/00	A 6 1 P 43/00	1 O 5
		4 C O 8 4
	審査請求 未請求 予備審査請求 有	(全 181 頁) 最終頁に続く

(21) 出願番号	特願2002-550081 (P2002-550081)	(71) 出願人	398021788
(86) (22) 出願日	平成13年12月11日 (2001.12.11)		ピーイー コーポレーション (エヌワイ)
(85) 翻訳文提出日	平成15年6月12日 (2003.6.12)		アメリカ合衆国 カリフォルニア州 94
(86) 国際出願番号	PCT/US2001/047559		404 フォスター シティ, リンカーン
(87) 国際公開番号	W02002/048366		センター ドライブ 850
(87) 国際公開日	平成14年6月20日 (2002.6.20)	(74) 代理人	100102978
(31) 優先権主張番号	60/254, 553		弁理士 清水 初志
(32) 優先日	平成12年12月12日 (2000.12.12)	(74) 代理人	100108774
(33) 優先権主張国	米国 (US)		弁理士 橋本 一憲
(31) 優先権主張番号	09/739, 457	(72) 発明者	マークロフ ジェナディ
(32) 優先日	平成12年12月19日 (2000.12.19)		アメリカ合衆国 メリーランド州 ロック
(33) 優先権主張国	米国 (US)		ビル シー2-4 #21 ウェスト グー
			デ ドライブ 45 セレラ ジェノミク
			ス内
			最終頁に続く

(54) 【発明の名称】 単離ヒト輸送体タンパク質、ヒト輸送体タンパク質をコードする核酸分子、およびそれらの使用方法

(57) 【要約】

本発明は、ヒトゲノム中の遺伝子によりコードされる、本発明の輸送体ペプチドのアミノ酸配列を提供する。本発明は、特に、単離ペプチドおよび核酸分子、輸送体ペプチドのオルソログおよびパラログを同定する方法、ならびに輸送体ペプチドのモジュレータを同定する方法を提供する。

【特許請求の範囲】

【請求項 1】

下記の群より選択されるアミノ酸配列からなる単離ペプチド：

- (a) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列；
- (b) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列の対立変異体のアミノ酸配列であって、該対立変異体が、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズする核酸分子によってコードされたアミノ酸配列；
- (c) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列のオルソログのアミノ酸配列であって、該オルソログが、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズする核酸分子によってコードされたアミノ酸配列；および
- (d) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列の断片であって、少なくとも10個の連続するアミノ酸を含むアミノ酸配列の断片。

【請求項 2】

下記の群より選択されるアミノ酸配列を含む単離ペプチド：

- (a) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列；
- (b) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列の対立変異体のアミノ酸配列であって、該対立変異体が、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズする核酸分子によってコードされたアミノ酸配列；
- (c) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列のオルソログのアミノ酸配列であって、該オルソログが、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズする核酸分子によってコードされたアミノ酸配列；および
- (d) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列の断片であって、少なくとも10個の連続するアミノ酸を含むアミノ酸配列の断片。

【請求項 3】

請求項2記載のペプチドに選択的に結合する単離抗体。

【請求項 4】

下記の群より選択されるヌクレオチド配列からなる単離核酸分子：

- (a) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列をコードするヌクレオチド配列；
- (b) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列の対立変異体のアミノ酸配列をコードするヌクレオチド配列であって、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズするヌクレオチド配列；
- (c) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列のオルソログをコードするヌクレオチド配列であって、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズするヌクレオチド配列；
- (d) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列の断片をコードするヌクレオチド配列であって、該断片が、少なくとも10個の連続するアミノ酸を含むヌクレオチド配列；および
- (e) (a)～(d)のヌクレオチド配列に相補的であるヌクレオチド配列。

【請求項 5】

下記の群より選択されるヌクレオチド配列を含む単離核酸分子：

- (a) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列をコードするヌクレオチド配列；
- (b) 配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列の対立変異体

のアミノ酸配列をコードするヌクレオチド配列であって、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズするヌクレオチド配列；

(c)配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列のオルソログをコードするヌクレオチド配列であって、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子の対向鎖に、ストリンジェントな条件下でハイブリダイズするヌクレオチド配列；

(d)配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列の断片をコードするヌクレオチド配列であって、該断片が、少なくとも10個の連続するアミノ酸を含むヌクレオチド配列；および

(e) (a)～(d)のヌクレオチド配列に相補的であるヌクレオチド配列。

10

【請求項6】

請求項5記載の核酸分子を含む遺伝子チップ。

【請求項7】

請求項5記載の核酸分子を含むヒト以外のトランスジェニック動物。

【請求項8】

請求項5記載の核酸分子を含む核酸ベクター。

【請求項9】

請求項8記載のベクターを含む宿主細胞。

【請求項10】

請求項1記載のいずれかのペプチドを製造する方法であって、(a)～(d)のいずれかのアミノ酸配列をコードするヌクレオチド配列を宿主細胞内に導入する段階、およびペプチドがヌクレオチド配列から発現される条件下で宿主細胞を培養する段階を含む方法。

20

【請求項11】

請求項2記載のいずれかのペプチドを製造する方法であって、(a)～(d)のいずれかのアミノ酸配列をコードするヌクレオチド配列を宿主細胞に導入する段階、およびペプチドがヌクレオチド配列から発現される条件下で宿主細胞を培養する段階を含む方法。

【請求項12】

試料中における請求項2記載のいずれかのペプチドの存在を検出する方法であって、試料中における該ペプチドの存在を特異的に検出する検出試薬と試料を接触させる段階、および該ペプチドの存在を検出する段階を含む方法。

30

【請求項13】

試料中における請求項5記載の核酸分子の存在を検出する方法であって、ストリンジェントな条件下で該核酸分子にハイブリダイズするオリゴヌクレオチドと試料を接触させる段階、および試料中の該核酸分子とオリゴヌクレオチドが結合するかどうかを判定する段階を含む方法。

【請求項14】

請求項2記載のペプチドのモジュレータを同定する方法であって、該ペプチドを試薬と接触させる段階、および該試薬が該ペプチドの機能または活性を調節したかどうかを判定する段階を含む方法。

40

【請求項15】

試薬がペプチドを発現する発現ベクターを含む宿主細胞に投与される、請求項14記載の方法。

【請求項16】

請求項2記載のいずれかのペプチドに結合する試薬を同定する方法であって、ペプチドと試薬を接触させる段階、および接触混合物をアッセイして、ペプチドと試薬との複合体が形成されるかどうかを判定する段階を含む方法。

【請求項17】

請求項16記載の方法により同定された試薬と、薬学的に許容されるそれらの担体とを含む薬学的組成物。

50

【請求項 18】

ヒト輸送体タンパク質により媒介される疾患または症状を治療する方法であって、薬学的に有効な量の請求項16記載の方法で同定された試薬を患者に投与する段階を含む方法。

【請求項 19】

請求項2記載のペプチドの発現のモジュレータを同定する方法であって、該ペプチドを発現する細胞と試薬とを接触させる段階、および該試薬が該ペプチドの発現を調節したかどうかを判定する段階を含む方法。

【請求項 20】

配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列と少なくとも70%の相同性を持つアミノ酸配列を有する、単離ヒト輸送体ペプチド。

10

【請求項 21】

配列番号：3および配列番号：4からなる群より選択されるアミノ酸配列と少なくとも90%の相同性を持つアミノ酸配列を有する、請求項20記載のペプチド。

【請求項 22】

ヒト輸送体ペプチドをコードする単離核酸分子であって、配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子と少なくとも80%の相同性を有する核酸分子。

【請求項 23】

配列番号：1、配列番号：2、および配列番号：5からなる群より選択される核酸分子と少なくとも90%の相同性を有している、請求項22記載の核酸分子。

20

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

本出願は、2000年12月12日に提出された米国仮出願第60/254,553号（代理人整理番号 CL001014-PROV）および2000年12月19日に提出された米国特許出願第09/739,457号（代理人整理番号 CL001014）の優先権を請求する。

【0002】

発明の分野

本発明は、糖輸送体サブファミリーと関連のある輸送体タンパク質、組換えDNA分子、およびタンパク質の産生の分野に属する。本発明は特に、リガンド輸送を行う新規ペプチドおよびタンパク質、ならびにこのようなペプチドおよびタンパク質分子をコードする核酸分子を提供し、これらは全てヒト治療法ならびに診断用の組成物および方法の開発に有用である。

30

【背景技術】

【0003】

発明の背景輸送体

輸送体タンパク質は、イオンおよび高分子などの分子の細胞の内外への流れを調節することにより、細胞増殖、分化およびシグナル伝達の過程を含む、細胞の多くの異なる機能を調節する。輸送体は真核生物において事実上全ての細胞の原形質膜に認められる。輸送体は、膜電位の調節ならびに細胞膜を介した分子およびイオンの吸収および分泌を含む、様々な細胞機能を媒介する。ゴルジ装置およびエンドサイトーシス小胞の細胞内膜に存在する場合、塩素イオンチャネルなどの輸送体は細胞小器官のpHも調節する。総説については、グレガー R. (Greger, R.) (1988) Annu. Rev. Physiol. 50: 111~122を参照されたい。

40

【0004】

輸送体は一般に構造および作用様式の種類によって分類される。さらに輸送体は時には輸送される分子の種類によって分類され、例えば、糖輸送体、塩素イオンチャネル、カリウムチャネルなどがある。単一種類の分子を輸送するチャネルには多くのクラスがある（チ

50

ヤネルの種類に関する詳細な総説は、Alexander, S.P.H.およびJ.A. Peters:「受容体および輸送体の命名法・増補版 (Receptor and transporter nomenclature supplement)」、Trends Pharmacol. Sci., Elsevier, pp. 65~68 (1997) ならびに<http://wwwbiology.ucsd.edu/~msaier/transport/titlepage2.html>に記載されている)。

【0005】

以下の一般的な分類方式が当技術分野では公知であり、これは本発見にも用いられている。

【0006】

チャンネル型輸送体

このクラスの膜貫通型チャンネルタンパク質は、細菌から高等真核生物までのあらゆる種類の生物の膜に遍在的に認められる。このタイプの輸送系は、担体を介した機構の証拠がない、膜貫通水性孔またはチャンネルを介した通過により、促進拡散(エネルギー非依存的過程による)を触媒する。これらのチャンネルタンパク質は通常、主に α -ヘリックス性貫通体からなるが、 β -ストランドも存在することがあり、チャンネルを構成することさえもある。しかし、外膜ポーリン型チャンネルタンパク質はこのクラスから除外され、代わりにクラス9に含められる。

10

【0007】

担体型輸送体

輸送系は、担体を介した過程を利用して単輸送(促進拡散により単一種が輸送される)、対向輸送(化学浸透圧エネルギー以外の直接的形態のエネルギーとは共役していない、強固に共役した過程により、2つまたはそれ以上の種が逆向きに輸送される)および/または共輸送(化学浸透圧エネルギー以外の直接的形態のエネルギーとは共役していない、強固に共役した過程により、2つまたはそれ以上の種が同じ向きに輸送される)を触媒する場合、このクラスに含まれる。

20

【0008】

ピロリン酸結合加水分解駆動型の能動輸送体

輸送系は、ATPまたは別のヌクレオシド三リン酸におけるピロリン酸または末端ピロリン酸結合を加水分解して、1つまたは複数の溶質の能動的な取り込みおよび/または排出を駆動する場合、このクラスに含まれる。輸送タンパク質は一過性にリン酸化されてもリン酸化されなくてもよいが、基質はリン酸化されない。

30

【0009】

PEP依存的なホスホリル転移駆動型グループ輸送体

細菌ホスホエノールピルビン酸：糖ホスホトランスフェラーゼ系の輸送系は、このクラスに含まれる。細胞外の糖に由来する反応産物は細胞質糖リン酸である。

【0010】

脱炭酸駆動型能動輸送体

細胞質基質の脱炭酸によって溶質(例えば、イオン)の取り込みまたは排出を駆動する輸送系は、このクラスに含まれる。

【0011】

酸化還元駆動型能動輸送体

還元型基質から酸化型基質への電子流によって賦活される溶質(例えば、イオン)の輸送を駆動する輸送系は、このクラスに含まれる。

40

【0012】

光駆動型能動輸送体

光エネルギーを利用して溶質(例えば、イオン)の輸送を駆動する輸送系は、このクラスに含まれる。

【0013】

機械的駆動型能動輸送体

輸送系は、イオン(または他の溶質)が膜を介して電気化学的勾配の下方に流れるのを可能にすることにより、細胞または細胞小器官の移動を駆動する場合、このクラスに含まれる

50

る。

【0014】

(b-構造の) 外膜ポーリン

これらのタンパク質は、溶質の膜を介したエネルギー非依存的な通過を通常可能にする、膜貫通孔またはチャネルを形成する。これらのタンパク質の膜貫通部分は、b-バレルを形成するb-ストランドのみからなる。これらのポーリン型タンパク質は、グラム陰性細菌、ミトコンドリアおよび真核生物色素体の外膜に認められる。

【0015】

メチルトランスフェラーゼ駆動型能動輸送体

現在、このカテゴリーに入るのは、唯一特徴づけられているタンパク質である、Na⁺輸送メチルトetraヒドロメタノプテリン：コエンザイムMメチルトランスフェラーゼである。

10

【0016】

非リボソーム合成型チャネル形成ペプチドまたはペプチド様分子

これらの分子は通常、L-アミノ酸およびD-アミノ酸ならびに乳酸などの他の低分子構成単位の連鎖であり、オリゴマー膜貫通型イオンチャネルを形成する。電位は膜貫通型チャネルの集合を促進することによって、チャネル形成を誘導しうる。これらのペプチドはしばしば細菌および真菌によって、生物闘争の作用物質として作られる。

【0017】

非タンパク質性輸送複合体

タンパク質またはペプチドから構成されず、いずれかに由来しない生体膜中のイオン伝導性物質は、このカテゴリーに入る。

20

【0018】

機能的な特徴づけはなされているが、配列データの無い輸送体

ファミリーの指定はできないが、特別な生理的意義のある輸送体はこのカテゴリーに含まれると考えられる。

【0019】

ファミリーのメンバーが輸送体として確立されていない、推定の輸送体

推定の輸送体タンパク質ファミリーはこの項目に分類され、メンバーの輸送機能が確立されれば別のものに分類され、提唱された輸送機能が反証されればTC分類系から除外されると考えられる。これらのファミリーは、輸送機能は示唆されているがそのような機能に関する証拠に説得力がない、1つまたは複数のメンバーを含む。

30

【0020】

補助輸送体タンパク質

1つまたは複数の生体膜を介した輸送を何らかの様式で促進するが、それ自体は輸送に直接関与しないタンパク質は、このクラスに含まれる。これらのタンパク質は常に1つまたは複数の輸送タンパク質とともに機能する。これらは輸送とのエネルギー共役と連結した機能を提供することができ、複合体形成において構造的役割を果たすことができ、調節機能を果たすことができる。

【0021】

分類不明の輸送体

分類不明の輸送タンパク質ファミリーはこの項目に分類され、輸送過程およびエネルギー共役機序が特徴づけられれば別のものに分類されると考えられる。これらのファミリーは、輸送機能は確立されているが輸送様式またはエネルギー共役機序が不明である、少なくとも1つのメンバーを含む。

40

【0022】

イオンチャネル

重要なタイプの輸送体はイオンチャネルである。イオンチャネルは、細胞の内外へのイオンの流れを調節することにより、多くの異なる細胞増殖、分化およびシグナル伝達の過程を調節する。イオンチャネルは真核生物において事実上全ての細胞の原形質膜に認められる。イオンチャネルは、膜電位の調節ならびに上皮膜を介したイオンの吸収および分泌を

50

含む、様々な細胞機能を媒介する。ゴルジ装置およびエンドサイトーシス小胞の細胞内膜に存在する場合、塩素イオンチャネルなどのイオンチャネルは細胞小器官のpHも調節する。総説については、グレガー R. (Greger, R.) (1988) *Annu. Rev. Physiol.* 50: 111~122を参照されたい。

【0023】

イオンチャネルは一般に構造および作用様式の種類によって分類される。例えば、細胞外リガンド依存性チャネル (ELG) は5つのポリペプチド性サブユニットから構成され (各サブユニットは4つの膜貫通ドメインを有する)、細胞外リガンドのチャネルとの結合によって活性化される。さらに、チャネルは時には輸送されるイオンの種類、例えば、塩素イオンチャネル、カリウムチャネルなどによっても分類される。単一種類のイオンを輸送するチャネルには多くのクラスがある (チャネルの種類に関する詳細な総説は、Alexander, S.P.H.およびJ.A. Peters: 「受容体およびイオンチャネルの命名法・増補版 (Receptor and ion channel nomenclature supplement)」、*Trends Pharmacol. Sci.*, Elsevier, pp. 65~68 (1997) ならびに<http://wwwbiology.ucsd.edu/~msaier/transport/toc.html>に認められる)。

10

【0024】

イオンチャネルには構造に基づく多くの種類がある。例えば、多くのイオンチャネルは以下の群の1つに入る: 細胞外リガンド依存性チャネル (ELG)、細胞内リガンド依存性チャネル (ILG)、内向き整流チャネル (INR)、細胞間 (ギャップ結合) チャネル、および電位依存性チャネル (VIC)。輸送されるイオンの種類、細胞の位置および薬剤感受性に

20

【0025】

細胞外リガンド依存性チャネル、すなわちELGは、一般に5つのポリペプチドサブユニットから構成される。アンウィン N. (Unwin, N.) (1993)、*Cell* 72: 31~41; アンウィン N. (Unwin, N.) (1995)、*Nature* 373: 37~43; ユチョウ F. (Hucho, F.) ら (1996) *J. Neurochem.* 66: 1781~1792; ユチョウ F. (Hucho, F.) ら (1996) *Eur. J. Biochem.* 239: 539~557; アレキサンダー S.P.H. (Alexander, S.P.H.) およびJ.A. ピーターズ (J.A. Peters) (1997)、*Trends Pharmacol. Sci.*, Elsevier, pp. 4~6; 36~40; 42~44; ならびにシュウ H. (Xue, H.) (1998) *J. Mol. Evol.* 47: 323~333。各サブユニットは4つの膜貫通領域を有する: これはELGファミリーのタンパク質の他のメンバーを同定する手段として役立つ。ELGはリガンドと結合し、それに反応してイオン流を調節する。ELGの例には、神経伝達物質受容体ファミリーのタンパク質のほとんどのメンバー、例えば、GABA_A受容体が含まれる。このファミリーのイオンチャネルの他のメンバーには、グリシン受容体、リアノジン受容体およびリガンド依存性カルシウムチャネルが含まれる。

30

【0026】

電位依存性イオンチャネル (VIC) スーパーファミリー

VICファミリーのタンパク質は、広範囲の細菌、古細菌および真核生物に認められるイオン選択的チャネルタンパク質である。ヒル B. (Hille, B.) (1992)、「興奮性膜のイオンチャネル (Ionic Channels of Excitable Membranes)」第2版、Sinaur Assoc. Inc., Pubs., サンダーランド、マサチューセッツの第9章: 「チャネルタンパク質の構造 (Structure of channel proteins)」; 第20章: 「進化および多様性 (Evolution and diversity)」; シグワース F.J. (Sigworth, F.J.) (1993)、*Quart. Rev. Biophys.* 27: 1~40; サルコフ L. (Salkoff, L.) およびT. イェーグラ (T. Jegla) (1995)、*Neuron* 15: 489~492; アレキサンダー S.P.H. (Alexander, S.P.H.) ら (1997)、*Trends Pharmacol. Sci.*, Elsevier, pp. 76~84; ヤン L.Y. (Jan, L.Y.) ら (1997)、*Annu. Rev. Neurosci.* 20: 91~123; ドイル D.A. (Doyle, D.A.) ら (1998) *Science* 280: 69~77; テルロー H. (Terlau, H.) およびW. スチューマー (W. Stuhmer) (1998)、*Naturwissenschaften* 85: 437~444。それらはホモオリゴマー構造またはいくつかの異なるサブユニ

40

50

ットを有するヘテロオリゴマー構造であることが多いが（例えば、a1-a2-d-b Ca^{2+} チャンネル、 ab_1b_2 Na^+ チャンネル、または $(\text{a})_4\text{-b}$ K^+ チャンネル）、チャンネルおよび一次受容体は通常、a（またはa1）サブユニットと結合する。機能的に特徴づけられたメンバーは K^+ 、 Na^+ または Ca^{2+} に対して特異的である。 K^+ チャンネルは通常、各々のa-サブユニットが6つの膜貫通体（TMS）を有するヘテロ四量体構造からなる。 Ca^{2+} および Na^+ チャンネルのa1およびaサブユニットはそれぞれほぼ4倍の大きさで、親水性ループによって隔てられた6個のTMSをそれぞれ備えた4個の単位を有し、合計24個のTMSを持つ。これらの大きなチャンネルタンパク質は、ほとんどの K^+ チャンネルのホモ四量体構造に相当する、ヘテロ四量体構造を形成する。 Ca^{2+} および Na^+ チャンネルの4つの単位はいずれも、ホモ四量体型 K^+ チャンネルの単一の単位と相同である。真核生物チャンネルを介したイオン流は、一般に膜内外電位差によって制御されるが（電位感受性という名称はこれによる）、リガンドまたは受容体の結合によって制御されるものもある。

10

【0027】

VICファミリーの K^+ 選択的チャンネルタンパク質と推定されるものが、いくつか原核生物で同定されている。その1つであるストレプトマイセス・リビダンス（*Streptomyces lividans*）のKcsA K^+ チャンネルは、分解能3.2 Åで解明されている。このタンパク質は、4つの同一なサブユニットを有し、それぞれが2つの膜貫通ヘリックスを有し、逆テント型または円錐型に配置されている。錐体は外端で「選択性フィルター」Pドメインを支える。選択性フィルターは細く、長さは12 Åに過ぎないが、チャンネルの残りの部分は幅広く、疎水性残基が並ぶ。水で満たされた大きな空洞とヘリックス双極子が孔内の K^+ を安定化する。選択性フィルターは互いに約7.5 Åを隔てて結合している2つの K^+ イオンを有する。イオンの伝導は静電的引力と斥力とのバランスに起因すると提唱されている。

20

【0028】

真核生物では、VICファミリーチャンネルの各タイプには、薬理学的および電気生理学的データに基づくサブタイプがいくつかある。すなわち、 Ca^{2+} チャンネルには5個のタイプがある（L、N、P、QおよびT）。 K^+ チャンネルには少なくとも10個のタイプがあり、それぞれは異なる刺激に対して異なる様式で反応する：電位感受性型 [Ka、Ky、Kvr、KvsおよびKsr]、 Ca^{2+} 感受性型 [BK_{Ca}、IK_{Ca}およびSK_{Ca}] ならびに受容体共役型 [K_MおよびK_{ACH}]。Na⁺チャンネルには少なくとも6個のタイプがある（I、II、III、μ1、H1およびPN3）。原核生物および真核生物の両方に由来する四量体チャンネルが公知であり、これらのチャンネルでは、各a-サブユニットが6個ではなく2個のTMSを有し、これら2つのTMSが電位感受性チャンネルタンパク質に認められる6つのTMS単位のうちTMS 5および6と相同である。S.リビダンス（*S. lividans*）のKcsAはこの種の2 TMS型チャンネルタンパク質の一例である。これらのチャンネルには、K_{Na}（Na⁺活性化）およびK_{Vo1}（細胞容積感受性） K^+ チャンネル、ならびに酵母のTok1 K^+ チャンネル、マウスのTWIK-1内向き整流 K^+ チャンネル、およびマウスのTREK-1 K^+ チャンネルのような、遠い関係にあるチャンネルも含まれる。VICファミリーのタンパク質との配列類似性が十分でないため、単一のPドメインおよび2つの隣接するTMSを有する内向き整流 K^+ IRKチャンネル（ATP調節性；G-タンパク質活性化型）は異なるファミリーに分類される。しかし、P領域における実質的な配列類似性により、それらは相同であることが示唆される。VICファミリーメンバーのb、gおよびdサブユニットは、存在する場合、チャンネルの活性化/不活性化において調節的役割を果たすことが多い。

30

40

【0029】

上皮Na⁺チャンネル（ENaC）ファミリー

ENaCファミリーは、配列が決定された24個を上回るタンパク質からなる（Canessa, C.M. ら（1994）、Nature 367: 463~467、Le, T.およびM.H. Saier, Jr.（1996）、Mol. Membr. Biol. 13: 149~157；Garty, H.およびL.G. Palmer（1997）、Physiol. Rev. 77: 359~396；Waldmann, R.ら（1997）、Nature 386: 173~177；Darboux, I.ら（1998）、J. Biol. Chem. 273: 9424~9429；Firsov, D.ら（1998）、EMBO J. 17: 344~352；Horisberger, J.-D.（1998）、Curr. Opin. Struct. Biol. 10: 443~449）。これらはいずれも動物由来であり、他の真核生物および細菌における相同体は認められていない。上皮細胞由

50

来の脊椎動物ENaCタンパク質は、系統樹の上で互いに密に集まっている：電位感受性のないENaC相同体も脳に認められる。デジェネリンを含む、配列が決定された11個の線虫 (*C. elegans*) タンパク質は、互いと同様に脊椎動物タンパク質と遠い関係にある。これらのタンパク質の少なくともいくつかは、接触感受性のための機械変換複合体の一部を形成する。相同なヘリックス・アスベルサ (*Helix aspersa*) (FMRF-アミド) 活性化型Na⁺チャネルは、配列が決定された最初のペプチド神経伝達物質依存性イオンチャネル型受容体である。

【0030】

このファミリーのタンパク質メンバーは全て、外見上同一の空間配置を示し、それぞれ、細胞の内側にN末端およびC末端、2つの両親媒性膜貫通部分、ならびに大きな細胞外ループを有する。細胞外ドメインは高度に保存された多数のシステイン残基を含む。これらは受容体機能に役立つと提唱されている。

10

【0031】

哺乳類ENaCはNa⁺バランスの維持および血圧の調節に重要である。3つの相同なENaCサブユニットである、およびが集合し、Na⁺選択性の高いチャネルを形成することが示されている。3つのサブユニットの化学量論は、1つのヘテロ四量体構造において 2、1、1である。

【0032】

グルタミン酸依存性イオンチャネル (GIC) ファミリーの神経伝達物質受容体

GICファミリーのメンバーはヘテロ五量体複合体であり、5つのサブユニットはそれぞれ800~1000アミノアシル残基長である (Nakanishi, N.ら (1990)、Neuron 5: 569~581; Unwin, N. (1993)、Cell 72: 31~41; Alexander, S.P.H.およびJ.A. Peters (1997) Trends Pharmacol. Sci., Elsevier, pp. 36~40)。これらのサブユニットは、推定のa-ヘリックスとして膜を3回または5回貫通し、N末端 (グルタミン酸結合ドメイン) は細胞外に位置し、C末端は細胞質に位置する。これらはリガンド依存性イオンチャネルと遠い関係にある可能性があり、そうであれば、これらは膜貫通領域中に実質的なb-構造を有する可能性がある。しかし、これらの2つのファミリー間の相同性は、配列比較のみを基にしては確定できない。サブユニットは6つのサブファミリー：a、b、g、d、eおよびzに分類される。

20

【0033】

GICチャネルは、以下の3つのタイプに分かれる：(1) a-アミノ-3-ヒドロキシ-5-メチル-4-イソキサゾールプロピオン酸 (AMPA) 選択的グルタミン酸受容体 (2) カイニン酸選択的グルタミン酸受容体および (3) N-メチル-D-アスパラギン酸 (NMDA) 選択的グルタミン酸受容体。AMPAおよびカイニン酸クラスのサブユニットは互いに35%~40%の同一性を示すが、NMDA受容体のサブユニットは前記サブユニットに22%~24%の同一性を示す。これらはグラム陰性細菌のABC型取り込み透過酵素の周辺質グルタミン受容体およびグルタミン酸受容体に相同な、細胞外グルタミン酸結合ドメインである、大きなN末端を有しする。GICファミリーの公知のメンバーは全て動物由来である。異なるタイプのチャネル (受容体) は異なるイオン選択性および伝導特性を示す。NMDA選択的な高伝導度チャネルは、一価陽イオンおよびCa²⁺の透過性が高い。AMPA選択的およびカイニン酸選択的なイオンチャネルは主として一価陽イオン透過性であり、Ca²⁺に対する透過性はわずかに過ぎない。

30

40

【0034】

塩素イオンチャネル (CIC) ファミリー

CICファミリーは、グラム陰性菌およびグラム陽性菌、シアノバクテリア、古細菌、酵母、植物ならびに動物に由来する、配列が決定された数十ものタンパク質からなる大規模なファミリーである (Steinmeyer, K.ら (1991)、Nature 354: 301~304; Uchida, S.ら (1993)、J. Biol. Chem. 268: 3821~3824; Huang, M.-E.ら (1994)、J. Mol. Biol. 242: 595~598; Kawasaki, M.ら (1994)、Neuron 12: 597~604; Fisher, W.E.ら (1995)、Genomics. 29: 598~606; およびFoskett, J.K. (1998)、Annu. Rev. Physiol. 60: 689~717)。これらのタンパク質は、インフルエンザ菌 (*Haemophilus influenzae*)、マ

50

イコプラズマ・ゲニタリウム (*Mycoplasma genitalium*) および肺炎マイコプラズマ (*Mycoplasma pneumoniae*) のゲノム中にはコードされていないが、本質的に遍在性である。配列が決定されたタンパク質のサイズは、395アミノアシル残基 (*M. jannaschii*) から988残基 (ヒト) までさまざまである。いくつかの生物は複数のCICファミリーパラログを含む。例えば、シネコキスチス (*Synechocystis*) は2つのパラログを含み、1つは451残基長であり、もう1つは899残基長である。シロイヌナズナ (*Arabidopsis thaliana*) は配列が決定されたパラログを少なくとも4つ含み (775~792残基)、ヒトも少なくとも5つのパラログを含んでおり (820~988残基)、線虫 (*C. elegans*) も少なくとも5つを含む (810~950残基)。哺乳類では9つのメンバーが公知であり、対応する遺伝子のうちの3つにおける変異はヒト疾患の原因となる。大腸菌 (*E. coli*)、メタン生成菌 (*Methanococcus jannaschii*) および出芽酵母 (*Saccharomyces cerevisiae*) は、各CICファミリーのメンバーを1つだけ有する。より長いシネコキスチスのパラログを例外として、細菌タンパク質は全て小型であるが (395~492残基) が、真核生物タンパク質は全て大型である (687~988残基)。これらのタンパク質は10個~12個の推定の α -ヘリックス性膜貫通体 (TMS) を示し、ホモ二量体として膜内に存在するようである。ファミリーのメンバーの1つであるシビレイのCIC-0は、サブユニット毎に1つずつ、計2つのチャネルを有すると報告されているが、他のものには1つしかないと考えられている。

【0035】

機能的に特徴付けられたCICファミリーのメンバーは全て塩素イオンを輸送し、電位で調節される過程で輸送するものもある。これらのチャネルは様々な生理機能に役立つ (細胞容積の調節; 膜電位の安定化; シグナル伝達; 経上皮輸送など)。ヒトにおける種々のホモログは異なる陰イオン選択性を示し、すなわち、CIC4およびCIC5は $\text{NO}_3^- > \text{Cl}^- > \text{Br}^- > \text{I}^-$ の順序の伝導度を共有するが、CIC3の選択性は $\text{I}^- > \text{Cl}^-$ である。CIC4およびCIC5チャネルなどは、+20mVよりも高い電位でのみ外向き整流電流を示す。

【0036】

動物の内向き整流 K^+ チャネル (IRK-C) ファミリー

IRKチャネルは、VICファミリーのチャネルタンパク質に特徴的な唯一のPDメイン、および2つの隣接した膜貫通体を備えた、「最小チャネル形成構造」を有する (Shuck, M.E.ら (1994)、J. Biol. Chem. 269: 24261~24270; Ashen, M.D.ら (1995)、Am. J. Physiol. 268: H506~H511; Salkoff, L.およびT. Jegla (1995)、Neuron 15: 489~492; Aguilar-Bryan, L.ら (1998)、Physiol. Rev. 78: 227~245; Ruknudin, A.ら (1998)、J. Biol. Chem. 273: 14165~14171)。これらは膜内にホモオリゴマーまたはヘテロオリゴマーとして存在しうる。これらは K^+ 流を細胞外に流出させるよりも細胞内に流入させる傾向が強い。電位依存性は外部 K^+ 、内部 Mg^{2+} 、内部ATPおよび/またはGタンパク質による調節を受ける。IRKチャネルのPDメインはVICファミリーのものとある程度の配列類似性を示すが、この配列類似性は相同性を確定するには不十分である。内向き整流は細胞膜電位の設定において役割を果たし、脱分極時のこれらのチャネルの閉鎖により、プラトー相を伴う活動電位が長時間持続することが可能になる。内向き整流には、VICファミリーチャネルに認められる固有の電位感知ヘリックスはない。少数の場合、例えばKir1.1aおよびKir6.2の場合は、ABCスーパーファミリーのメンバーとの直接的な相互作用により、ATP感受性を含む特有の機能的および調節的な特性がヘテロマー複合体に付与されると提唱されている。SUR1スルホニル尿素受容体 (spQ09428) はATPに反応してKir6.2チャネルを調節するABCタンパク質であり、CFTRはKir1.1aを調節しうる。SUR1における変異は、膵臓における未制御のインスリン分泌を特徴とする、常染色体劣性障害である、新生児期の家族性持続性高インスリン性低血糖 (PHHI) の原因である。

【0037】

ATP依存性陽イオンチャネル (ACC) ファミリー

ACCファミリー (P2X受容体とも呼ばれる) のメンバーは、多くの種類のニューロンからエキソサイトーシスによって放出される機能性神経伝達物質である、ATPに反応する (North, R.A. (1996)、Curr. Opin. Cell Biol. 8: 474~483; Soto, F., M. Garcia-Guzmanお

よびW. Stuhmer (1997)、J. Membr. Biol. 160: 91~100)。これらは薬理学的特性に基づいて7つの群 (P2X₁ ~ P2X₇) に分類されている。これらのチャネルはニューロン-ニューロン接続およびニューロン-筋接合部で働き、血圧および痛覚の制御において役割を果たしう。これらはリンパ球および血小板の生理的作用においても機能する。これらは動物でのみ認められる。

【0038】

ACCファミリーのタンパク質は配列が非常に類似しているが(35%を超える同一性)、主としてC末端ドメインに位置する、380個~1000個の様々な長さのアミノアシル残基を各サブユニットごとに有する。これらは2つの膜貫通体を有し、1つはN末端からの約30~50残基であり、もう1つは近傍の320~340残基である。これら2つの膜貫通体(約270残基)の間にある細胞外受容体ドメインはよく保存されており、保存されたグリシルおよびシステイル残基を数多く有する。親水性C末端の長さは25~240残基とさまざまである。これらは、(a)細胞内に位置するN末端およびC末端、(b)2つの推定の膜貫通体、(c)大きな細胞外ループドメイン、ならびに(d)保存された多くの細胞外システイル残基を有する点で、上皮Na⁺チャネル(ENaC)タンパク質と空間配置が類似している。しかし、ACCファミリーのメンバーにはそれらとの明らかな相同性はない。ACCチャネルはおそらくヘテロ多量体またはホモ多量体であり、小型の一価陽イオン(Me⁺)を輸送する。Ca²⁺を輸送するものもある;少数は低分子代謝産物も輸送する。

10

【0039】

リアノジン-イノシトール1,4,5-三リン酸受容体Ca²⁺チャネル(RIR-CaC)ファミリー
リアノジン(Ry)感受性およびイノシトール1,4,5-三リン酸(IP3)感受性のCa²⁺放出チャネルは、動物細胞内の細胞内貯蔵部位からのCa²⁺の放出において機能し、それによって様々なCa²⁺依存的生理過程を調節する(Hasan, G.ら(1992) Development 116: 967~975; Michikawa, T.ら(1994)、J. Biol. Chem. 269: 9184~9189; Tunwell, R.E.A. (1996)、Biochem. J. 318: 477~487; Lee, A.G. (1996) Biomembranes, Vol. 6、「膜貫通型受容体およびチャネル(Transmembrane Receptors and Channels)」(A.G. Lee編)、J Al Press, Denver, CO., pp 291~326; Mikoshiba, K.ら(1996) J. Biochem. Biomem. 6: 273~289)。Ry受容体は主に筋細胞の筋小胞体(SR)膜に存在し、IP3受容体は主に脳細胞の小胞体(ER)膜に存在し、ここでチャネルが活性化(開口)するとCa²⁺を細胞質に放出させる。

20

30

【0040】

Ry受容体はジヒドロピリジン感受性Ca²⁺チャネルの活動の結果として活性化される。後者は電位感受性イオンチャネル(VIC)ファミリーのメンバーである。ジヒドロピリジン感受性チャネルは筋組織のT管系に存在する。

【0041】

Ry受容体はホモ四量体複合体であり、各サブユニットは500,000ダルトン(約5,000アミノアシル残基)を上回る分子サイズを示す。これらは推定のα-ヘリックス性膜貫通体(TMS)を6つ備えたC末端ドメインを有する。VICファミリーのメンバーについて示唆されているように、推定の孔形成配列が5番目および6番目のTMSの間に存在する。大きなN末端親水性ドメインおよび小さなC末端親水性ドメインが細胞質に位置する。低分解能の三次元構造データが利用できる。哺乳類は、哺乳類種の分岐進化の前におそらく遺伝子の重複および分岐によって生じた、少なくとも3つのアイソフォームを有する。ヒトおよび線虫(Caenorabditis elegans)にはホモログが存在する。

40

【0042】

IP₃受容体は多くの点でRy受容体と類似している。(1)これらはホモ四量体複合体であり、各サブユニットは300,000ダルトン(約2,700アミノアシル残基)を上回る分子サイズを示す。(2)これらはRy受容体のものと相同なC末端チャネルドメインを有する。(3)チャネルドメインに推定のTMSが6つあり、TMS 5および6の間に推定のチャネル内層領域がある。(4)大きなN末端ドメインおよび小さなC末端尾部が細胞質に面している。(5)これらはチャネルドメインの細胞質外ループ上に共有結合した糖質を有する。(6)これらに

50

は現在、異なる調節を受け、組織分布の異なる哺乳類で認められている3つのアイソフォーム（1型、2型および3型）がある。

【0043】

IP₃受容体は、N末端IP₃結合ドメイン、中央の結合ドメインまたは調節ドメイン、およびC末端チャネルドメインという、3つのドメインを有する。チャネルはIP₃結合によって活性化され、Ry受容体と同じく、IP₃受容体チャネルの活性は、種々のプロテインキナーゼによって触媒される調節ドメインのリン酸化によって調節される。これらは脳内の様々な細胞型の小胞体膜において主体となっているが、種々の組織に由来するいくつかの神経細胞の原形質膜でも認められている。

【0044】

Ry受容体およびIP₃受容体のチャネルドメインは一貫性のあるファミリーを構成し、外見上の構造類似性にもかかわらず、VICファミリーのタンパク質とはそれほど高い配列類似性は示さない。Ry受容体およびIP₃受容体はRIR-CaCファミリーの系統樹上に別々にクラスターを形成する。それらはいずれもショウジョウバエ（*Drosophila*）にホモログがある。ファミリーの系統樹に基づき、このファミリーはおそらく以下の順序で進化した：（1）遺伝子重複事象が起こり、無脊椎動物でRy受容体およびIP₃受容体が生じた、（2）無脊椎動物から脊椎動物が進化した、（3）2つの異なる遺伝子の重複事象の結果として、各受容体の3つのアイソフォームが生じた、（4）これらのアイソフォームは哺乳類種の分岐の前に哺乳類に移行した。

【0045】

細胞小器官の塩素イオンチャネル（O-ClC）ファミリー

O-ClCファミリーのタンパク質は、動物細胞の細胞内膜に認められるが原形質膜には認められない、電位感受性塩素イオンチャネルである（Landry, Dら（1993）、*J. Biol. Chem.* 268: 14948~14955; Valenzuela, Sら（1997）、*J. Biol. Chem.* 272: 12575~12582; およびDuncan, R.R.ら（1997）、*J. Biol. Chem.* 272: 23880~23886）。

【0046】

これらはヒト核膜内に認められ、ウシタンパク質をアフリカツメガエル（*Xenopus laevis*）卵母細胞で発現させると、ミクロソームに向かうが原形質膜には向かわない。これらのタンパク質は膜電位の調節ならびに腎臓における経上皮イオン吸収および分泌において機能すると考えられる。これらは推定の細胞質N末端およびC末端、ならびにグリコシル化されうる大きな管腔ループを備えた、 α -ヘリックス性膜貫通体（TMS）を2つ有する。ウシタンパク質は437アミノアシル残基長であり、223位~239位および367位~385位に推定のTMSを2つ有する。ヒト核タンパク質ははるかに小さい（241残基）。線虫（*C. elegans*）ホモログは260残基長である。

【0047】

糖輸送体

有機基質（糖、アミノ酸、カルボン酸および神経伝達物質）は、Na⁺共輸送によって真核細胞内へ能動的に輸送される。輸送タンパク質には同定されているものもあり（例えば、腸管刷子縁Na⁺/グルコース輸送体およびNa⁺/プロリン輸送体、ならびに脳Na⁺/Cl⁻/GABA輸送体）、それらの活性部位を特定し、それらの高次構造の状態を調べることに於いて進展があった。典型的なNa⁺駆動型輸送体は腸管刷子縁Na⁺/グルコース共輸送体であり、共輸送体における欠陥は先天性グルコース-ガラクトース吸収不良症候群の原因である。

【0048】

共輸送体は、膜タンパク質の主要なクラスであり、これは典型的には13個の膜貫通ヘリックスを持つ。それらは、膜を隔てた分子（栄養素、神経伝達物質、オスモリート（osmolyte）およびイオン）の濃度の原因となる。例えば、アミノ酸、糖、ヌクレオシドおよびビタミンの共輸送体がある。

【0049】

Na⁺/グルコース共輸送体（SGLT1）は、ボブ クレイン（Bob Crane）によって1960年に報告された。ナトリウム依存性グルコース輸送は、動物の腎臓および腸の両方で起こる。こ

10

20

30

40

50

これらの輸送体の両方は、互いに近い類似性を示す。

【0050】

これらの輸送体は、多機能性であることが報告され、以下の4つの方法で機能することが示されている：1) 脱共役型受動Na⁺輸送、2) 下り坂水輸送 (Downhill water transport)、3) Na⁺輸送および基質輸送、4) Na⁺輸送、水輸送および基質輸送。本発明に関する更なる情報については、マツオ (Matsuo) ら、Biochem Biophys Res Commun 1997 Sep 8; 238 (1): 126-9を参照のこと。

【0051】

哺乳動物細胞へのヘキソース輸送は、特異的な機能を有し、かつ組織分布が異なる44kD~55kDの膜タンパク質の小さいファミリーのメンバーによって触媒されている。観察されるヘキソース輸送体は、12個の膜貫通ヘリックス、およびいくつかの重要な保存残基を有する。保存されたGLUT配列を含むクローンについてESTデータベースを検索し、その後ラットの組織についてのスクリーニング、および5'末端RACEを行うことによって、イバーソン (Ibberson) らおよびドージェ (Doege) らは、新規のグルコース輸送体をコードする齧歯類とヒトのcDNAを同定した (Ibberson, M. ら、J. Biol. Chem. 275: 4607-4612、2000、PubMed ID: 10671487; およびDoege, H. ら、J. Biol. Chem. 275: 16275-16280、2000、PubMed ID: 10821868)。ヒトcDNAは、GLUT8またはGLUTX1と指定された、477アミノ酸と推測されるタンパク質をコードし、マウス配列と85%の配列相同性を示す。イバーソンらによって、内部移行シグナルとして働くことができるN末端のジロイシンモチーフがアラニンに変異導入されない限り、カエル卵母細胞で発現する約37kDのラットGlutx1はグルコースの取り込みが不可能であることが見出された。免疫蛍光解析によって、Glutx1 (LL-AA) が原形質膜において発現しているのに対して、Glutx1が細胞内で発現していることが実証された。明らかに対照的に、ドージェらは、GLUT8を発現する細胞からの膜調製物がグルコース存在下でサイトカラシンBを結合できず、リボソームに再構成された場合にD-グルコース輸送活性が上昇することを見出した。ウェスタンブロット解析によって、ドージェらは、ヒトGLUT8が42kDタンパク質として発現することを決定した。ノーザンブロット解析により、精巣において最も強い発現をし、甲状腺を除く他の組織において中程度の発現をする、2.4kbの転写物の発現が明らかとなった。さらに、ドージェらは、2人の精巣癌患者において、またはエストロゲンで治療した4人の患者の精巣組織において、GLUT8が検出不可能であることを見出した。彼らは、思春期および成体ではあるが思春期前ではないラットの精巣でGlut8 mRNAが検出可能であることを見出した。

【0052】

移植前マウス初期胚におけるグルコース輸送活性は、公知の促進性 (facilitative) グルコース輸送体GLUT1 (SLC2A1; 138140)、GLUT2 (SLC2A2; 138160)、およびGLUT3 (SLC2A3; 138170) に起因していた。GLUT1は、1細胞胚期から始まり胚盤胞期で終わる、移植前の期間を通じて存在する。GLUT2およびGLUT3は、最初に8細胞期後期に発現し、残りの移植前の期間で存在し続ける。3つの輸送体すべての同時出現は、胚燃料代謝 (embryonic fuel metabolism) がクレブス回路および酸化的リン酸化によるラクテートおよびピルベートの酸化から、糖新生によるグルコースの嫌氣的代謝に切り換わる、哺乳類の発生における重要な時期と一致する。哺乳類の移植前胚盤胞は、唯一公知のインシュリン制御輸送体であるGLUT4 (SLC2A4; 138190) が欠如しているにもかかわらず、インシュリン刺激性のグルコース取り込みを示す。カラヤノポウロス (Carayannopoulos) らは、マウスのGlut8がGlut1、Glut3、およびGlut4と20%~25%のアミノ酸配列同一性を示すことを見出した (Carayannopoulos, M. O. ら、Proc. Nat. Acad. Sci. 97: 7313-7318、2000、PubMed ID: 10860996)。インシュリンは、このタンパク質の細胞内局在の変化を誘導し、これは胚盤胞へのグルコース取り込みを増加させ、この過程はアンチセンスオリゴプロープによって阻害される。この輸送体の存在は、胚盤胞発生、燃料代謝、およびその後の移植の成功に必要な場合であり得る。代替の輸送体の存在により、Glut4非存在下のインシュリン制御グルコース輸送の、他の組織における例を説明できる。

【0053】

ドージェらは、国際放射ハイブリッドマッピングコンソーシアム (International Radiation Hybrid Mapping Consortium) が第9染色体にGLUT8遺伝子が存在すること (A005N15) を突き止めたことに注目した。

【0054】

輸送体タンパク質、特に糖輸送体サブファミリーのメンバーは、薬物の作用および開発の主な標的である。したがって、これまで知られていない輸送体タンパク質の同定および特徴づけを行うことは、医薬品開発の分野にとって有意義である。本発明は、これまで同定されていないヒト輸送体タンパク質を提供することにより、最先端技術を前進させる。

【発明の開示】

【0055】

10

発明の概要

本発明は一部には、糖輸送体サブファミリーと関連のあるヒト輸送体ペプチドおよびタンパク質、ならびにそれらの対立遺伝子変異体および他の哺乳類でのオルソログの、アミノ酸配列の同定に基づく。これらの特有のペプチド配列、およびこれらのペプチドをコードする核酸配列は、ヒトの治療標的の開発のためのモデルとして用いることができ、治療用タンパク質の同定を補助することができ、ならびに輸送体を発現する細胞および組織における輸送体活性を調整する、ヒト治療薬の開発のための標的として働くことができる。図1に提供した実験データは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳での発現を示している。

【0056】

20

発明の詳細な説明

概論

本発明は、ヒトゲノムの配列決定に基づいている。ヒトゲノムの配列決定および構築に際して、配列情報を解析することによって、当技術分野において輸送体タンパク質、または輸送体タンパク質の一部であると同定および特徴付けされ、また糖輸送体サブファミリーに関連付けられるタンパク質/ペプチド/ドメインに対して、構造および/または配列の相同性を有するペプチドをコードする、ヒトゲノムの未同定の断片が明らかになった。これらの配列を用いて、付加的なゲノム配列を構築、転写し、および/またはcDNA配列を単離し、特徴付けた。この解析に基づき、本発明は、糖輸送体サブファミリーに関連するヒト輸送体ペプチドおよびタンパク質のアミノ酸配列、これらの輸送体ペプチドおよびタンパク質をコードする転写配列、cDNA配列、および/またはゲノム配列形態における核酸配列、核酸変異（対立遺伝子情報）、発現の組織分布、ならびに本発明の輸送体に対して構造または配列の相同性を有する、最も関連性の高い既知のタンパク質/ペプチド/ドメインに関する情報を提供するものである。

30

【0057】

本発明において提供されるペプチドは、従来より未知であることに加えて、商業的に重要な製品およびサービスの開発にとって有用であるという能力に基づいて、選択され得る。特に、本発明のペプチドは、糖輸送体サブファミリーにおける既知の輸送体タンパク質に対して相同性および/または構造上の相関性を有し、ならびに発現パターンが観察されることに基づいて選択される。図1の実験データは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳での発現を示す。この技術は、このファミリーのタンパク質、および本発明の遺伝子に類似した発現パターンを有するタンパク質の商業的な重要性を明確に確立するものである。本発明のペプチドについてのより特異的な性質、およびその使用については、本明細書、特に発明の背景、図面の注釈に記載され、および/または既知の糖輸送体ファミリーもしくは輸送体タンパク質サブファミリーのそれぞれについては、当技術分野で周知である。

40

【0058】

特定の態様

ペプチド分子

本発明は、輸送体ファミリーのタンパク質のメンバーであると同定されたタンパク質分子

50

をコードする核酸配列を提供するものであり、これらは糖輸送体サブファミリー(図2にタンパク質配列、図1に転写/cDNA配列、図3にゲノム配列を示す)に関連付けられる。図2にはペプチド配列が記載され、明らかな変異体、特に本明細書および図3の情報をを用いて同定される対立遺伝子変異体も記載されており、これらは、本明細書において、本発明の輸送体ペプチド、輸送体ペプチド、または本発明のペプチド/タンパク質と呼ばれる。

【0059】

本発明は、図2に示す輸送体ペプチド(図1の転写/cDNA、または図3のゲノム配列に示される核酸分子によりコードされる)のアミノ酸配列からなる、または実質的にこれらからなる、またはこれらを含む、単離ペプチドおよびタンパク質分子を提供するとともに、本技術に含まれ、作製および使用されるこれらのペプチドの全ての明らかな変異体を提供するものである。これらの変異体については、以下で詳述する。

10

【0060】

本明細書で使用されているように、ペプチドが細胞物質を実質的に含まない、または化学前駆物質もしくは他の化学物質を含まない場合に、ペプチドは「単離」または「精製」されたという。本発明のペプチドは、均一、または他の純度になるまで精製することができる。精製のレベルは使用目的に基づくと考えられる。重要な性質は、調製物中に他の成分が多量に存在していたとしても、所望のペプチドの機能を発揮できるということである(単離核酸分子の性質については、後述する)。

【0061】

いくつかの使用では、「実質的に細胞物質を含まない」とは、他のタンパク質(すなわち汚染タンパク質)を約30%(乾燥重量)未満、他のタンパク質を約20%未満、他のタンパク質を約10%未満、または、他のタンパク質を約5%未満有するペプチド調製物を含む。ペプチドが組換えにより製造される場合、培地がタンパク質調製物の容量に対して20%未満の場合には、実質的に培地を含まないとすることができる。

20

【0062】

「実質的に化学前駆物質または他の化学物質を含まない」という用語は、合成に関与した化学前駆物質または他の化学物質から分離されたペプチド調製物を含む。ある態様においては、「実質的に化学前駆物質または他の化学物質を含まない」という用語は、化学前駆物質もしくは他の化学物質を約30%(乾燥重量)未満、化学前駆物質もしくは他の化学物質を約20%未満、化学前駆物質もしくは他の化学物質を約10%未満、または化学前駆物質もしくは他の化学物質を約5%未満有する輸送体ペプチド調製物を含む。

30

【0063】

単離輸送体ペプチドは、それを天然に発現する細胞、それを発現させるために変化させた(組換えられた)細胞から精製するか、または、既知のタンパク質合成方法を用いて合成することができる。図1の実験データは、卵巣(腺癌組織)、子宮(平滑筋肉腫組織)、頸部、腎臓、腎臓癌組織(副腎腫)、胚中心B細胞、結腸、および乳児脳での発現を示す。例えば、輸送体ペプチドをコードする核酸分子は、発現ベクター中にクローニングされ、さらにこの発現ベクターは宿主細胞に導入されて、タンパク質が宿主細胞内で発現する。その後、タンパク質は標準のタンパク質精製技術を用いた適当な精製スキームによって、細胞から単離することができる。これらの多くの技術については、以下で詳述する。

40

【0064】

したがって、本発明は、図2に示されるアミノ酸配列(配列番号:3および配列番号:4)からなるタンパク質、例えば、図1に示される転写/cDNA核酸配列(配列番号:1および配列番号:2)ならびに図3に示されるゲノム配列(配列番号:5)によりコードされるタンパク質を提供するものである。このようなタンパク質のアミノ酸配列を図2に示す。このようなタンパク質の最終的なアミノ酸配列がこのアミノ酸配列である場合、タンパク質はアミノ酸配列からなる。

【0065】

本発明はさらに、図2に示されるアミノ酸配列(配列番号:3および配列番号:4)から実質的になるタンパク質、例えば、図1に示される転写/cDNA核酸配列(配列番号:1および配

50

列番号：2)ならびに図3に示されるゲノム配列（配列番号：5）によりコードされるタンパク質を提供するものである。このようなアミノ酸配列に数個の付加アミノ酸残基、例えば、最終的なタンパク質中に約1個～約100個程度の付加残基、一般的には1個～約20個の付加残基が存在する場合、タンパク質はアミノ酸配列から実質的になる。

【0066】

本発明はさらに、図2に示されるアミノ酸配列（配列番号：3および配列番号：4）を含むタンパク質、例えば、図1に示される転写/cDNA核酸配列（配列番号：1および配列番号：2）ならびに図3に示されるゲノム配列（配列番号：5）によりコードされるタンパク質を提供するものである。このアミノ酸配列が、タンパク質の最終的なアミノ酸配列の少なくとも一部である場合、タンパク質はアミノ酸配列を含む。このような場合、タンパク質はペプチドのみであるか、またはタンパク質と天然に結合しているアミノ酸残基（連続するコード配列）もしくは非相同アミノ酸残基/ペプチド配列のような付加アミノ酸分子を有することができる。このようなタンパク質は、数個の付加アミノ酸残基を有するか、または数百もしくはそれ以上の付加アミノ酸を含むことができる。本発明の輸送体ペプチドが含まれるタンパク質の好ましい種として、天然の成熟タンパク質がある。これらの様々な種のタンパク質を調製/単離する方法について、以下に簡単に述べる。

10

【0067】

本発明の輸送体ペプチドは、キメラまたは融合タンパク質を形成するために、非相同性の配列に結合することができる。このようなキメラおよび融合タンパク質は、輸送体ペプチドに対して実質的に相同性のないアミノ酸配列を有する非相同タンパク質に、機能的に結合される輸送体ペプチドを含む。「機能的に結合される」とは、輸送体ペプチドと非相同タンパク質がフレーム中で融合していることを意味する。非相同タンパク質は、輸送体ペプチドのN末端またはC末端に融合されることができる。

20

【0068】

いくつかの使用において、融合タンパク質は、輸送体ペプチド自体の活性に影響を及ぼさない。例えば、融合タンパク質には、ガラクトシダーゼ融合、酵母ツーハイブリッドGAL融合、ポリHis融合、MYC標識、HI標識およびIg融合などの酵素融合タンパク質が含まれるが、これらに限定されるものではない。このような融合タンパク質、特にポリHis融合は、組換え輸送体ペプチドの精製を容易にすることができる。ある種の宿主細胞（例えば哺乳類の宿主細胞）においては、タンパク質の発現および/または分泌は、非相同シグナル配列を用いることにより増加させることができる。

30

【0069】

キメラまたは融合タンパク質は、標準の組換えDNA技術により製造することができる。例えば、異なるタンパク質配列をコードするDNA断片は、従来技術に従ってフレーム中に共に連結される。他の態様では、融合遺伝子は、自動DNA合成機を含む従来技術により合成することが可能である。あるいは、遺伝子断片のPCR増幅にアンカープライマーを用い、2つの連続的な遺伝子断片間に相補的な突出部を形成し、その後アニーリングし、再増幅して、キメラ遺伝子配列を作製することができる（Ausubelら、「分子生物学の最新プロトコル（Current Protocols in Molecular Biology）」、1992参照）。さらに、既に融合部分（例えばGSTタンパク質）をコードした多くの発現ベクターが市販されている。輸送体ペプチドをコードする核酸を、融合部がフレーム中で輸送体ペプチドに結合するようにして、このような発現ベクター中にクローニングすることができる。

40

【0070】

以上説明したように、本発明はまた、天然のペプチド成熟形態、ペプチドの対立遺伝子/配列変異体、非天然のペプチドの組換え誘導変異体、ならびにペプチドのオルソログおよびパラログなど、本発明のタンパク質のアミノ酸配列における明らかな変異体を提供、および実施可能にするものである。このような変異体は、核酸組換え技術およびタンパク質生化学の分野で公知の技術を用いることにより、容易に生成することができる。しかし、当然のことながら、この変異体には、本発明以前に開示されているいずれのアミノ酸配列も含まれないものである。

50

【0071】

このような変異体は、本明細書に示される分子技術および配列情報を用いることにより、容易に同定/製造することが可能である。さらに、このような変異体は、本発明の輸送体ペプチドに対する配列および/または構造上の相同性に基づいて、他のペプチドと容易に区別することができる。この相同性/同一性の程度は、主に、ペプチドが機能的な変異体であるか非機能的な変異体であるか、パラログファミリー中に存在する相違量、およびオルソログ間の進化距離に基づいて判断される。

【0072】

2つのアミノ酸配列、または2つの核酸配列の同一性割合(%)を決定するために、最適な比較を行う目的で配列は整列される(例えば、最適なアライメントのために、ギャップが第一および第二アミノ酸または核酸配列の一方または両方に導入され、非相同性配列は比較を行う目的のために無視することができる)。好ましい態様としては、基準配列の少なくとも30%、40%、50%、60%、70%、80%、90%またはそれ以上が、比較目的に応じて整列化される。その後、対応するアミノ酸の位置またはヌクレオチドの位置上のアミノ酸残基またはヌクレオチドが比較される。第一配列での位置が、第二配列において対応する位置と同じアミノ酸残基またはヌクレオチドによって占められている場合、分子はその位置と同一である(ここで用いられているアミノ酸または核酸の「同一性」は、アミノ酸または核酸の「相同性」と同等である)。2つの配列間の同一性割合(%)は、配列において共有される同一配置数の関数であり、ギャップ数および各ギャップ長さを考慮し、ギャップは2つの配列の最適なアライメントのために導入される必要がある。

【0073】

2つの配列間における、配列の比較ならびに同一性割合(%)および類似性割合(%)の決定は、数学的アルゴリズムを用いて行うことができる(「計算分子生物学(Computational Molecular Biology)」、Lesk, A.M.編、Oxford University Press、New York、1988;「バイオコンピューティング:情報学およびゲノムプロジェクト(Biocomputing: Informatics and Genome Projects)」、Smith, D.W.編、Academic Press、New York、1993;「配列データのコンピュータ解析、パート1(Computer Analysis of Sequence Data, Part 1)」、Griffin, A.M.,およびGriffin, H.G.編、Humana Press、New Jersey、1994;「分子生物学における配列解析(Sequence Analysis in Molecular Biology)」、von Heinje, G., Academic Press、1987;ならびに「配列解析プライマー(Sequence Analysis Primer)」、Gribskov, M.およびDevereux, J.編、M Stockton Press、New York、1991)。好ましい態様として、2つのアミノ酸配列間の同一性割合(%)はGCGソフトウェアパッケージ(<http://www.gcg.com>で入手可能)のGAPプログラムに組み込まれたニードルマン(Needl eman)およびウンシュ(Wunsch)アルゴリズム(J. Mol. Biol. (48):444-453 (1970))を用い、Blossom62マトリックスまたはPAM250マトリックスのいずれか、ならびにギャップ重量16、14、12、10、8、6または4、および長さ重量1、2、3、4、5または6を用いて決定される。さらに好ましい態様としては、2つのヌクレオチド配列間の同一性割合(%)は、GCGソフトウェアパッケージ(<http://www.gcg.com>で入手可能)のGAPプログラム(De vereux, J.,ら、(Nucleic Acids Res. 12(1): 387 (1984))を用い、NWSgapdna. CMPマトリックス、ならびにギャップ重量40、50、60、70または80、および長さ重量1、2、3、4、5または6を用いて決定される。他の態様としては、2つのアミノ酸またはヌクレオチド配列間の同一性割合(%)は、ALIGNプログラム(バージョン2.0)に組み込まれたE. メイヤー(E. Myers)およびW. ミラー(W. Miller)のアルゴリズム(CABIOS、4:11-17 (1989))を用い、PAM120重量残基表、ギャップ長ペナルティ12、およびギャップペナルティ4を用いて決定される。

【0074】

本発明の核酸およびタンパク質配列を、例えば他のファミリーまたは関連した配列を同定するために、配列データベースに対して検索を行う「クエリー配列」としてさらに使用することができる。このような検索は、アルツシュル(Altschul)らのNBLAST、およびXBLASTプログラム(バージョン2.0)(J. Mol. Biol. 215:403-10 (1990))を用いて行うことが

できる。BLASTヌクレオチド検索は、本発明の核酸分子に相同性のあるヌクレオチド配列を得るために、NBLASTプログラムを用い、スコア (score) =100、ワード長 (wordlength) =12で行うことができる。BLASTタンパク質検索は、本発明のタンパク質に相同性のあるアミノ酸配列を得るために、XBLASTプログラムを用い、スコア=50、ワード長=3で行うことができる。比較目的のギャップアライメントを得るために、アルツシュル (Altschul) らの記載のように、ギャップBLAST (Gapped BLAST) (Nucleic Acids Res. 25(17):3389-3402 (1997))を用いることができる。BLASTおよびギャップBLASTプログラムを用いる際には、各プログラム (例えばXBLASTおよびNBLAST) の既定のパラメータを用いることができる。

【0075】

本発明のペプチドの1つを含む、プロセッシング前の全長型、ならびにプロセッシングを受けた成熟型のタンパク質は、本発明の輸送体ペプチドの1つと完全な配列同一性を有するものとして、ならびに本明細書に提供する輸送体ペプチドと同じ遺伝子座によってコードされるものとして、容易に同定することができる。図3に提示したデータが示すように、マップ位置は第1染色体上にあることが決定された。

【0076】

輸送体ペプチドの対立遺伝子変異体は、輸送体ペプチドの少なくとも一部に対して高度の (有意な) 配列相同性/同一性を有するヒトタンパク質として、ならびに本明細書に提供する輸送体ペプチドと同じ遺伝子座によってコードされるものとして、容易に同定することができる。遺伝子座は、ヒト基準 (reference human) に対して位置づけられたゲノム配列などの、図3に提供したゲノム情報に基づいて容易に決定しうる。図3に提示したデータが示すように、マップ位置は第1染色体上にあることが決定された。本明細書で用いる場合、アミノ酸配列が典型的には少なくとも約70%~80%、80%~90%、より典型的には少なくとも約90%~95%またはそれ以上相同なとき、2つのタンパク質 (またはタンパク質の領域) は有意な相同性を有する。本発明による有意に相同なアミノ酸配列は、輸送体ペプチドをコードする核酸分子と、以下にさらに詳細に説明するストリンジェントな条件下でハイブリダイズする核酸配列によってコードされると考えられる。

【0077】

図3は、本発明の輸送体タンパク質をコードする遺伝子内で見いだされた、SNPに関する情報を提供する。イントロン内ならびにORFの5'および3'領域内において、42個の異なるヌクレオチド位置でSNPが同定された。イントロン内およびORFの外側のそのようなSNPは、制御/調節要素に影響を及ぼしうる。エキソンにおける2つのSNPのうち、1つはアミノ酸配列の変化をもたらす (すなわち、非同義SNP)。これらのSNPがもたらすアミノ酸配列の変化は図3に示され、普遍的遺伝暗号および参照として図2に提供されるタンパク質配列を用いて容易に決定することができる。

【0078】

輸送体ペプチドのパラログは、輸送体ペプチドの少なくとも一部に対して、ある程度の著しい配列相同性/同一性を有し、ヒト由来の遺伝子によってコードされ、かつ同様の活性または機能を有しているものとして、容易に同定することができる。アミノ酸配列が、所与の領域またはドメインを通じて、典型的に少なくとも約60%またはそれ以上、さらに典型的には少なくとも約70%またはそれ以上の相同性を有する場合、2つのタンパク質は典型的にはパラログであると考えられる。このようなパラログは、より詳細には以下に述べられるような穏やかな条件からストリンジェントな条件下で、輸送体ペプチドをコードする核酸分子とハイブリダイズする核酸配列によりコードされると考えられる。

【0079】

輸送体ペプチドのオルソログは、輸送体ペプチドの少なくとも一部に対してある程度の著しい配列相同性/同一性を有し、他の生物由来の遺伝子によってコードされているものとして、容易に同定することができる。好ましいオルソログは、哺乳類、好ましくは霊長類から単離され、ヒトの治療標的および治療薬剤の開発のために用いられる。このようなオルソログは、より詳細には以下に述べられるような、穏やかな条件からストリンジェント

10

20

30

40

50

な条件下で、輸送体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うな核酸配列によりコードされると考えられ、これはタンパク質を生成する2つの生物の関連性の程度に依存する。

【0080】

本発明の輸送体ペプチドの非天然の変異体は、組換え技術を用いて容易に生成することができる。このような変異体には、輸送体ペプチドのアミノ酸配列中における欠失、付加、および置換によるものが含まれるが、これらに限定されるものではない。例えば、置換の1種として、保存的アミノ酸置換が挙げられる。この置換は、輸送体ペプチドにおける所与のアミノ酸が同様の特徴を持つ他のアミノ酸によって置換されるものである。保存的置換として典型的に見られるものには、脂肪族のアミノ酸Ala、Val、LeuおよびIleの中の一つから他の一つへの置換、ヒドロキシル残基SerとThr間の置換、酸性残基AspとGluとの置換、アミド残基AsnとGln間の置換、塩基性残基LysとArgとの置換、ならびに芳香族残基PheとTyrとの置換がある。どのアミノ酸変化が表現型としてサイレントになる可能性を有するかに関する指針については、ボウイ (Bowie) ら、Science 247:1306-1310 (1990) に述べられている。

10

【0081】

変異輸送体ペプチドは、完全に機能しているか、または例えばリガンド結合能、リガンド輸送能、シグナル伝達調節能などの一つもしくは複数の活性において機能が欠失していることがある。完全に機能的な変異体には、典型的に、保存的な変異、または致命的でない残基における変異もしくは致命的でない領域内での変異のみが含まれる。図2は、タンパク質分析の結果を示しており、致命的ドメイン/領域を同定するのに使用することができる。機能的変異体には、機能が変化しない、または著しい機能変化の無い類似アミノ酸の置換も含まれる。他方、このような置換は、ある程度機能に対して正または負の影響を及ぼすことがある。

20

【0082】

非機能的変異体には、典型的に、一つもしくは複数の非保存的なアミノ酸の置換、欠失、挿入、反転もしくは切断、または致命的な残基もしくは致命的な領域内での置換、挿入、反転もしくは欠失が含まれる。

【0083】

機能において必須のアミノ酸は、例えば、特定部位の突然変異誘発、またはアラニンスキャニング突然変異誘発 (Cunninghamら、Science 244:1081-1085 (1989)) 等の当技術分野における既知の方法により、特に図2に示す結果を用いて同定することができる。アラニンスキャニング突然変異誘発では、分子内の全ての残基において、単独のアラニン突然変異を導入する。この結果生じた変異分子は、その後、輸送体活性のような生物活性、またはインビトロ増殖活性分析のようなアッセイのために試験される。結合対象/基質結合にとって重要な部位は、結晶化、核磁気共鳴、または光学的親和性標識等の構造解析によって決定される (Smithら、J. Mol. Biol. 224:899-904 (1992) ; de Vosら、Science 255:306-312 (1992)) 。

30

【0084】

本発明はさらに、輸送体ペプチドの断片を提供し、このような断片を含む、およびこのような断片からなるタンパク質およびペプチドに加え、特に図2に同定された残基を含むタンパク質およびペプチドを提供するものである。しかしながら、本発明に関連する断片は、本発明より以前に公開されている断片を含むものとは見なされない。

40

【0085】

本明細書で使用されるように、断片は、輸送体ペプチドの少なくとも8個、10個、12個、14個、16個またはそれ以上の連続するアミノ酸を含む。このような断片は、輸送体ペプチドの一つもしくは複数の生物活性を保持する能力に基づいて選択されるか、または基質との結合もしくは抗原としての作用等の機能を果たす能力によって選択され得る。特に重要な断片は生物活性断片であり、これは例えば、約8個またはそれ以上の長さのアミノ酸のペプチドである。このような断片は、典型的には、例えば活性部位、膜貫通ドメインまた

50

は基質結合ドメインのような、輸送体ペプチドのドメインまたはモチーフを含むと考えられる。さらに、可能な断片としては、ドメインまたはモチーフ含有断片、可溶性ペプチド断片、免疫原性構造含有断片を含むが、これらに限定されるものではない。推定されるドメインおよび機能性部位は、当業者にとって容易に入手可能な公知のコンピュータプログラム（例えばPROSITE分析）により、容易に確認することができる。このような分析の1つによる結果を図2に示す。

【0086】

ポリペプチドは、一般に、20天然アミノ酸と呼ばれている20種のアミノ酸以外のアミノ酸をしばしば含む。さらに、末端アミノ酸を含む多くのアミノ酸は、プロセシングおよび他の翻訳後修飾等の天然の過程、または当技術分野において公知の化学修飾技術によって修飾され得る。輸送体ペプチドにおいて天然に生じる一般的な修飾については、基本的なテキスト、詳細な文献および研究論文に記述されており、これは当業者に周知である。（これらの特性のいくつかは図2において確認される）。

10

【0087】

既知の修飾としては、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付加、ヘム部分の共有結合付加、ヌクレオチドまたはヌクレオチド誘導体の共有結合付加、脂質または脂質誘導体の共有結合付加、ホスファチジルイノシトールの共有結合付加、架橋結合、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、ピログルタミン酸塩の形成、ホルミル化、 α -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、ヨウ素化、メチル化、ミリストイル化、酸化、タンパク質分解プロセシング、リン酸化、プレニル化、ラセミ化、セレノイル化、硫酸化、アルギニン化などのタンパク質へのアミノ酸の転写RNA媒介付加、およびユビキチン化を含むが、これらに限定されるものではない。

20

【0088】

このような修飾は、当業者には周知であり、科学文献に非常に詳細に記載されてきた。グリコシル化、脂質付加、硫酸化、グルタミン酸残基の α -カルボキシル化、ヒドロキシル化、およびADPリボシル化など、いくつかの特に一般的な修飾は、「タンパク質-構造と分子特性 (Proteins - Structure and Molecular Properties)」、第2版、T.E. Creighton、W. H. Freeman and Company、New York (1993) のような多くの基本テキストに記載されている。この点に関する詳細な総説としては、ウォルド F. (Wold, F.)、「タンパク質の翻訳後共有結合修飾 (Posttranslational Covalent Modification of Proteins)」、B.C. ジョンソン (B.C. Johnson) 編、Academic Press、New York 1-12 (1983); セイフター (Seifter) ら、(Meth. Enzymol. 182: 626-646 (1990)) およびラタン (Rattan) ら、(Ann. N.Y. Acad. Sci. 663:48-62 (1992)) のような多くの総説を利用することができる。

30

【0089】

したがって、本発明の輸送体ペプチドは、誘導体または類似体をも包括するものであり、ここで、置換されたアミノ酸残基は遺伝子コードによってコードされるものではなく、置換基が含まれ、成熟輸送体ペプチドが、輸送体ペプチドの半減期を増加させる化合物（例えば、ポリエチレングリコール）のような他の化合物と融合するか、または付加アミノ酸が、リーダー配列、分泌配列、成熟輸送体ペプチドの精製配列、またはプロ-タンパク質 (pro-protein) 配列のような成熟輸送体ペプチドと融合する。

40

【0090】

タンパク質/ペプチドの使用

本発明のタンパク質は、図面に示される機能情報に関連した、実質的かつ特異的なアッセイ法において、例えば、抗体を産生させる、または他の免疫反応を誘導するため；生物液中におけるタンパク質（またはその結合対象、またはリガンド）レベルの定量のためのアッセイ法に用いる試薬（標識試薬を含む）として；および対応するタンパク質を選択的に発現する（組織の分化もしくは発達または疾患の状態において、構成的もしくは特定の段階のいずれかで発現する）組織のマーカーとして使用することができる。タンパク質が、

50

別のタンパク質もしくはリガンドと結合するか、または結合する可能性を有する場合、例えば、輸送体-エフェクタータンパク質の相互作用、または輸送体-リガンドの相互作用において、このタンパク質を用いて結合対象/リガンドを特定し、結合相互作用の阻害因子を同定するシステムを開発することができる。これらの一部または全ての使用により、商業製品として製品化するための試薬グレードまたはキット形式へと発展させることが可能となる。

【0091】

上に列記した使用を実施する方法は、当業者に周知である。このような方法を開示している参考文献としては、「分子クローニング：実験マニュアル (Molecular Cloning: A Laboratory Manual)」、第2版、Cold Spring Harbor Laboratory Press、サムブルック J. (Sambrook, J.)、E.F. フリッツ (E. F. Fritsch) および T. マニアティス (T. Maniatis) 編、1989、ならびに「酵素学の方法：分子クローニング技術へのガイド (Methods in Enzymology: Guide to Molecular Cloning Techniques)」、Academic Press、バーガー S.L. (Berger, S. L.) および A.R. キメル (A. R. Kimmel) 編、1987がある。

10

【0092】

本明細書に記載される糖輸送タンパク質と新生児マウスの海馬で発現する糖輸送体との間で実質的な化学的および構造的相同性がある (図1参照)。背景で議論されているように、新生児マウスの海馬で発現する糖輸送体が、糖吸収に関与することは当技術分野において公知である。したがって、本発明によって提供される糖輸送タンパク質、およびこれをコードする遺伝子は、神経精神医学的障害、海馬の糖輸送体に関連する糖吸収不良および他の障害の治療、予防、ならびに/または診断に有用である。

20

【0093】

本発明のペプチドの可能な用途は、主としてタンパク質の供給源ならびにタンパク質のクラス/作用に基づく。例えば、ヒトから単離された輸送体およびそのヒト/哺乳類オルソログは、哺乳類の治療的応用、例えばヒト用薬物、特に輸送体を発現する細胞または組織における生物反応または病的反応の調整に用いられる、作用物質を同定するための標的として役立つ。図1に提供した実験データは、バーチャル・ノーザンブロットにより、本発明の輸送体タンパク質が、卵巣 (腺癌組織)、子宮 (平滑筋肉腫組織)、頸部、腎臓癌組織 (副腎腫)、胚中心B細胞、結腸、および乳児脳において発現されることを示す。さらに、PCRに基づく組織スクリーニングパネルにより、腎臓における発現が示される。輸送体タンパク質、特に糖輸送体サブファミリーのメンバーの活性を調整する医薬品がかなり高い比率で開発されつつある (「背景技術」を参照)。背景技術および図面に提供した構造情報および機能情報は、特に図1に提供した発現情報と共に、本発明の分子の具体的および実質的な用途を提供する。図1に提供した実験データは、卵巣 (腺癌組織)、子宮 (平滑筋肉腫組織)、頸部、腎臓、腎臓癌組織 (副腎腫)、胚中心B細胞、結腸、および乳児脳での発現を示す。このような用途は、当技術分野および日常的な実験において公知の、本明細書に提供した情報を用いて、容易に判断することができる。

30

【0094】

本発明のタンパク質 (本発明の前に開示された可能性のある変異体および断片も含む) は、糖輸送体サブファミリーのメンバーと関連のある輸送体に関連する、生物学的アッセイ法に有用である。このようなアッセイ法は、本発明が属する輸送体のサブファミリーに対して特異的な輸送体関連状態、特に輸送体を発現する細胞および組織における輸送体関連状態の診断および治療に有用な、公知の輸送体機能もしくは活性または特性のいずれかを必要とする。図1に提供した実験データは、バーチャル・ノーザンブロットにより、本発明の輸送体タンパク質が、卵巣 (腺癌組織)、子宮 (平滑筋肉腫組織)、頸部、腎臓癌組織 (副腎腫)、胚中心B細胞、結腸、および乳児脳において発現されることを示す。さらに、PCRに基づく組織スクリーニングパネルにより、腎臓における発現が示される。本発明のタンパク質は、細胞系または無細胞系における薬物スクリーニングアッセイにも有用である (Hodgson, Bio/technology, 1992, Sept 10 (9); 973~80)。細胞系は自然状態のもの、すなわち、生検材料または細胞培養下で増殖した、輸送体を通常発現する細胞で

40

50

よい。図1に提供した実験データは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳における発現を示す。1つの代替的な態様では、細胞系アッセイ法は、輸送体タンパク質を発現する組換え宿主細胞を伴う。

【0095】

ポリペプチドは、天然の状態、または輸送体に関連する特定の疾患もしくは症状を引き起こす改変された形態における、タンパク質の輸送体活性を調節する化合物を同定するために用いることができる。本発明の輸送体、ならびに適当な変異体および断片はいずれも、この輸送体に対して結合能力を持つ候補化合物をアッセイするためのハイスループット・スクリーニングにおいて使用することができる。これらの化合物は、さらに、これらの輸送体活性に対する化合物の作用を判定するために、機能性の輸送体に対してスクリーニングを行うことができる。さらにこれらの化合物は、動物または無脊椎動物系において、活性/効果を判定するために試験することができる。化合物は、輸送体を望ましい程度まで活性化（アゴニスト）または不活化（アンタゴニスト）するかどうかと同定される。

10

【0096】

さらに、本発明のタンパク質は、輸送体タンパク質と、該輸送体タンパク質と通常相互作用する分子（例えば、輸送体タンパク質が通常相互作用するシグナル経路の基質または構成要素（例えば別の輸送体））との間での相互作用を刺激または阻害する能力について、化合物をスクリーニングするために用いることができる。このようなアッセイ法には、一般的に、輸送体タンパク質もしくは断片が標的分子と相互作用し、かつタンパク質と標的との複合物形成を検出することが可能な条件、または膜電位変化、タンパク質リン酸化、cAMPの代謝回転、およびアデニル酸シクラーゼの活性化などの、シグナル伝達の関連作用のいずれかのような、輸送体タンパク質と標的との相互作用の生化学的結果を検出することが可能な条件で、輸送体タンパク質と候補化合物が結合される工程が含まれる。

20

【0097】

候補化合物としては、例えば、1)最終部がIgの融合ペプチド、およびランダムペプチドライブラリーのメンバーを含む可溶性ペプチド（例えば、Lamら、Nature 354:82-84 (1991) ; Houghtenら、Nature 354:84-86 (1991)参照）、ならびにD型および/またはL型アミノ酸から構成されるコンビナトリアルケミストリーに由来の分子ライブラリーのメンバーを含むペプチド；2)ホスホペプチド（例えば、ランダムおよび部分的に変更されたホスホペプチドライブラリーのメンバー；例えば、Songyangら、Cell 72:767-778 (1993)参照）；3)抗体（例えば、ポリクローナル抗体、モノクローナル抗体、ヒト化抗体、抗イディオタイプ抗体、キメラ抗体、および単鎖抗体、ならびにFab、F(ab')₂、Fab発現ライブラリー断片、および抗体のエピトープ結合断片）；ならびに4)小型の有機分子および無機分子（例えば、コンビナトリアルおよび天然生成物ライブラリーから得られる分子）が含まれる。

30

【0098】

ある候補化合物は、リガンド結合において競合する受容体の可溶性断片である。他の候補化合物には、変異輸送体、または輸送体機能に影響を及ぼす変異を含む適切な断片が含まれ、このため、リガンドと競合する。したがって、例えば高い親和性を有するか、または断片がリガンドと結合するが解離しないような、リガンドと競合する断片が本発明に包含される。

40

【0099】

本発明はさらに、輸送体活性を調節（刺激または阻害）する化合物を同定するための、他のエンドポイントアッセイ法を含む。このアッセイ法は、一般的に、輸送体活性を示すシグナル伝達経路における事象のアッセイを伴う。このため、輸送体タンパク質依存性シグナルカスケードに対する応答を促進または抑制するよう調節される、リガンドの輸送、細胞膜電位の変化、タンパク質の活性化、遺伝子発現の変化についてのアッセイが行われる。

【0100】

輸送体により媒介される生物学的または生化学的な機能は、いずれもエンドポイントアッ

50

セイ法として使用されうる。これらは、本明細書に記載されている全ての生化学的または生化学的/生物学的な事象を含み、本明細書に引用される文献には、これらのエンドポイントアッセイ法の標的が参照として本明細書に組み入れられ、また、これらは、当業者に公知であるか、または図面、特に図2の情報を用いて、容易に同定することができる他の機能を含む。特に、輸送体を発現する細胞または組織の生物学的機能についてアッセイを行うことができる。図1の実験データは、バーチャル・ノーザンプロットにより、本発明の輸送体タンパク質が、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳において発現されることを示す。加えて、PCRに基づく組織スクリーニングパネルにより、腎臓における発現が示されている。

【0101】

10

結合および/または活性化化合物はまた、キメラ輸送体タンパク質を用いることによりスクリーニングを行うことができ、それはアミノ末端細胞外ドメインまたはその一部、7回膜貫通セグメントまたは細胞内もしくは細胞外ループのような膜貫通ドメイン全体または小領域、およびカルボキシル末端細胞内ドメインまたはその一部において、異種ドメインもしくは小領域に置換され得る。例えば、異なるリガンドと相互作用しさらに未処理の輸送体によって認識される、リガンド結合領域を用いることができる。したがって、異なるセットのシグナル伝達構成要素を活性化のエンドポイントアッセイ法として利用することができる。このような方法により、輸送体が由来する特定の宿主細胞以外でアッセイを行うことが可能となる。

【0102】

20

本発明のタンパク質はまた、輸送体と相互作用する化合物（例えば、結合対象および/またはリガンド）を発見するために設計された方法である、競争結合アッセイ法にも有用である。このために、化合物がポリペプチドと結合または相互作用可能な条件下で、化合物を輸送体ポリペプチドと接触させる。可溶性輸送体ポリペプチドもまた混合物中に加えられる。被験化合物が可溶性輸送体ポリペプチドと相互作用する場合、輸送体標的から形成される複合体の量、または活性は減少する。このタイプのアッセイ法は特に輸送体の特定領域と相互作用する化合物を検索する場合に有用である。したがって、標的の輸送体領域と競合する可溶性ポリペプチドは、対象となる領域に対応したペプチド配列を含むように設計されている。

【0103】

30

無細胞系の薬物スクリーニングアッセイを行うためには、タンパク質の一方または両方の非複合形態からの複合体の分離を促進し、アッセイの自動化に適応させるために、輸送体タンパク質もしくは断片、またはその標的分子のいずれかを固定化することが望ましい場合がある。

【0104】

薬物スクリーニングアッセイ法においては、マトリックスにタンパク質を固定化する技術を使用することができる。ある態様では、融合タンパク質にはタンパク質をマトリックスに結合することのできるドメインを付加することができる。例えばグルタチオン-S-トランスフェラーゼ融合タンパク質を、グルタチオンセファロースビーズ（Sigma Chemical、St. Louis、MO）またはグルタチオン誘導マイクロタイタープレート上に吸着することができ、次いで細胞溶解物（例えば、³⁵S標識）と候補化合物とが結合され、複合体形成誘導条件（例えば、塩およびpHの生理学的条件）の下で混合物がインキュベーションされる。インキュベーションの後、非結合標識の除去のためにビーズを洗浄し、マトリックスを固定化して、放射性標識を直接、または複合体を分離した後の上澄みを測定する。あるいは、複合体はSDS-PAGEによりマトリックスから分離することができ、標準の電気泳動技術を用いることによって、ゲルからビーズ画分中の輸送体結合タンパク質のレベルを定量することができる。例えば、ポリペプチドまたはその標的分子のいずれかは、当技術分野に周知の技術を利用して、ピオチンおよびストレプトアビジンの結合を用いて固定化される。あるいは、タンパク質と反応し、タンパク質と標的分子との結合を妨げない抗体は、プレートのウェルに誘導化され、このタンパク質は抗体との結合によりそのウェルの中に捕

50

らえられる。輸送体結合タンパク質および候補化合物の調製物は、輸送体タンパク質の存在するウェル中で培養され、ウェルに捕らえられた複合体の量を定量することができる。このような複合体を検出する方法としては、GST固定複合体による前述の方法に加えて、輸送体タンパク質標的分子に反応性のある抗体、または、輸送体タンパク質に反応性があり標的分子と競合する抗体を用いた複合体の免疫検出法、および標的分子に関連する酵素活性の検出に基づく酵素結合アッセイ法が含まれる。

【0105】

本発明の輸送体のうちの1つを調節する物質は、上述のアッセイ法の1つまたは複数を単独または組み合わせて用いることにより同定することができる。一般的には、最初に細胞系または無細胞系を用い、次に動物または他のモデル系における活性を確認することが好ましい。このようなモデル系は、当技術分野に周知であり、本記載において容易に用いることができる。

10

【0106】

これらの薬物スクリーニングアッセイ法によって同定される輸送体タンパク質活性のモジュレータは、輸送体を発現する細胞または組織に処理することによって、輸送体経路により媒介される疾患に罹患する患者の治療に用いることができる。図1の実験データは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳における発現を示す。これらの治療方法には、薬学的組成物中の輸送体活性のモジュレータを患者の治療に必要な量投与する工程が含まれており、このモジュレータは本明細書に記載のようにして同定される。

20

【0107】

本発明の他の局面では、輸送体と結合または相互作用し、輸送体活性に関連している他のタンパク質を同定するために、ツーハイブリッドアッセイ法またはスリーハイブリッドアッセイ法（米国特許第5,283,317号；Zervosら、(1993) Cell 72:223-232；Maduraら、(1993) J. Biol. Chem. 268:12046-12054；Bartelら、(1993) Biotechniques 14:920-924；Iwabuchiら、(1993) Oncogene 8:1693-1696；およびBrent国際公開公報第94/10300号参照）において、輸送体タンパク質を「ベイト（bait）タンパク質」として使用することができる。このような輸送体結合タンパク質は、例えば、輸送体媒介シグナル伝達経路の下流要素としての、輸送体タンパク質または輸送体標的によるシグナル伝達に関与している可能性がある。あるいは、このような輸送体結合タンパク質は、輸送体阻害因子である可能性も考えられる。

30

【0108】

ツーハイブリッドシステムは、分離可能なDNA結合ドメインおよび活性化ドメインからなる大部分の転写因子のモジュラー性に基づいている。簡単に言うと、このアッセイ法では2つの異なるDNA構造を利用する。一方の構造においては、輸送体タンパク質をコードする遺伝子は、既知の転写因子（例えばGAL4）のDNA結合ドメインをコードする遺伝子に融合される。他方の構造においては、DNA配列ライブラリーから得られ、未知のタンパク質（「プレイ（pray）」または「サンプル（sample）」）をコードするDNA配列が既知の転写因子の活性化ドメインをコードする遺伝子に融合される。「ベイトタンパク質」および「プレイタンパク質」がインビボで相互作用することができ、輸送体依存性の複合体を形成する場合、転写因子のDNA結合ドメインおよび活性化ドメインは近接する。この近接により、転写因子に反応する転写調節部位に機能的に結合するレポーター遺伝子（例えばLacZ）の転写を行うことができる。レポーター遺伝子の発現を検出することが可能であり、機能的転写調節因子を含む細胞コロニーを単離および使用して、輸送体タンパク質と相互作用するタンパク質をコードするクローン遺伝子を得ることができる。

40

【0109】

本発明はさらに、前述のスクリーニングアッセイ法によって同定される新規の物質にも関係する。したがって、本明細書に記載されるようにして同定された物質を適当な動物のモデルに使用することも本発明の範囲内である。例えば、本明細書に記載のように同定された物質（例えば輸送体調節物質、アンチセンス輸送体核酸分子、輸送体特異的抗体、また

50

は輸送体結合対象)を、これらの物質による治療の有効性、毒性、または副作用を判定するために、動物、または他のモデルで用いることができる。あるいは、本明細書に記載のように同定された物質を、このような物質の作用機構を決定するために、動物または他のモデルで用いることができる。さらに、本発明は、本明細書に記載のように治療のための前記スクリーニングアッセイ法により同定された新規の薬物の使用に関する。

【0110】

本発明の輸送体タンパク質は、ペプチドにより媒介される疾患または素因の診断のための標的を提供するのに有用である。したがって、本発明は、細胞、組織、もしくは生体中のタンパク質(またはコードするmRNA)の存在、またはそのレベルを検出する方法を提供するものである。図1の実験データは、卵巣(腺癌組織)、子宮(平滑筋肉腫組織)、頸部、腎臓、腎臓癌組織(副腎腫)、胚中心B細胞、結腸、および乳児脳での発現を示す。方法には、輸送体タンパク質との相互作用能を有し、その相互作用が検出可能な化合物と生物試料とを接触させる工程が含まれる。このようなアッセイ法は、単一の検出形態、または抗体チップアレイのようなマルチ検出形態で提供される。

10

【0111】

試料中のタンパク質を検出する1つの物質は、タンパク質に選択的に結合することのできる抗体である。生物試料には、被験者から単離された組織、細胞、および体液、ならびに被験者の内部に存在する組織、細胞、および体液が含まれる。

【0112】

本発明のペプチドはまた、変異ペプチドを持つ患者における、タンパク質の活性、疾患または素因、特に現存するタンパク質ファミリーの他のメンバーで知られる活性および症状の診断に用いるための標的を提供するものである。したがって、ペプチドを生物試料から単離することができ、かつ異常ペプチドを生じる遺伝子突然変異の存在についてアッセイを行うことができる。これは、アミノ酸の置換、欠失、挿入、再配置(異常なスプライシング事象の結果生じる)、および不適当な翻訳後の修飾を含む。分析方法としては、電気泳動移動度の変化、トリプシンペプチド消化の変化、細胞系または無細胞のアッセイ法による輸送体活性の変化、リガンドまたは抗体の結合パターンの変化、等電点の変化、直接アミノ酸配列決定、およびタンパク質の変異の検出に有用な他の公知のアッセイ技術を含む。このようなアッセイ法は、単一の検出形態、または抗体チップアレイのような、マルチ検出形態で提供される。

20

30

【0113】

ペプチドのインビトロ検出技術としては、酵素結合免疫吸着アッセイ法(ELISA)、ウェスタンブロット、抗体、またはタンパク質結合剤のような検出試薬を用いた免疫沈降および免疫蛍光検査法を含む。あるいは、標識された抗ペプチド抗体、または他のタイプの検出物質を被験者に導入することにより、被験者中でペプチドのインビボ検出を行うことができる。例えば、抗体は放射性マーカーにより標識することができ、被験者中のこのマーカーの存在および位置は、標準画像化技術によって検出することができる。被験者において発現されたペプチドの対立遺伝子変異体を検出する方法、および試料中のペプチド断片を検出する方法は、特に有用である。

【0114】

ペプチドはまた、薬理遺伝学的分析においても有用である。薬理遺伝学では、薬物の変化の傾向と、影響を受けたヒトの異常作用に従って、薬物に対する応答における臨床的に著しい遺伝的変異について取り扱う。例えば、アイヒェルバウム M.(Eichelbaum, M.) (Clin. Exp. Pharmacol. Physiol. 23(10-11):983-985 (1996))、およびリンダー M.W.(Linder, M.W.) (Clin. Chem. 43(2):254-266 (1997))参照。これらの変異の臨床的な結果は、個体の代謝変異の結果として、ある個体に対しては治療薬物が重い毒性をもたらし、またはある個体に対しては治療の失敗に終わる。このように、個体の遺伝子型は、体内で治療化合物を作用させる方法、または体が化合物を代謝する方法を決定することができる。さらに、酵素を代謝させる薬物の活性は、薬物作用の強度と期間の両方に影響する。このように、個体の薬理遺伝学は、個体の遺伝子型に基づいた予防、または治療的な処置にお

40

50

いて、効果的な化合物、およびこのような化合物の効果的な投与量の選択を可能とする。酵素代謝性の薬物における、遺伝子多形性の発見により、ある患者は期待される薬効を得られない、過度の薬効を示す、または標準の投薬量から重大な毒性を被るといったことの原因を説明することができる。多形性は、代謝能の高い個体 (extensive metabolizer) の表現型と代謝能の低い個体 (poor metabolizer) の表現型で表されることができる。したがって、遺伝子の多形性は、ある集団の輸送体機能の1つまたは複数が他の集団のそれと異なるような、輸送体タンパク質の対立遺伝子タンパク質変異に至るかもしれない。このように、ペプチドは治療法に影響しうる遺伝子の素因を確認するための標的となり得る。このため、リガンドベースの治療において、多形性により、リガンド結合活性および輸送体活性化活性がより高いまたはより低い、アミノ末端細胞外ドメインおよび/または他のリガンド結合領域が生じうる。したがって、多形性を含む所与の集団においては、治療効果を最大にするように、リガンド投与量は必然的に修正されることが考えられる。遺伝子型同定に代わるものとしては、特定の多形性のペプチドを同定することができる。

10

【0115】

ペプチドはまた、タンパク質の発現がない、タンパク質の発現が不適当である、またはタンパク質の発現が望ましくないことによって特徴づけられる障害を治療するために有用である。図1の実験データは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳での発現を示す。したがって、治療方法には、輸送体タンパク質または断片の使用が含まれる。

20

【0116】

抗体

本発明はまた、本発明のペプチド、このようなペプチドを含むタンパク質、それらの変異体およびその断片の1つに選択的に結合する抗体を提供するものである。本明細書で用いられているように、抗体が標的ペプチドと結合し、無関係なタンパク質と強く結合しないような場合、抗体は標的ペプチドと選択的に結合している。標的ペプチドと実質的に相同性の無い他のタンパク質と結合していても、そのタンパク質が抗体の標的となるペプチドの断片またはドメインと相同性を有している限り、抗体は選択的にペプチドと結合すると考えられる。この場合、ペプチドに結合している抗体は、ある程度の交差反応性を持つにも関わらず、なお選択的であると理解される。

30

【0117】

本明細書で用いられるように、抗体は当技術分野で認められているものと同じ用語で定義され、これらは、抗原の投与に应答して哺乳類生物により生成されるマルチサブユニットタンパク質である。本発明の抗体は、ポリクローナル抗体、モノクローナル抗体、およびこれらの抗体の断片を含み、FabまたはF(ab')₂、およびFv断片を含むが、これに限定されるものではない。

【0118】

所与の標的ペプチドに対する抗体の生成および/または同定について、多くの方法が知られている。このような方法のいくつかは、ハーロウ (Harlow)、「抗体 (Antibodies)」、Cold Spring Harbor Press、(1989) に記載されている。

40

【0119】

一般に、抗体を生成するためには、単離ペプチドを免疫原として用い、例えばラット、ウサギ、またはマウスのような哺乳類生物に投与する。全長タンパク質、抗原性ペプチド断片または融合タンパク質を用いることができる。特に重要な断片は、図2において同定されるドメインのような、機能ドメインを含むものであり、タンパク質アライメント方法を使用して容易に同定することができ、図面に示されているようなファミリーと配列相同性または相違性を持つドメインである。

【0120】

抗体は、好ましくは輸送体タンパク質の領域、または単離された断片から調製される。抗体は、本明細書に記載されるように、ペプチドのいかなる領域からでも調製することができる。しかしながら、好ましい領域には、機能/活性、および/または輸送体/結合対象の

50

相互作用に関係している領域が含まれると考えられる。図2は特に重要な領域を同定するのに用いることができ、この時、配列アライメントは保存された特有の配列断片を同定するのに用いることができる。

【0121】

抗原性断片は、一般的に、少なくとも8個の連続するアミノ酸残基を含むと考えられる。抗原性ペプチドは、少なくとも10個、12個、14個、16個またはそれ以上のアミノ酸残基を含むことができる。このような断片は、例えば、タンパク質の表面上に位置する領域、例えば、親水性の領域に対応する断片のような物理的な性質、または配列の特異性（図2参照）に基づいて選択することができる。

【0122】

本発明の抗体の検出は、検出可能な物質と抗体とのカップリング（すなわち、物理的な結合）によって容易に行うことができる。検出可能な物質の例としては、種々の酵素、補欠分子族、蛍光性物質、発光性物質、生物発光性物質、および放射性物質が含まれる。好適な酵素の例としては、西洋ワサビペルオキシダーゼ、アルカリホスファターゼ、ガラクトシダーゼ、またはアセチルコリンエステラーゼを含み、好適な補欠分子族複合体の例としては、ストレプトアビジン/ビオチン、およびアビジン/ビオチンを含み、好適な蛍光性物質の例としては、ウンベリフェロン、フルオレセイン、フルオレセイン・イソチオシアネート、ローダミン、ジクロロトリアジニルアミンフルオレセイン、ダンシルクロライド、またはフィコエリトリンを含み、発光性物質の例としては、ルミノールを含み、生物発光性物質の例としては、ルシフェラーゼ、ルシフェリン、およびエクオリンを含み、なら

10

20

【0123】

抗体の使用

抗体は、本発明のタンパク質の1つを、アフィニティークロマトグラフィ、または免疫沈降のような標準の技術によって単離するために用いることができる。抗体は、細胞からの天然型タンパク質、および宿主細胞で発現される組換えによって産生されたタンパク質の精製を容易にすることができる。さらに、このような抗体は、生体内の様々な組織または通常の発達工程におけるタンパク質の発現パターンを決定するため、細胞または組織内における本発明のタンパク質の存在の検出に有用である。図1の実験データは、バーチャル・ノーザンブロットにより、本発明の輸送体タンパク質が、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳において発現されることを示す。加えて、PCRに基づく組織スクリーニングパネルにより、腎臓における発現が示されている。さらに、このような抗体は、発現の量およびパターンを評価するための、インサイチュー、インビトロ、細胞溶解物中、および上澄み中でのタンパク質の検出に用いることができる。また、このような抗体は、生物学的状態の発達または進行の間、異常な組織分布または異常な発現を評価するのに用いることができる。全長タンパク質の循環断片における抗体検出は、代謝回転を同定するのに用いることができる。

30

40

【0124】

さらに、抗体は、タンパク質に関連した疾患の活発な段階、または該疾患素因を持つ個体などの、疾患状態における発現を評価するのに用いることができる。障害が不適当な組織分布、発生における発現、タンパク質の発現レベル、または発現/進行状態に起因する場合、抗体は通常のタンパク質に対して調製される。図1の実験データは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳での発現を示す。障害がタンパク質の特定の変異により特徴づけられる場合、この変異タンパク質に特異的な抗体を、特定の変異タンパク質の存在についてアッセイするために用いることができる。

【0125】

抗体はまた、生体内の各種組織における、細胞の正常または異常な細胞内局在を評価するのに用いることができる。図1の実験データは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳での発現

50

を示す。診断としての使用は、遺伝子の試験だけでなく、治療法をモニターすることにも適用することができる。したがって、治療が最終的に、発現レベル、または異常配列および異常組織分布の存在、または発生における発現を修正することを目指すものである場合、タンパク質または関連する断片に対して誘導された抗体を、治療の有効性をモニターするのに用いることができる。

【0126】

さらに、抗体は薬理遺伝学的分析に有用である。このように、多形性のタンパク質に対して調製される抗体は、治療法の修正を必要とする個体を特定するために用いることができる。抗体は、また、電気泳動移動度、等電点、トリプシンペプチド消化、および当業者に周知の他の物理的なアッセイ法によって分析される異常タンパク質の免疫学的なマーカーのような診断上の手段としても有用である。

10

【0127】

抗体はまた、組織型の分類にも有用である。図1の実験データにより、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳での発現が示されている。このように、特定のタンパク質が特定の組織中の発現と関連していた場合、このタンパク質に特異的である抗体を、組織型を同定するために用いることができる。

【0128】

抗体はまた、タンパク質機能を阻害するのに有用であり、例えば、リガンドまたはタンパク質結合対象のような結合対象への輸送体ペプチドの結合を妨害する。これらの使用はタンパク質の機能阻害に関連する治療状況において適用されることができる。抗体は、例えば、結合を妨害し、ペプチド活性を調節（アゴナイズまたはアンタゴナイズ）することに用いることができる。抗体は、機能のために必要な部位を含む特定の断片に対して、または細胞もしくは細胞膜と関係している完全タンパク質に対して調製される。図2に、本発明のタンパク質に関する構造情報を示す。

20

【0129】

本発明はまた、生物試料中のタンパク質の存在を検出するために抗体を用いたキットを包含する。キットには、標識された抗体または標識可能な抗体、および生物試料中でタンパク質を検出するための化合物または試薬；試料中のタンパク質量を決定する手段；試料中のタンパク質量と標準の量とを比較する手段；ならびに使用のための説明を含む。このようなキットは、単一のタンパク質もしくはエピトープを検出するために提供されるか、または抗体検出アレイのように、多数のエピトープのうちの1つを検出するように設定されることができる。アレイとしては、核酸アレイが詳細に後述され、抗体アレイのための同様の方法も開発されている。

30

【0130】

核酸分子

本発明は、さらに本発明の輸送体ペプチドまたはタンパク質をコードする単離核酸分子（cDNA、転写、およびゲノム配列）を提供するものである。このような核酸分子は、本発明の輸送体ペプチドの1つをコードするヌクレオチド配列、これらの対立遺伝子変異体、またはこれらのオルソログもしくはパラログからなる、本質的になる、または含むと考えられる。

40

【0131】

本明細書に用いられているように、「単離」核酸分子は、核酸の天然起源に存在する他の核酸から分離されたものである。好ましくは、「単離」核酸はその核酸の由来となる生物のゲノムDNAにおいて、核酸に天然に隣接する配列（すなわち、核酸の5'末端および3'末端に位置する配列）は含まない。しかしながら、例えば、約5KB、4KB、3KB、2KBまたは1KB未満まで、特に連続するペプチドをコードする配列、および同一遺伝子内であるが、ゲノム配列中のイントロンにより分離されているペプチドをコードする配列のような、いくつかの隣接ヌクレオチド配列がある。重要な点は、核酸が、本明細書に記載されるような特定の操作、例えば、組換え発現、プローブやプライマーの調製、および核酸配列のため

50

の他の特定の使用等に取り扱うことができるように、離れた重要でない隣接配列から分離されているということである。

【0132】

さらに、例えば、転写/cDNA分子のような「単離」核酸分子は、他の細胞物質、組換え技術により製造される場合には培地、または化学的に合成される場合には化学前駆体もしくは他の化学物質を、実質的に含まない。しかしながら、この核酸分子は、他のコード配列または他の調節配列に融合されることができるが、これは単離されたものとして考えられる。

【0133】

例えば、ベクターに含まれる組換えDNA分子は、単離されたものとして考えられる。さらなる単離DNA分子の例には、非相同性の宿主細胞中に保持された組換えDNA分子、または溶液中の精製（部分的または実質的に）されたDNA分子が含まれる。単離されたRNA分子は、本発明の単離DNA分子の、インピボまたはインピトロでのRNA転写産物を含む。本発明による単離核酸分子としては、合成的に製造された分子をさらに含む。

10

【0134】

したがって、本発明は、図1もしくは図3（配列番号：1および配列番号：2、転写配列、ならびに配列番号：5、ゲノム配列）に記載のヌクレオチド配列からなる核酸分子、または図2（配列番号：3および配列番号：4）に記載のタンパク質をコードする任意の核酸分子を提供するものである。ヌクレオチド配列がこの核酸分子の完全なヌクレオチド配列であるとき、核酸分子はヌクレオチド配列からなる。

20

【0135】

本発明はさらに、図1もしくは図3（配列番号：1および配列番号：2、転写配列、ならびに配列番号：5、ゲノム配列）に記載のヌクレオチド配列から実質的になる核酸分子、または図2（配列番号：3および配列番号：4）に記載のタンパク質をコードする任意の核酸分子を提供するものである。最終的な核酸分子において、このようなヌクレオチド配列がごくわずかの付加核酸残基とともに存在するとき、核酸分子はヌクレオチド配列から実質的になる。

【0136】

本発明はさらに、図1または図3（配列番号：1および配列番号：2、転写配列、ならびに配列番号：5、ゲノム配列）に記載のヌクレオチド配列を含む核酸分子、または図2（配列番号：3および配列番号：4）に記載のタンパク質をコードする任意の核酸分子を提供するものである。ヌクレオチド配列が核酸分子の最終的なヌクレオチド配列の少なくとも一部である場合、核酸分子はヌクレオチド配列を含む。これによると、核酸分子は、そのヌクレオチド配列だけであるか、または付加的な核酸残基、例えば、それに天然に関連する核酸残基、または非相同性のヌクレオチド配列を有することもできる。このような核酸分子は、ごくわずかの付加的なヌクレオチドを有するか、または数百もしくはそれ以上の付加的なヌクレオチドを含むこともできる。これらの種々のタイプの核酸分子を容易に生成/単離する方法について、以下に簡単に述べる。

30

【0137】

図1および図3に、コード配列および非コードの配列の両者が示される。本発明の起源である、ヒトゲノム配列（図3）、およびcDNA/転写配列（図1）のため、図面中の核酸分子は、ゲノムイントロン配列、5'と3'の非コード配列、遺伝子調節領域、および非コード遺伝子間配列を含むと考えられる。一般に、このような配列の特徴は、図1および図3の両方において記載されているか、または当技術分野において公知の計算手段を用いて容易に同定することができる。以下で議論されるように、いくつかの非コード領域、特にプロモーターのような遺伝子調節要素は、例えば、非相同性の遺伝子発現の制御、遺伝子活性を調節する化合物同定のための標的等の種々の目的にとって有用であり、また特に、本明細書で提供されるゲノム配列の断片として主張されている。

40

【0138】

単離核酸分子は、成熟したタンパク質と付加的アミノ末端もしくはカルボキシル末端アミ

50

ノ酸、または成熟ペプチド内のアミノ酸（例えば、成熟形態が1つより多くのペプチド鎖を有する場合）をコードすることができる。このような配列は、前駆体から成熟した形態へのタンパク質のプロセッシングにおいて、タンパク質搬送の促進、タンパク質半減期の延長もしくは短縮、またはタンパク質のアッセイもしくは製造の際の操作の効率化、または他の事象における役割を果たし得る。一般に、インサイチューの場合、付加アミノ酸は細胞酵素によって成熟したタンパク質へとプロセッシングされてもよい。

【0139】

上述したように、単離核酸分子は、輸送体ペプチドのみをコードする配列、成熟したペプチドをコードする配列、およびリーダー配列または分泌配列（例えば、プレ-プロ（pre-pro）、プロ-タンパク質配列）のような付加的なコード配列を含むが、これに限定されるものではなく、付加的なコード配列および付加的な非コード配列、例えば、イントロンと非コード5'配列および3'配列のような、転写されるが翻訳はされない、転写、mRNAプロセッシング（スプライシングおよびポリアデニル化シグナルを含む）、リボソームの結合、およびmRNAの安定性の役割を果たすものを含んでも含まなくても良い。加えて、核酸分子は、例えば、精製を容易にするペプチドをコードするマーカー配列と融合されることもできる。

10

【0140】

単離核酸分子は、mRNAのようなRNAの形態、またはクローニングによって得られるかもしくは化学合成技術もしくはその組み合わせによって生成されるcDNAおよびゲノムDNAを含む、DNAの形態をとり得る。核酸、特にDNAは、二本鎖、または一本鎖であり得る。一本鎖の核酸は、コード鎖（センス鎖）、または非コード鎖（アンチセンス鎖）であり得る。

20

【0141】

本発明はさらに、本発明のペプチドの断片をコードする核酸分子と同様に、上記したような本発明の輸送体タンパク質の明らかな変異体をコードする核酸分子を提供するものである。このような核酸分子は、対立遺伝子変異体（同一遺伝子座）、パラログ（異なる遺伝子座）、およびオルソログ（異なる生物）のように天然に発生するか、または組換えDNA法もしくは化学合成によって生成され得る。このような非天然に発生する変異体は、核酸分子、細胞または生物に適用される技術を含む突然変異誘発技術によって生成され得る。したがって、上述したように、変異体にはヌクレオチドの置換、欠失、反転、および挿入が含まれうる。変異は、コード領域および非コード領域のいずれか、または両方で起こりうる。変異は、保存的アミノ酸置換および非保存的アミノ酸置換の両方を生じることができる。

30

【0142】

本発明はさらに、図1および図3に示される核酸分子の非コードの断片を提供するものである。好ましい非コードの断片としては、プロモーター配列、エンハンサー配列、遺伝子調節配列、および遺伝子終結配列が含まれるが、これに限定されるものではない。このような断片は、非相同性の遺伝子発現の制御、および遺伝子調節物質の同定を行うためのスクリーニングの開発において有用である。プロモーターは、図3のゲノム配列における5'からATG開始部位において容易に同定される。

【0143】

断片は、12個またはそれ以上のヌクレオチドの連続するヌクレオチド配列を含む。さらに、断片は少なくとも30個、40個、50個、100個、250個、または500個のヌクレオチド長であり得る。断片の長さは使用目的に基づく。例えば断片は、ペプチドのエピトープ関連領域をコードすることができるか、またはDNAプローブおよびDNAプライマーとして有用である。このような断片は、オリゴヌクレオチドプローブを合成するための既知のヌクレオチド配列を用いて単離することができる。標識されたプローブは、コード領域と対応する核酸を単離するため、cDNAライブラリー、ゲノムDNAライブラリー、またはmRNAのスクリーニングに用いることができる。さらに、プライマーは、遺伝子の特定領域をクローニングするためのPCR反応に用いることができる。

40

【0144】

50

プローブ/プライマーは一般的に、実質的に精製されたオリゴヌクレオチドまたはオリゴヌクレオチド対を含む。オリゴヌクレオチドは、一般に、少なくとも約12個、20個、25個、40個、50個またはそれ以上の連続するヌクレオチドに、ストリンジェントな条件下でハイブリダイズされたヌクレオチド配列領域を含む。

【0145】

オルソログ、ホモログ、および対立遺伝子変異体は、当技術分野において周知の方法を用いて同定することができる。ペプチドの項で述べたように、これらの変異体は、ペプチドをコードするヌクレオチド配列を含み、図面に示されるヌクレオチド配列、またはこの配列の断片に対して、典型的には、60%～70%、70%～80%、80%～90%、より典型的には、少なくとも約90%～95%またはそれ以上の相同性を有するものである。このような核酸分子は、穏やかな条件からストリンジェントな条件の下で、図面に示されるヌクレオチド配列またはこの配列の断片に対してハイブリダイズが可能なものとして、容易に同定することができる。対立遺伝子変異体は、コードする遺伝子の遺伝子座で容易に決定されることができる。図3に提示したデータが示すように、マップ位置は第1染色体上にあることが決定された。

10

【0146】

図3に、本発明の輸送体タンパク質をコードする遺伝子内において見出された、SNPについての情報が示される。イントロン内ならびにORFの5'および3'領域内において、42個の異なるヌクレオチド位置でSNPが同定された。イントロン内およびORFの外側のそのようなSNPは、制御/調節要素に影響を及ぼしうる。

20

エキソンにおける2つのSNPのうち、1つはアミノ酸配列の変化をもたらす（すなわち、非同義SNP）。これらのSNPがもたらすアミノ酸配列の変化は図3に示され、普遍的遺伝暗号および参照として図2に提供されるタンパク質配列を用いて容易に決定することができる。

【0147】

本明細書に用いられるように、「ストリンジェントな条件下でハイブリダイズする」という用語は、ペプチドをコードするヌクレオチド配列が、互いに少なくとも60%～70%の相同性を有し、互いにハイブリダイズしたままである程度にハイブリダイズおよび洗浄が行われる条件を意味している。この条件は、互いに少なくとも約60%、少なくとも約70%、少なくとも約80%、またはそれ以上の配列相同性を有するような配列が、典型的には互いにハイブリダイズしたままであるような条件でありうる。このようなストリンジェントな条件は、当業者に周知であり、「分子生物学の最新プロトコル（Current Protocols in Molecular Biology）」、John Wiley & Sons、N.Y. (1989)、6.3.1-6.3.6. に記載されている。ストリンジェントなハイブリダイズ条件の1つの例では、6×塩化ナトリウム/クエン酸ナトリウム（SSC）中、約45℃でハイブリダイズし、その後、0.2×SSC、0.1% SDS中、50℃～65℃で1回または複数回洗浄する。穏やかな、低ストリンジェントのハイブリダイズ条件の例は、当業者において周知である。

30

【0148】

核酸分子の使用

本発明の核酸分子は、プローブ、プライマー、化学合成中間体、および生物学的アッセイ法において有用である。核酸分子は、図2に示されているペプチドをコードする全長cDNAおよびゲノムクローンを単離するため、ならびに図2に示すペプチドと同一または関連したペプチドを生成する変異体（対立遺伝子、オルソログ等）に対応するcDNAおよびゲノムクローンを単離するために、メッセンジャーRNA、転写/cDNA、およびゲノムDNAのハイブリダイゼーションプローブとして有用である。図3に説明されているように、挿入/欠失変異体（「インデル」）を含むSNPが、42個の異なるヌクレオチド位置で同定された。

40

【0149】

プローブは、図面に示されている核酸分子の全長において、いかなる配列とも対応することができる。したがって、それは5'非コード領域、コード領域、および3'非コード領域から誘導することができる。しかしながら、すでに述べたように、断片は、本発明以前に開

50

示された断片を含むものとして見なされることはない。

【0150】

核酸分子はまた、核酸分子のいずれかの所与の領域を増幅するPCRのプライマーとしても有用であり、所望の長さおよび配列のアンチセンス分子の合成においても有用である。

【0151】

核酸分子はまた、組換えベクターの構築にも有用である。このようなベクターには、ペプチド配列の一部または全部を発現する発現ベクターが含まれる。ベクターはまた、挿入ベクターも含み、これは例えば細胞ゲノム中のような他の核酸分子中に組み込まれ、遺伝子および/または遺伝子産物のインサイチュウ発現を変化させるために用いられる。例えば、内因性コード配列では、1つまたは複数の特異的に導入された変異を含むコード領域の全部または一部との相同組換えを経て置換され得る。

10

【0152】

核酸分子はまた、タンパク質の抗原部分を発現するためにも有用である。

【0153】

核酸分子はまた、インサイチュウハイブリダイゼーション法により、核酸分子の染色体位置を決定するためのプローブとしても有用である。図3に提示したデータが示すように、マップ位置は第1染色体上にあることが決定された。

【0154】

核酸分子はまた、本発明の核酸分子の遺伝子調節領域を含むベクターの製造にも有用である。

20

【0155】

核酸分子はまた、本明細書に記載される核酸分子から生成されるmRNAの全部または一部と対応しているリボザイムの設計にも有用である。

【0156】

核酸分子はまた、ペプチドの一部または全部を発現するベクターの製造にも有用である。

【0157】

核酸分子はまた、核酸分子およびペプチドの一部または全部を発現する宿主細胞の構築にも有用である。

【0158】

核酸分子はまた、核酸分子およびペプチドの一部または全部を発現するトランスジェニック動物の製造にも有用である。

30

【0159】

核酸分子はまた、核酸発現の存在、レベル、形態、および分布を決定するためのハイブリダイゼーションプローブとしても有用である。図1の実験データは、バーチャル・ノーザンプロットにより、本発明の輸送体タンパク質が、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳において発現されることを示す。

【0160】

したがって、このプローブは、細胞、組織および生物中での特定の核酸分子の存在を検出するか、またはそのレベルを測定するために使用することができる。レベルが測定される核酸は、DNAまたはRNAでありうる。したがって、本明細書で述べられるペプチドに対応するプローブは、所与の細胞、組織および生物における発現、ならびに/または遺伝子コピー数の評価に用いることができる。これらの使用は、正常値と比較して上昇または低下している輸送体タンパク質の発現を含む障害の診断に適當である。

40

【0161】

mRNAを検出するインビトロの技術には、ノーザンハイブリダイゼーションおよびインサイチュウハイブリダイゼーションが含まれる。DNAを検出するインビトロの技術には、サザンハイブリダイゼーションおよびインサイチュウハイブリダイゼーションが含まれる。

【0162】

プローブは、例えばmRNAもしくはゲノムDNAなどの被験者由来の試料細胞中で輸送体をコ

50

ードする核酸のレベルを測定したり、または輸送体遺伝子に変異しているかどうかを確認することにより、輸送体タンパク質を発現する細胞もしくは組織を同定する診断試験キットの一部として使用することができる。図1の実験データは、バーチャル・ノーザンブロットにより、本発明の輸送体タンパク質が、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳において発現されることを示す。加えて、PCRに基づく組織スクリーニングパネルにより、腎臓における発現が示されている。

【0163】

核酸発現アッセイ法は、輸送体の核酸発現を調節する化合物を同定する薬物スクリーニングに有用である。

10

【0164】

したがって、本発明は、輸送体遺伝子の核酸発現に関連した障害、特にそれを発現する細胞および組織において、輸送体が媒介する生物学的過程および病理学的過程に関連した障害の治療に使用可能な化合物を同定する方法を提供する。図1の実験データにより、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳での発現を示す。この方法は、典型的には、輸送体核酸の発現を調節する化合物の能力についてアッセイを行う工程、および望ましくない輸送体核酸発現により特徴づけられる障害を治療するのに用いることができる化合物を同定する工程を含む。このアッセイ法は、細胞系および無細胞系において実施することができる。細胞系のアッセイ法には、天然に輸送体核酸を発現する細胞、または特定の核酸配列を発現するために遺伝子操作された組換え細胞が含まれる。

20

【0165】

輸送体核酸発現のアッセイ法は、例えばmRNAレベルのような核酸レベル、またはシグナル経路に関連する副次化合物の直接的なアッセイ法と関連している。さらに、輸送体タンパク質シグナル経路における応答性を上方調節または下方調節する遺伝子の発現についてもアッセイされる。この態様において、これらの遺伝子調節領域は、ルシフェラーゼのようなレポーター遺伝子に機能的に結合することができる。

【0166】

したがって、輸送体遺伝子発現のモジュレータは、細胞と候補化合物とを接触させ、mRNAの発現を判定する方法により同定されうる。候補化合物の存在下での輸送体mRNAの発現レベルは、候補化合物非存在下での輸送体mRNAの発現レベルと比較される。この比較に基づいて、候補化合物は核酸発現のモジュレータとして同定され、例えば、異常核酸発現により特徴付けられる障害の治療に用いることができる。候補化合物存在下でのmRNAの発現が、非存在下のものと比較して統計的に有意に大きい場合、候補化合物は核酸発現の刺激因子として同定される。候補化合物存在下での核酸発現が、非存在下のものと比較して統計的に有意に小さい場合、候補化合物は核酸発現の阻害因子として同定される。

30

【0167】

本発明はさらに、輸送体を発現する細胞および組織において輸送体核酸発現を調節する遺伝子モジュレータとしての薬物スクリーニングを経て同定された化合物を用い、標的として核酸を用いる治療方法を提供するものである。図1の実験データが、バーチャル・ノーザンブロットにより、本発明の輸送体タンパク質が、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳において発現されることを示す。加えて、PCRに基づく組織スクリーニングパネルにより、腎臓における発現が示されている。調節は、上方調節（即ち、活性化もしくはアゴニゼーション）もしくは下方調節（抑制もしくはアンタゴニゼーション）の両者、または核酸発現を含む。

40

【0168】

あるいは、薬物または小分子がタンパク質を発現する細胞および組織中で輸送体核酸発現を阻害するものである限り、輸送体核酸発現のモジュレータは、本明細書に記載されるスクリーニングアッセイ法を用いて同定される小分子または薬物であり得る。図1の実験デ

50

ータは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳での発現を示す。

【0169】

核酸分子はまた、臨床試験または治療方法において、輸送体遺伝子の発現または活性に対する調節化合物の効果をモニターするのに有用である。したがって、遺伝子発現パターンは、化合物、特に患者の耐性を向上させる化合物を用いた治療における、継続的効果のパロメータとなり得る。遺伝子発現パターンはまた、化合物に対して影響を受けた細胞の生理的反応を示すマーカーとなり得る。したがって、このようなモニタリングにより、化合物の投与量の増加、または患者が耐性を示さない代替化合物の投与を行うことができる。同様に、核酸発現のレベルが望ましいレベルまで低下した場合には、化合物の投与をこれに比例して減少することができる。 10

【0170】

核酸分子はまた、輸送体核酸発現の質的变化、特に疾患に至る質的变化の診断アッセイ法にも有用である。核酸分子は、輸送体遺伝子およびmRNAのような遺伝子発現産物における突然変異の検出に用いることができる。核酸分子は、輸送体遺伝子において天然に発生した遺伝子突然変異を検出し、それによって、その変異を持つ被験者が変異により生じる障害の危険性を有しているかどうかを判定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。突然変異は、遺伝子内の1つもしくは複数のヌクレオチドの欠失、付加、または置換、反転もしくは転位のような染色体の再編成、異常メチル化パターンのようなゲノムDNAの修飾、または増幅のような遺伝子コピー数の変化を含む。機能障害に関連する輸送体遺伝子の変異型の検出は、疾患が輸送体タンパク質の過剰発現、過小発現、または発現の変化の結果生じる場合に、疾患の活性または感受性の診断手段を提供するものである。 20

【0171】

輸送体遺伝子内に突然変異を有する個体は、種々の技術によって核酸レベルにおいて検出されうる。図3に、本発明の輸送体タンパク質をコードする遺伝子内において見出されたSNPについての情報が示される。イントロン内ならびにORFの5'および3'領域内において、42個の異なるヌクレオチド位置でSNPが同定された。イントロン内およびORFの外側のそのようなSNPは、制御/調節要素に影響を及ぼしうる。エキソンにおける2つのSNPのうち、1つはアミノ酸配列の変化をもたらす（すなわち、非同義SNP）。これらのSNPがもたらすアミノ酸配列の変化は図3に示され、普遍的遺伝暗号および参照として図2に提供されるタンパク質配列を用いて容易に決定することができる。図3に提示したデータが示すように、マップ位置は第1染色体上にあることが決定された。ゲノムDNAは直接分析してもよく、または予めPCRを用いて増幅した後で分析してもよい。RNAまたはcDNAも、同様に用いることができる。ある使用においては、突然変異の検出は、アンカーPCRもしくはRACE PCRのような、ポリメラーゼ連鎖反応（PCR）（例えば、米国特許第4,683,195号および同第4,683,202号参照）、または他のものとして、ライゲーション連鎖反応（LCR）（例えば、Landegranら、Science 241:1077-1080 (1988); および Nakazawaら、PNAS 91:360-364 (1994) 参照）において、プローブ/プライマーの使用に関連し、後者は遺伝子内の点突然変異の検出に特に有用である（Abravayaら、Nucleic Acids Res. 23:675-682 (1995)参照）。この方法には、患者から細胞試料を回収する工程；試料細胞から核酸（例えば、ゲノム、mRNA、またはその両方）を単離する工程；遺伝子（存在する場合）のハイブリダイズおよび増幅が起こる条件下で遺伝子に特異的にハイブリダイズする1つまたは複数のプライマーと、核酸試料とを接触させる工程；ならびに増幅産物の存在の有無を検出するか、または増幅産物のサイズを検出し、対照試料の長さと比較する工程を含む。欠失および挿入は、増幅産物のサイズの変化を、正常な遺伝子型のものと比較することにより検出することができる。点突然変異は、増幅DNAと正常なRNAまたはアンチセンスDNA配列とハイブリダイズすることによって同定することができる。 30 40

【0172】

あるいは、輸送体遺伝子の突然変異は、例えば、ゲル電気泳動により決定される制限酵素 50

消化パターンの変化により、直接的に同定することができる。

【0173】

さらに、配列特異的リボザイム（米国特許第5,498,531号）を、リボザイム開裂部位の発生または減少により、特定の変異の存在のスコア化に用いることができる。完全に一致する配列は、ヌクレアーゼ開裂消化アッセイ法、または融解温度の違いによって、不一致の配列から識別することができる。

【0174】

特定位置での配列変化はまた、RNA分解酵素およびS1保護、または化学開裂法のようなヌクレアーゼ保護アッセイ法によって評価することができる。さらに、変異輸送体遺伝子と野生型遺伝子との配列の相違は、直接DNA配列決定によって決定することができる。種々の自動化された配列決定手段は、診断アッセイ法（Naeve, C.W., (1995) *Biotechniques* 19:448）の実施に有用することができ、これらには、質量分析による配列決定（例えば、国際公開公報第94/16101号；Cohenら、*Adv. Chromatogr.* 36:127-162 (1996)；およびGriffinら、*Appl. Biochem. Biotechnol.* 38:147-159 (1993) 参照）も含まれる。

【0175】

遺伝子内の突然変異を検出する他の方法には、RNA/RNAまたはRNA/DNA二本鎖から不一致の塩基を検出するために使用される、開裂試薬から保護する方法（Myersら、*Science* 230:1242 (1985)）；Cottonら、*PNAS* 85:4397 (1988)；Saleebaら、*Meth. Enzymol.* 217:286-295 (1992)）、変異体と野生型の核酸の電気泳動移動度を比較する方法（Oritaら、*PNAS* 86:2766 (1989)；Cottonら、*Mutat. Res.* 285:125-144 (1993)；およびHayashiら、*Genet. Anal. Tech. Appl.* 9:73-79 (1992)）、および変性剤の勾配を含むポリアクリルアミドゲル中での変異体または野生型の断片の動きを、変性勾配ゲル電気泳動を用いてアッセイする方法（Myersら、*Nature* 313:495 (1985)）が含まれる。点突然変異を検出する他の技術の例としては、選択的オリゴヌクレオチドハイブリダイゼーション、選択的増幅、および選択的プライマー伸長が含まれる。

【0176】

核酸分子は、治療方法としての効果を持つにも関わらず、必ずしも疾患を引き起こすわけではない遺伝子型のための個体試験においても有用である。このため、核酸分子は、個体の遺伝子型と、治療に用いられる化合物に対する個体の応答との相関（薬理遺伝学的相関）についての研究に用いることができる。したがって、本明細書に記載される核酸分子は、治療のための適切な化合物または投与計画を選択するために、個体における輸送体遺伝子の変異含量の評価に用いることができる。図3に、本発明の輸送体タンパク質をコードする遺伝子内において見出

イントロン内ならびにORFの5'および3'領域内において、42個の異なるヌクレオチド位置でSNPが同定された。イントロン内およびORFの外側のそのようなSNPは、制御/調節要素に影響を及ぼしうる。エキソンにおける2つのSNPのうち、1つはアミノ酸配列の変化をもたらす（すなわち、非同義SNP）。これらのSNPがもたらすアミノ酸配列の変化は図3に示され、普遍的遺伝暗号および参照として図2に提供されるタンパク質配列を用いて容易に決定することができる。

【0177】

このように、治療に影響する遺伝子変異を示す核酸分子は、個体における目的に適合させた治療に使用可能な診断標的を提供するものである。したがって、これらの多形性を含む組換え細胞および組換え動物の製造は、治療化合物および投与計画についての効果的な臨床設計を可能とする。

【0178】

したがって、核酸分子は、細胞、組織、および生物における輸送体遺伝子発現を制御するためのアンチセンス構築物として有用である。DNAアンチセンス核酸分子は、転写に関連する遺伝子領域に対して相補的になるよう設計され、輸送体タンパク質の転写および産生を阻害する。アンチセンスRNAまたはDNA核酸分子はmRNAとハイブリダイズし、これにより輸送体タンパク質へのmRNAの翻訳が妨害される。

【0179】

あるいは、あるクラスのアンチセンス分子は、輸送体核酸の発現を減少させるためのmRNAの不活性化に用いることができる。したがって、これらの分子は、異常または望ましくない輸送体核酸の発現により特徴づけられる障害の治療に用いることができる。この技術は、mRNAの翻訳能力を減少させるような、mRNAの1つまたは複数の領域に相補的なヌクレオチド配列を含むリボザイム手段による開裂に関連している。可能な領域としては、コード領域、特に、リガンド結合のような輸送体タンパク質の触媒活性および他の機能活性に対応したコード領域を含む。

【0180】

核酸分子はまた、輸送体遺伝子発現において異常な細胞を持つ患者の遺伝子治療のためのベクターを提供するものである。エキソピボで操作され患者に戻される患者の細胞を含む組換え細胞は、個体の体内に導入され、個体細胞内で、個体の治療のために所望の輸送体タンパク質を産生する。

【0181】

本発明は、生物試料中の輸送体核酸の存在を検出するためのキットも包含する。図1の実験は、バーチャル・ノーザンブロットにより、本発明の輸送体タンパク質が、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓癌組織（副腎腫）、胚中心B細胞、結腸、および乳児脳において発現されることを示す。加えて、PCRに基づく組織スクリーニングパネルにより、腎臓における発現が示されている。例えば、キットは、標識された核酸もしくは標識可能な核酸、または生物試料中で輸送体核酸を検出可能な物質を含む試薬；試料中の輸送体核酸量を決定する手段；および試料中の輸送体核酸量と標準の量とを比較する手段を含むことができる。この化合物または物質は適当な容器に封入することができる。このキットは、輸送体タンパク質mRNAまたはDNAの検出キットとして使用するための説明をさらに含むことができる。

【0182】

核酸アレイ

本発明はさらに、核酸検出キットを提供するものであり、これらは、例えば、図1および図3（配列番号：1、2、および5）に示される配列情報に基づいた核酸分子のアレイまたはマイクロアレイである。

【0183】

本明細書に用いられる「アレイ」または「マイクロアレイ」は、紙、ナイロンまたは他のタイプの膜、フィルタ、チップ、ガラススライド、または他の適当な固形支持体のような基板上で合成された異なるポリヌクレオチドまたはオリゴヌクレオチドのアレイを意味する。1つの態様において、マイクロアレイは、米国特許第5,837,832号、チー（Chee）ら、国際公開公報第95/11995号（Cheeら）、ロックハート D.J.（Lockhart, D. J.）ら（1996; Nat. Biotech. 14: 1675-1680）、およびスキーナ M.（Skena, M.）ら（1996; Proc. Natl. Acad. Sci. 93: 10614-10619）に記載される方法にしたがって調製および使用され、これらの全ては参照として本明細書に組み入れられる。他の態様において、このようなアレイは、ブラウン（Brown）ら、米国特許第5,807,522号に記載される方法により製造される。

【0184】

マイクロアレイまたは検出キットは、好ましくは、多数の特有の一本鎖核酸配列により構成され、通常は合成アンチセンスオリゴヌクレオチドか、またはcDNAの断片のいずれかが固体支持体上に固定される。オリゴヌクレオチドは、好ましくは約6個～60個のヌクレオチド長、より好ましくは15個～30個のヌクレオチド長、最も好ましくは約20個～25個のヌクレオチド長である。あるタイプのマイクロアレイまたは検出キットのためには、7個～20個のみのヌクレオチド長であるオリゴヌクレオチドを使うことが好適であり得る。マイクロアレイまたは検出キットは、既知の5'配列または3'配列を含むオリゴヌクレオチド、全長配列を含む連続的なオリゴヌクレオチド、または配列の長さにおいて特定領域から選択された特有のオリゴヌクレオチドを含むものであり得る。マイクロアレイまたは検出キットにおいて用いられるポリヌクレオチドは、遺伝子または対象となる遺伝子に対して特

異的なオリゴヌクレオチドであり得る。

【0185】

マイクロアレイまたは検出キットのための既知の配列のオリゴヌクレオチドを製造するために、対象となる遺伝子（または本発明のコンティグから同定されたORF）は典型的にはコンピュータアルゴリズムを用いて試験され、ヌクレオチド配列の5'末端または3'末端から開始される。典型的なアルゴリズムでは、遺伝子に特有である規定された長さのオリゴマーが同定され、ハイブリダイゼーションに好適な範囲のGC含量を持ち、ハイブリダイゼーションを妨害しうる予測される二次構造を持たない。ある条件では、マイクロアレイまたは検出キットにおいて、オリゴヌクレオチド対を用いることが好適であり得る。オリゴヌクレオチドの「対」は、好ましくは、配列の中央に位置する1つのヌクレオチドを除いて、同一である。第二の対のオリゴヌクレオチド（一方とは不一致）は対照として用いられる。オリゴヌクレオチド対の数は、2から100万の間でありうる。オリゴマーは、光誘導化学プロセスを用いて、基板上の指定領域で合成される。基板は、紙、ナイロンまたは他のタイプの膜、フィルタ、チップ、ガラススライド、または他の適当な固形支持体である。

10

【0186】

他の局面において、オリゴヌクレオチドは、国際公開公報第95/251116号(Baldeschweiler)に記載されるように、化学カップリング手段、およびインクジェットアプリケーション装置を用いて基板の表面上で合成され、これらの全ては参照として本明細書に組み入れられる。他の局面において、ドット（またはスロット）プロットと類似した「グリッド」アレイでは、真空システム、加熱、UV、力学的または化学的結合工程を用いて、cDNA断片、またはオリゴヌクレオチドを基板の表面上に配列し、結合させることができる。上記のようなアレイは、手工または利用可能な装置（スロットプロット、またはドットプロット装置）、材料（任意の適当な固形支持体）、および機械（ロボット装置を含む）を用いて製造され、8、24、96、384、1536、6144もしくはこれ以上、または市販の装置に効果的に使用される2から100万の間の他の数のオリゴヌクレオチドを含んでもいても良い。

20

【0187】

マイクロアレイまたは検出キットを用いて試料の分析を行うために、生物試料から得られたRNAまたはDNAは、ハイブリダイゼーションプローブに調製される。mRNAが単離され、cDNAが調製され、アンチセンスのRNA(aRNA)を調製するためのテンプレートとして用いられる。aRNAを蛍光性ヌクレオチドの存在下で増幅し、標識されたプローブをマイクロアレイまたは検出キットと共にインキュベートし、プローブの配列がマイクロアレイまたは検出キット中の相補的なオリゴヌクレオチドとハイブリダイズする。インキュベーション条件は、正確に相補的に一致しているか、または様々な程度のより低い相補性でハイブリダイゼーションが起こるように調節される。ハイブリダイズしていないプローブを除去した後、蛍光のレベルおよびパターンを判定するためにスキャナが用いられる。スキャンされた画像は、マイクロアレイまたは検出キット上の、相補性の程度および各々のオリゴヌクレオチド配列の相補性の程度および相対的な量を決定するために試験される。生物試料は、任意の体液（例えば血、尿、唾液、痰、胃液など）、培養細胞、生検材料、または他の組織調製物から得られる。検出システムでは、全ての異なる配列において、ハイブリダイゼーションの存在、非存在、および量を、同時に測定するために用いられる。このデータは、試料間での、配列、発現パターン、変異、変異体、または多形性といった、大規模な相関性の研究に用いられる。

30

40

【0188】

本発明は、このようなアレイを用いて、本発明の輸送体タンパク質/ペプチドの発現を同定するための方法を提供するものである。詳細には、このような方法は、被験試料と1つまたは複数の核酸分子とをインキュベートする工程、および被験試料中の構成要素と核酸分子との結合についてアッセイを行う工程を含む。このようなアッセイ法は、典型的には、遺伝子の少なくとも1つが本発明の遺伝子および/または本発明の輸送体遺伝子の対立遺伝子である、多くの遺伝子を含むアレイに関連している。図3に、本発明の輸送体タンパ

50

ク質をコードする遺伝子内において見出された、SNPについての情報が示される。イントロン内ならびにORFの5'および3'領域内において、42個の異なるヌクレオチド位置でSNPが同定された。イントロン内およびORFの外側のそのようなSNPは、制御/調節要素に影響を及ぼしうる。エキソンにおける2つのSNPのうち、1つはアミノ酸配列の変化をもたらす(すなわち、非同義SNP)。これらのSNPがもたらすアミノ酸配列の変化は図3に示され、普遍的遺伝暗号および参照として図2に提供されるタンパク質配列を用いて容易に決定することができる。

【0189】

被験試料と核酸分子のインキュベーション条件は変化する。インキュベーション条件は、使用されるアッセイ法の形式、使用される検出方法、およびアッセイ法に使用される核酸分子のタイプおよび性質に依存する。当業者は、一般的に利用可能なハイブリダイゼーション、増幅、またはアレイアッセイ法の形式を認識していると思われ、これらは本明細書に開示されるヒトゲノムの新規断片を使用するために容易に適用することができる。このようなアッセイ法の例は、チャード T. (Chard, T.)、「放射標識免疫アッセイ法および関連する技術の概論 (An Introduction to Radioimmunoassay and Related Techniques)」、Elsevier Science Publishers、Amsterdam、The Netherlands (1986); ブロック G.R. (Bullock, G. R.) ら、「免疫細胞化学における技術 (Techniques in Immunocytochemistry)」、Academic Press、Orlando, FL、第1巻(1982)、第2巻(1983)、第3巻(1985); ティユセン P. (Tijssen, P.)、「酵素免疫アッセイ法の実践と理論: 生化学および分子生物学の実験技術 (Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology)」、Elsevier Science Publishers、Amsterdam、The Netherlands (1985) に記載されている。

10

20

【0190】

本発明の被験試料は、細胞、タンパク質、または細胞からの膜抽出物を含む。上記の方法に用いられる被験試料は、アッセイ法の形式、検出方法の性質、およびアッセイ法の試料として用いられる組織、細胞、またはその抽出物に基づいて変化する。核酸抽出物または細胞抽出物の調製方法は当技術分野において周知であり、使用されるシステムと適合する試料を得られるように、容易に適用させることができる。

【0191】

本発明の他の態様としては、本発明のアッセイ法を行うために必要な試薬を含むキットが提供される。

30

【0192】

特に、本発明は、(a) 本明細書に開示されるヒトゲノムの断片と結合可能な核酸分子を含む第一の容器と、(b) 1つまたは複数の洗浄試薬、結合核酸の存在を検出可能な試薬を含む、1つまたは複数の他の容器とを含む、1つまたは複数の容器に閉鎖的に封入され、区分されたキットを提供するものである。

【0193】

詳細には、区分されたキットには、試薬が別々の容器に含まれている任意のキットを含む。このような容器としては、小さいガラスの容器、プラスチック容器、帯状のプラスチック、ガラスもしくは紙、または二酸化ケイ素のようなアレイ材料を含む。このような容器は、試料と試薬が交叉汚染しないように、1つの区分から他の区分へと試薬を効率的に移動させることができ、それぞれの容器の試薬または溶液は、1つの区分から他の区分へと定量的に添加されることができる。このような容器には、被験試料を入れる容器、核酸プローブを含む容器、洗浄試薬(例えば、リン酸緩衝食塩水、トリス緩衝液等)を含む容器、および結合プローブの検出に用いられる試薬を含む容器を含むと考えられる。当業者は、本発明にかかる従来より未知の輸送体遺伝子を認識し、本明細書に開示される配列情報を用いて日常的に同定することができ、さらにこれを当技術分野において周知の確立されたキット形式、特に発現アレイに容易に組み込むことができる。

40

【0194】

ベクター/宿主細胞

50

本発明はまた、本明細書に記載される核酸分子を含むベクターを提供するものである。「ベクター」という用語は、媒体のことを言い、好ましくは核酸分子であり、核酸分子を輸送することができるものである。ベクターが核酸分子である場合、核酸分子はベクターの核酸と共有結合している。本発明のこの局面では、ベクターは、プラスミド、一本鎖もしくは二本鎖のファージ、一本鎖もしくは二本鎖のDNAもしくはRNAウイルスベクター、またはBAC、PAC、YACもしくはMACのような人工染色体を含む。

【0195】

ベクターは宿主細胞中に染色体外の要素として保持されてもよく、そこで核酸分子の付加的なコピーを複製および生成する。あるいは、ベクターは宿主細胞のゲノム中に組み込まれてもよく、宿主細胞の複製の際に核酸分子の付加的なコピーを生成する。

10

【0196】

本発明は、核酸分子の維持のためのベクター（クローニングベクター）、または核酸分子の発現のためのベクター（発現ベクター）を提供するものである。このベクターは、原核生物細胞もしくは真核生物細胞、またはその両方で機能することができる（シャトルベクター）。

【0197】

発現ベクターは、ベクター中で核酸分子と機能的に結合されたシス作用性調節領域を含み、これにより宿主細胞中での核酸分子の転写が可能となる。この核酸分子は転写に影響を及ぼしうる核酸分子と別々に、宿主細胞に導入されることができる。したがって、第二の核酸分子は、ベクターからの核酸分子の転写を可能にするシス調節制御領域と相互作用するトランス作用性因子を提供するものである。あるいは、トランス作用性因子は宿主細胞により提供されてもよい。最終的に、トランス作用性因子は、ベクター自身から作り出すことができる。しかし、いくつかの態様では、核酸分子の転写および/または翻訳は無細胞系でも起こり得ることが理解される。

20

【0198】

本明細書に記載される核酸分子が機能的に結合できる調節配列は、mRNA転写を誘導するためのプロモーターを含む。これらには、バクテリオファージからの左部プロモーター、大腸菌（*E. coli*）から得られたlac、TRP、およびTACプロモーター、SV40から得られた初期および後期プロモーター、CMV極初期プロモーター、アデノウイルス初期および後期プロモーター、ならびにレトロウイルスの末端反復配列が含まれるが、これに限定されるものではない。

30

【0199】

転写を促進する制御領域に加えて、発現ベクターはまた、リプレッサー結合部位およびエンハンサーのような転写を調節する領域を含むものであり得る。この例としては、SV40エンハンサー、サイトメガロウイルスの極初期のエンハンサー、ポリオーマエンハンサー、アデノウイルスエンハンサー、およびレトロウイルスLTRエンハンサーが含まれる。

【0200】

転写の開始および制御部位に加えて、発現ベクターはまた、転写終了のために必要な配列、および転写領域における転写のためのリボソーム結合部位を含むことができる。他の発現調節制御要素としては、ポリアデニル化シグナルと同様に、開始および停止コドンが含まれる。当業者には、発現ベクターに有用な多数の調節配列が既知であると思われる。このような調節配列は、例えば、サムブルック（Sambrook）ら、「分子クローニング：実験マニュアル（Molecular Cloning: A Laboratory Manual）」、第2版、Cold Spring Harbor Laboratory Press、Cold Spring Harbor、NY、(1989)に記載されている。

40

【0201】

様々な発現ベクターを、核酸分子の発現に用いることができる。このようなベクターには、染色体、エピソーム、およびウイルス由来のベクターが含まれ、これらは例えば、細菌プラスミド、バクテリオファージ、酵母エピソーム、人工酵母染色体を含む酵母染色体要素、バキュロウイルス、SV40のようなパポバウイルス、ワクシニアウイルス、アデノウイルス、ポックスウイルス、仮性狂犬病ウイルス、およびレトロウイルスのようなウイルス

50

由来のベクターである。ベクターはまた、これらの起源の組み合わせから得ることができ、例えば、コスミドとファージミドのようなプラスミドとバクテリオファージの遺伝子要素から得ることができる。原核および真核生物の宿主細胞のための適切なクローニングベクターおよび発現ベクターは、サムブルック (Sambrook) ら、「分子クローニング：実験マニュアル (Molecular Cloning: A Laboratory Manual)」、第2版、Cold Spring Harbor Laboratory Press、Cold Spring Harbor、NY、(1989) に記載されている。

【0202】

調節配列は、1つもしくは複数の宿主細胞の構成的な発現 (すなわち組織特異性) または、温度、養分添加、またはホルモンもしくは他のリガンドのような外因性因子による1つもしくは複数の細胞タイプでの誘導性の発現を提供するものである。原核および真核生物の宿主細胞において構成的および誘導的に発現する種々のベクターは、当業者に周知である。

10

【0203】

核酸分子を、周知の方法によってベクター核酸内に導入することができる。一般に、最終的に発現するDNA配列は、1つまたは複数の制限酵素によりDNA配列と発現ベクターとが開裂し、断片が互いにライゲーションすることによって、発現ベクターと結合される。制限酵素の消化およびライゲーションの手順は、当業者に周知である。

【0204】

適切な核酸分子を含むベクターは、公知の技術を用いて、増殖または発現のために適切な宿主細胞内へ導入することができる。細菌細胞には、大腸菌 (*E. coli*)、放線菌 (*Streptomyces*)、およびネズミチフス菌 (*Salmonella typhimurium*) が含まれるが、これに限定されるものではない。真核生物細胞には、酵母、ショウジョウバエ (*Drosophila*) のような昆虫細胞、COSおよびCHO細胞のような動物細胞、ならびに植物細胞が含まれるが、これらに限定されるものではない。

20

【0205】

本明細書に記載のように、融合タンパク質としてのペプチドの発現が望ましいと考えられる。したがって、本発明はペプチドの産生を可能にする融合ベクターを提供するものである。融合ベクターは組換えタンパク質の発現および組換えタンパク質の溶解性を向上させることができ、また、例えば、アフィニティ精製のためのリガンドの作用によってタンパク質精製を促進することができる。タンパク質分解性開裂部位は、融合部分との結合位置に導入され、このために、所望のペプチドを最終的に融合部分から分離することができる。タンパク質分解酵素は、ファクターXa、トロンピン、およびエンテロトランスポーター (enterotransporter) を含むが、これらに限定されない。典型的な融合発現ベクターとしては、グルタチオン-S-トランスフェラーゼ (GST)、マルトース結合タンパク質、またはタンパク質Aのそれぞれを、標的組換えタンパク質に融合した、pGEX (Smithら、Gene 67:31-40 (1988))、pMAL (New England Biolabs、Beverly、MA)、およびpRIT5 (Pharmacia、Piscataway、NJ) が含まれるが、これに限定されるものではない。好適な誘導性の非融合大腸菌 (*E. coli*) 発現ベクターの例としては、pTrc (Amannら、Gene 69:301-315 (1988))、およびpET 11d (Studierら、「遺伝子発現技術：酵素学における方法 (Gene Expression Technology: Methods in Enzymology)」、185:60-89 (1990)) が含まれる。

30

40

【0206】

組換えタンパク質の発現は、宿主細胞において、組換えタンパク質のタンパク質分解性の開裂欠損能力を持つ遺伝的背景を提供することによって、宿主細胞において最大化することができる (Gottesman, S., 「遺伝子発現技術：酵素学における方法 (Gene Expression Technology: Methods in Enzymology)」、185、Academic Press、San Diego、California (1990) 119-128)。あるいは、対象となる核酸分子の配列は、例えば、大腸菌 (*E. coli*) のような特定の宿主細胞のために優先的に使用されるコドンとなるように変更されることができる (Wadaら、Nucleic Acids Res. 20:2111-2118 (1992))。

【0207】

核酸分子はまた、酵母において作用する発現ベクターにより発現されることもできる。S.

50

セレビスエ (*S. cerevisiae*) のような酵母中で発現するベクターの例としては、pYepSec 1 (Baldariら、EMBO J. 6:229-234 (1987))、pMFa (Kurjanら、Cell 30:933-943(1982))、pJRY88 (Schultzら、Gene 54:113-123 (1987))、およびpYES2 (Invitrogen Corporation、San Diego、CA) を含む。

【0208】

核酸分子はまた、例えば、バキュロウイルス発現ベクターを用いて、昆虫細胞内で発現されることもできる。培養昆虫細胞 (例えば、Sf9細胞) 中のタンパク質の発現に利用されるバキュロウイルスベクターは、pAcシリーズ (Smithら、Mol. Cell Biol. 3:2156-2165 (1983)) およびpVLシリーズ (Lucklowら、Virology 170:31-39 (1989))を含む。

【0209】

本発明のある態様においては、本明細書に記載される核酸分子は、哺乳類発現ベクターを用いて哺乳類細胞内で発現される。哺乳類発現ベクターの例としては、pCDM8 (Seed, B. Nature 329:840(1987))およびpMT2PC (Kaufmanら、EMBO J. 6:187-195 (1987))を含む。

【0210】

本明細書に列記されている発現ベクターとしては、核酸分子を発現するために有用であり、当業者が利用可能な周知のベクターのみが例示されている。本明細書に記載される核酸分子の維持増殖または発現において、好適な他のベクターは、当業者に周知であると思われる。これらは、例えば、サムブルック J. (Sambrook, J.)、フリッツ E.F. (Fritsh, E. F.)、およびマニアティス T. (Maniatis, T.)、「分子クローニング：実験マニュアル (Molecular Cloning: A Laboratory Manual)」、第2版、Cold Spring Harbor Laboratory、Cold Spring Harbor Laboratory Press、Cold Spring Harbor、NY、1989に記載されている。

【0211】

本発明はまた、本明細書に記載される核酸配列がベクター中に逆方向にクローニングされたベクターを含むものであり、このベクターは、アンチセンスRNAの転写を可能にする調節配列と機能的に結合される。このように、アンチセンス転写は、コード領域および非コード領域の両方が含まれ、本明細書に記載される核酸分子配列の全部または一部を産生することができる。このアンチセンスRNAの発現は、センスRNAの発現 (調節配列、構成的または誘導性の発現、組織特異的発現) に関して、前記した各パラメータに対応する。

【0212】

本発明はまた、本明細書に記載されるベクターを含む組換え宿主細胞に関連するものである。したがって、宿主細胞は、原核生物細胞、酵母のような下等真核生物細胞、昆虫細胞のような他の真核生物細胞、および哺乳類細胞のような高等真核生物細胞を含む。

【0213】

組換え宿主細胞は、当業者が容易に利用可能な技術により、本明細書に記載されるベクター構築物を細胞中に導入することにより調製することができる。これらには、リン酸カルシウムトランスフェクション、DEAEデキストラン媒介トランスフェクション、陽イオン性脂質媒介トランスフェクション、エレクトロポレーション、トランスダクション、インフュクション、リポフェクション、およびサムブルック (Sambrook) ら、「分子クローニング：実験マニュアル (Molecular Cloning: A Laboratory Manual)」、第2版、Cold Spring Harbor Laboratory、Cold Spring Harbor Laboratory Press、Cold Spring Harbor、NY、1989) に記載されるような他の技術が含まれるが、これらに限定されるものではない。

【0214】

宿主細胞は、1つまたは複数のベクターを含むことができる。このため、異なるヌクレオチド配列が、同じ細胞の異なるベクター中に導入されることができる。同様に、核酸分子は、単独で、または発現ベクターのトランス作用性因子を提供しているような関連のない他の核酸分子と共に導入されることができる。1つまたは複数のベクターが細胞内に導入される場合、ベクターは単独で導入されるか、共に導入されるか、または核酸分子ベクターに結合して導入されることができる。

10

20

30

40

50

【0215】

バクテリオファージおよびウイルスベクターの場合、これらはインフェクションおよびトランスダクションの標準的な操作により、封入またはカプセル化されたウイルスとして細胞内に導入されることができる。ウイルスベクターは、複製可能、または複製欠陥であり得る。ウイルスの複製に欠陥がある場合、複製は欠陥を相補する機能が提供される宿主細胞内で起こり得る。

【0216】

ベクターは一般に、組換えベクターの構築物を含む細胞の部分母集団の選択を可能とする選択マーカーを含む。このマーカーは、本明細書に記載される核酸分子を含む同一のベクター内か、または別のベクター中に含まれることができる。マーカーは、原核生物宿主細胞のためのテトラサイクリンまたはアンピシリン耐性遺伝子、および真核生物宿主細胞のためのジヒドロ葉酸還元酵素またはネオマイシン耐性を含む。しかしながら、表現型特性の選択性を提供するマーカーはいずれも有効であると考えられる。

10

【0217】

成熟タンパク質は、適切な調節配列の制御下で、細菌、酵母、哺乳類細胞、および他の細胞で産生されることができるが、無細胞転写系および翻訳系もまた、本明細書に記載されるDNA構築物に由来のRNAを用い、これらのタンパク質を産生するために用いることができる。

【0218】

ペプチドの分泌が必要とされる場合、輸送体のようなタンパク質を含む複数回膜貫通ドメインで達成することは困難であり、適切な分泌シグナルがベクター中に組み込まれる。シグナル配列は、これらのペプチドに対して内因性であるか、またはペプチドに対して非相溶性であり得る。

20

【0219】

ペプチドが培地中に分泌されない場合、典型的には輸送体の場合、タンパク質を、凍結融解、超音波処理、機械的破壊、分解物質の使用等を含む標準的な破壊操作によって、宿主細胞から単離することができる。ペプチドは、硫酸アンモニウム沈殿、酸抽出、または陰イオンもしくは陽イオン交換クロマトグラフィ、ホスホセルロースクロマトグラフィ、疎水性相互作用クロマトグラフィ、アフィニティクロマトグラフィ、ヒドロキシルアパタイトクロマトグラフィ、レクチンクロマトグラフィ、または高速液体クロマトグラフィを含む、公知の精製方法によって、回収および精製されることができる。

30

【0220】

また、本明細書に記載されるペプチドの組換え産生においては宿主細胞に依存しており、ペプチドは細胞に依存して種々のグリコシル化パターンを持ち、細菌内で産生される場合にはグリコシル化されないであろうことが理解される。さらに、ペプチドは、宿主を媒介する過程の結果として、いくつかの場合で最初に修飾されたメチオニンを含むものであり得る。

【0221】

ベクターおよび宿主細胞の使用

本明細書に記載されるペプチドを発現する組換え宿主細胞には、種々の用途がある。まず、この細胞は、所望の量の輸送体タンパク質または断片を産生するため、さらに精製を行うことのできる輸送体タンパク質またはペプチドの産生に有用である。このため、発現ベクターを含む宿主細胞は、ペプチドの産生に有用である。

40

【0222】

宿主細胞はまた、輸送体タンパク質または輸送体タンパク質断片に関連している細胞系のアッセイ法、例えば上記したものおよび当技術分野に周知の他の形態のものの実施において有用である。このため、天然の輸送体タンパク質を発現する組換え宿主細胞は、輸送体タンパク質機能を刺激または阻害する化合物のアッセイに有用である。

【0223】

宿主細胞はまた、機能が影響を受ける輸送体タンパク質変異体を同定するために有用であ

50

る。変異が天然に生じ病理を引き起こすような場合、突然変異を含む宿主細胞は、天然の輸送体タンパク質の効果を示さずに、輸送体タンパク質変異体に望ましい効果（例えば、機能を刺激または阻害）を持つ化合物のアッセイに有用である。

【0224】

遺伝的に操作された宿主細胞は、さらにヒト以外のトランスジェニック動物を産生するために用いることができる。遺伝子組換え動物は、好ましくは哺乳類であり、例えば、この動物の1つまたは複数の細胞が導入遺伝子を含むラットまたはマウスのような齧歯類動物である。導入遺伝子は発達中のトランスジェニック動物の細胞のゲノムに組み込まれ、1つまたは複数の細胞型または組織において、成熟した動物のゲノム中に残存する外因性のDNAである。これらの動物は、輸送体タンパク質の機能の研究、ならびに輸送体タンパク質活性のモジュレータの同定および評価に有用である。トランスジェニック動物の他の例としては、ヒト以外の霊長類、ヒツジ、イヌ、ウシ、ヤギ、ニワトリ、および両生類が含まれる。

10

【0225】

トランスジェニック動物は、例えば、マイクロインジェクション、レトロウイルス感染によって、受精卵母細胞の雄性前核内に核酸を導入し、卵母細胞を偽妊娠性の雌性育成動物中で発達させることにより作製される。任意の輸送体タンパク質ヌクレオチド配列を、マウスのようなヒト以外の動物のゲノム中に導入遺伝子として導入することができる。

【0226】

発現ベクターに有用な調節配列または他の配列は、いずれも導入遺伝子配列の一部を形成することができる。これには、既に含まれない場合、イントロン配列およびポリアダニル化シグナルが含まれる。組織特異性調節配列は、特定の細胞に対し輸送体タンパク質が直接発現するために、導入遺伝子に機能的に結合されることができる。

20

【0227】

胚操作およびマイクロインジェクションを通して、トランスジェニック動物を産生する方法、特にマウスのような動物を用いる方法は、当技術分野において一般化されており、例えば、リーダー（Leder）らの米国特許第4,736,866号および同第4,870,009号、ワグナー（Wagner）らの米国特許第4,873,191号、およびホーガン B.（Hogan, B.）、「マウス胚の操作（Manipulating the Mouse Embryo）」、（Cold Spring Harbor Laboratory Press、Cold Spring Harbor, N.Y., 1986）に記載されている。同様の方法が、他のトランスジェニック動物の産生のために用いられている。トランスジェニック創始動物を、ゲノム中の導入遺伝子の存在および/または動物の組織もしくは細胞内でのトランスジェニックmRNAの発現に基づいて確認することができる。その後、トランスジェニック創始動物を、さらに導入遺伝子を有する動物を繁殖させるために用いることができる。さらに、導入遺伝子を有するトランスジェニック動物を、さらに他の導入遺伝子を有する他のトランスジェニック動物と交配することができる。トランスジェニック動物はまた、動物全体または動物の組織が本明細書に記載される相同的な組換え宿主細胞を用いて作製された動物を含む。

30

【0228】

他の態様では、ヒト以外のトランスジェニック動物は、導入遺伝子の調節された発現を行う選択システムを含むものとして産生されることができる。このようなシステムの1つの例は、バクテリオファージP1のcre/loxPリコンビナーゼシステムである。cre/loxPリコンビナーゼシステムについての記載は、例えば、ラクソ（Lakso）ら、PNAS 89:6232-6236（1992）参照。リコンビナーゼシステムのもう一つの例は、S.セレビシエ（*S. cerevisiae*）のFLPリコンビナーゼシステムである（O'Gormanら、Science 251:1351-1355（1991）。cre/loxPリコンビナーゼシステムが導入遺伝子の発現の調節に用いられる場合は、動物において、creリコンビナーゼおよび選択されたタンパク質の両者をコードする導入遺伝子が含まれていることが必要である。このような動物は、例えば、一方は選択されたタンパク質をコードする導入遺伝子を持ち、他方はリコンビナーゼをコードする導入遺伝子を持った2つのトランスジェニック動物を交配させることにより、「二重」トランスジェニッ

40

50

ク動物を作製することによって提供される。

【0229】

本明細書に記載されるヒト以外の遺伝子組換え動物のクローンはまた、ウィルムト I. (Wilmut, I.) ら、Nature 385:810-813 (1997) および国際公開公報第 97/07668 号および国際公開公報第 97/07669 号に記載される方法に従って産生されることができる。簡単に述べると、トランスジェニック動物からの細胞、例えば体細胞を単離し、増殖周期から出て G₀ 期に入るように誘導することができる。静止細胞を、例えば電気パルスの使用によって、静止細胞を単離した動物と同種の動物の除核した卵母細胞と融合することができる。再構成された卵母細胞は、桑実胚または胚盤胞に発達するよう培養され、その後、偽妊娠性の雌性育成動物中に移される。この雌性育成動物から誕生する子孫は、細胞、例えば体細胞を単離した動物のクローンとなる。

10

【0230】

本明細書に記載されるペプチドを発現する組換え細胞を含むトランスジェニック動物は、インビボの環境で、本明細書に記載したようなアッセイを行うために有用である。したがって、インビボに存在し、リガンド結合、輸送体タンパク質活性化、シグナル伝達に影響を与えうる各種の生理学的因子は、インビトロの無細胞または細胞系のアッセイ法では明らかにならない可能性がある。したがって、これらは、リガンド相互作用、輸送体タンパク質機能およびリガンド相互作用に対する特定の変異輸送体タンパク質の影響、ならびにキメラ輸送体タンパク質の影響を含む、インビボでの輸送体タンパク質機能をアッセイするための、ヒト以外のトランスジェニック動物を提供するために有用である。また、実質的にまたは完全に 1 つまたは複数の輸送体タンパク質機能を除去する突然変異である、ヌル変異の影響を評価することも可能である。

20

【0231】

本明細書において、上記の全ての刊行物および特許は参照として本明細書に組み入れられる。本発明に記載された方法およびシステムの各種修正および変形は、本発明の範囲および趣旨から逸脱しない限り、当業者において明らかなものであると思われる。本発明は、特定の好ましい態様に関連して記述されているが、特許請求の範囲に記載された発明は、このような特定の態様に不当に限定されないと理解されるべきである。実際に、本発明を実施するための上記方法の各種変形は、分子生物学または関連する分野の業者において明らかであり、このようなものも特許請求の範囲に含まれることが意図される。

30

【図面の簡単な説明】

【0232】

【図 1】本発明の輸送体タンパク質をコードする cDNA 分子および転写配列ののヌクレオチド配列を提供する。加えて、この分子配列に基づいて本発明の具体的な用途を容易に決定することを可能にする、ATG 開始コドン、停止コドン、および組織分布などの入手可能な構造および機能情報も提供する。図 1 に提供した実験データは、卵巣（腺癌組織）、子宮（平滑筋肉腫組織）、頸部、腎臓、腎臓癌組織（副腎腫）、胚中心 B 細胞、結腸、および乳児脳での発現を示す。

【図 2】本発明の輸送体の予想されるアミノ酸配列を提供する。加えて、この分子配列に基づいて本発明の具体的な用途を容易に決定することを可能にする、タンパク質ファミリー、機能、および修飾部位などの入手可能な構造および機能情報も提供する。

40

【図 3】本発明の輸送体タンパク質をコードする遺伝子の全域にわたるゲノム配列を提供する。加えて、この分子配列に基づいて本発明の具体的な用途を容易に決定することを可能にする、イントロン/エキソン構造、プロモーター位置などの入手可能な構造および機能情報も提供する。図 3 に示すように、挿入/欠失変異体（「インデル」）を含む SNP が、42 個の異なるヌクレオチド位置で同定された。

【図 1 A】

配列番号 1:
(配列番号 1)
1 ATGACCCMGG GGAAGAGGA GAAACGGCC CCGAACCGA STATACGT
51 GCGCAGAGG ATCATCTTA AGACGGAGG CAGCTCTCA GGAATAGCT
101 CCGTATAGT GTATATGCA GTATTTGGA GTTTTTCCT
151 TGGGACATAT TGACAGACG CAGCTTGCT GTATTACAG AACCTTTTC
201 TAACATACA TTCTCTGGA AGCTTATAT TGAGAGGTA AGGCTTTCT
251 TGCTATCTT TAGTGGCCCG CTATTGCTG CTATTCTGA TGTGTGGCG
301 GGAATATCT TTCTGCTGT AACGTGCTT TTCACTATG CCGCAATTC
351 TTATAGAGG ATGACCGCG GGGGTATCT TGTGTATAT TGTGTTCCT
401 GGGTTTTCG AGTACATCT TTCTGTGAT TTCACTATG AGCAATATA
451 ACCAGAGAG ATGAGAGAG TAGTGTCTT GAACTGTCTT CAGCAATAT
501 TCTCTAGAT TTATGTCGA GTTCTGCAAT TGAGCTTAT CTGGAGGAG
551 TATATAGGA CAGCTTGCTG GTAGCTTAG CTACAGCAT AGCTTCTTA
601 GAAATATCT TTATCTCTG TCTCTGCGA GATCTGTCT CTGAGAAAT
651 GGGGCGGCA TCTTGGGCG CAGCTTCTT CTGGAGGTA GCTGTGCTT
701 TTGCTCTCT AAAAAGAGT GGGCAGAT CTATATCTG GTGTATCTG
751 ATTACAGCT TTCTCTCTA CTATCTGAG GAGGCGCAT ATTCTAGCT
801 TTTTATAGC CTACAGACA TATGCAAT TTCTGCGGA AGTGTGCGA
851 CCGTATAGC AGTCTTTCG ATTCTCTCA TTATGTCGA GAGCTATCT
901 TTGCTTATC TTATGCTG ATTTGGAAT AGAAGCGCA TTATCTGCG
951 TCTGCACTT GAAATATCT AGTGGAGAT GTATGCTCT GGTTCAGAG
1001 CTGCAATGAT GGGGCTCTT GGGGCGATG CAGCTATCT TAGCATCAC
1051 TTCTCTCTG TACGCTACT TTCTCTGCA AGTCTGTAG CTGATACCA
1101 GGGTCTCTT CAGGAGTCA TAGAGAGAT TGAGGCTTA TCGATGCTG
1151 TGGGACCGG CCGTATGGA TTATTTCTT AGATATCTA TGTGAGCTT
1201 AAGAAGCTC CAAATACAG AACAGCTTG GAGAGACCA GAGCTACCA
1251 GAGCAGCTT GAGAGAGAT CAGCAGCGG TGCGCTCTT TCTATCTG
1301 GAGCTTCTT AGTATCTG GCTCTGCTG TTCTCTCTT TATCTCTGA
1351 CATACATAT TAGCTTAG CTACAGCT AGGAGAGAG ACTGTGCGC
1401 TCAAGAGAT CCGTATATA CAGAGCGCG AGGAGCGCG AAGAGACTT
1451 TACTCGAGA CCAATATCTG TGA

転写物の特徴:
開始コドン: 1
終止コドン: 1471

転写物のトランスクリプト組立:
gi16880221|ref|NP_032272.1| hippocampus abundant gene transcrip... 977 0-0
gi17292310|gb|AAAF7717.1| (AA003477) GCL1537 gene product [alt... 557 e-157
gi17508400|pir|J23372 hypothetical protein T25D3.1 - Caenorhab... 474 e-132
gi16802881|emb|CA870819.1| (AJ137576) hypothetical protein [Rom... 388 e-106
gi17296325|gb|AAAF5161.1| (AA003591) GCL1537 gene product [Dros... 382 e-105
gi17296325|gb|AAAF5161.1| (AA003591) GCL1537 gene product [Dros... 183 3e-45
gi17296325|gb|AAAF5161.1| (AA003591) GCL1537 gene product [Dros... 183 4e-45
gi17473952|pir|J27515 tetracycline-efflux transporter - Deinoc... 139 8e-32
gi13860032|gb|AA009860.1| (AF090987) drug efflux protein TetA (... 132 6e-30
gi12078521|gb|AAAF7664.1| (M02189) tet protein [synthetic const... 132 6e-30

【図 1 C】

転写物のトランスクリプト組立:
gb|BF036183|BF036183 601458251F1 NIH_MSC_66 Homo sapiens cDNA c... 1197 0-0
gb|B0384890|B0384890 601506823F1 NIH_MSC_71 Homo sapiens cDNA c... 1150 0-0
gb|B0539469|B0539469 601060131F2 NIH_MSC_10 Homo sapiens cDNA c... 1013 0-0
gb|BF029950|BF029950 601557355F1 NIH_MSC_58 Homo sapiens cDNA c... 912 0-0
gb|AA223840|AA223840 r12209.r1 Stragatine hnt neuron (8597233)... 876 0-0
gb|AA641767|AA641767 na12008.r1 NCI-OSM C638 Homo sapiens cDNA... 837 0-0
gb|B67165|B67165 yf08-08.r2 Soares infant brain N18B Homo sapie... 753 0-0
gb|AA099754|AA099754 n179a03.r1 Stragatine colon (4537204) Homo... 755 0-0

転写物使用のための変異情報:
ウィルソン変異:
gb|B036183|Ovary, adenocarcinoma
gb|B0384890|Uterus Leiomyocarcinoma
gb|B0539469|Cervix
gb|B029950|Kidney hypernephroma
gb|AA223840|Neuron
gb|AA641767|germinal center B Cell
gb|B67165|R Infant brain
gb|AA099754|Colon
変異を用いた相対スクリーニングハザルカでの変異情報:
変異情報

【図 1 B】

配列番号 2:
(配列番号 1)
1 atgacccagg ggaagagga gaaacggcc ccgaaacga statatcgt
51 ggcagagag atcatctta agacggagc cagctctca ggaatagct
101 cgcgtaggt gtatattga gtatttggg gtttttctt
151 ttgttgtat tocttagtc cccgttatt ggtgtcttt ctatgtttg
201 gggcgaaa tocttttgc tgacagagc gttttcaga tggcgaaa
251 tcccttaat gaagatagc caatgttgt acrttgtgt tatctcttt
301 ttgggggtt tggagtagc ttittctgt gtatttgt acpiagaga
351 taaacaaa gagaagaaa gagtattgc ttatagagc gtttcaga
401 catttgcg aagttagc acagtctgt caatggagc ttatctgt
451 cgatattgt gggagagct ggtgtgtgc ttatctagc caatagctt
501 gtagatatt tttttatcc ttgttgtgc gtagagtag ttgcgaga
551 aaatggcgc agatctctg ggaacaaaa ttcttgga acaagtagc
601 ccttttgtc ctttaaaaa agtggccca gttccatag tcttgtgat
651 ctgattaca gttttctct cttccctacc gggggagc caatattcc
701 gctttttt atccctaga cagatattg aatttccac agaatgttt
751 cgagcttta tagactctt tggactctt tccattatt cacagacct
801 agcttgagc ttactatga gttcaattg aatagagc acctttcc
851 tgggtcagg atttcaaat ttacatttg catgtatg ctttgttca
901 gaacttga tgaatggc tctgtggga gtagcaga tctctgat
951 cactttcct gctgtcgt cactgttcc agaatgct gctgtgat
1001 aacagpctt cttcaagga atattacag gaattcgt atttccat
1051 gctctgggc cggcctcta tggattcatt tctacatc tccatgga
1101 acttataga ctgcaataa cagagacga cttgggaga aacacagc
1151 ctgagaca cttgagagc aattccatc tcttgggc tcttctca
1201 tttagagct gttcagct cttgtctgt ctttgtct tttttatcc
1251 ggaacatcc aatttagc taagttagc cagtggga agaatgag
1301 gtagtcag catctctc aattacag cttcgaga ggaacagga
1351 cctttatcc agaacacaa tgttga

転写物の特徴:
開始コドン: 1
終止コドン: 1375

転写物のトランスクリプト組立:
gi16880221|ref|NP_032272.1| hippocampus abundant gene transcrip... 895 0-0
gi17292310|gb|AAAF7717.1| (AA003477) GCL1537 gene product [alt... 509 e-143
gi17508400|pir|J23372 hypothetical protein T25D3.1 - Caenorhab... 437 e-121
gi11360183|pir|J74272 hypothetical protein DFP6410864.1 - ... 388 e-107
gi17296327|gb|AAAF5161.1| (AA003591) GCL1537 gene product [Dros... 381 e-105
gi17296326|gb|AAAF5161.1| (AA003591) GCL1537 gene product [Dros... 145 1e-33
gi17296325|gb|AAAF5161.1| (AA003591) GCL1537 gene product [Dros... 145 1e-33
gi17473952|pir|J27515 tetracycline-efflux transporter - Deinoc... 138 8e-32
gi13860032|gb|AA009860.1| (AF090987) drug efflux protein TetA (... 132 6e-30
gi12078521|gb|AAAF7664.1| (M02189) tet protein [synthetic const... 132 6e-30
gi1184318|gb|AA003625.1| (U01194) tetracycline resistance prote... 132 8e-30
gi1236519|gb|AA0092917.1| (U48855) tetracycline resistance prot... 131 1e-29
gi1507411|ref|NP_052111.1| protein responsible for tetracycline... 131 1e-29
gi1221494|emb|CA84021.1| (AA00346) Mapc protein [Entococcu... 127 2e-29
gi10946601|ref|NP_052571.1| tetracycline resistance protein Te... 126 4e-28

【図 2 A】

転写物のトランスクリプト組立:
(転写物 1)
1 MTGCKKKA AHSRGLAKK IIRKGOTPO GTSRPSVHA VIVITLEFFA
51 WLLTAPET VIRETTEPT FLNGLIOGV KGLSLFAP LIGLSLDWG
101 KSLTLTVE TPCFIRLME TSPWYRVR SVSUVATVE SVFVAFADI
151 TORHERMA GLVSATFAS LVTSFAGY LGVYGDVLY VVLAATALL
201 DCFILWAV ESLPERRA SWAPINMO ADFASLKY SGRSLVLC
251 IVFVLTME AGPSRPTF LQCHMSSE SVAFVAVL LSLVAGTV
301 LSLMSIGE WLLGLGGE QILGLAYCF GSEFMDKRA GAVAMSEIT
351 FPAVSALVR TADADQGVV QGNITIGEL ONGLCPALV FFFIFHVE
401 KSLPTTDE GTWSPGHF EGNITIGP TPCACVLL ALVAFVPE
451 HTNLSISS WDCCHSHR RHTVAGGA RFLVQTHV

転写物使用のための変異情報:
ウィルソン変異:
1 45 65 1.705 4e-11
2 85 105 1.209 4e-11
3 108 128 1.726 4e-11
4 131 151 2.108 4e-11
5 163 183 1.582 4e-11
6 185 215 1.732 4e-11
7 248 268 1.518 4e-11
8 287 307 1.974 4e-11
9 317 337 1.176 4e-11
10 340 360 1.861 4e-11
11 385 405 1.393 4e-11
12 435 455 2.247 4e-11

【図 2 B】

タンパク質配列に対する上位ヒットとのBLASTクエリメント:
 >gi|668022|ref|NP_032272.1| hippocampus abundant gene transcript 1
 (Mus musculus) >gi|751396|pir|JC5641 sugar transporter
 protein NIA1 - mouse >gi|1506078|dbj|BA22622.1|
 (DB8315) tetracycline transporter-like protein [Mus
 musculus]
 Length = 490

Score = 977 bits (2498), Expect = 0.0
 Identities = 486/490 (99%), Positives = 490/490 (99%)

クエリ: 1 MTQKKEKRAANSIMAAKLIIRKGGTPOGISPSVHIVIFLEFFWMLGTAFTLV 60
 MTQKKEKRAANSIMAAKLIIRKGGTPOGISPSVHIVIFLEFFWMLGTAFTLV 60
 MTQKKEKRAANSIMAAKLIIRKGGTPOGISPSVHIVIFLEFFWMLGTAFTLV 60

サブジェクト: 61 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120
 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120
 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120

クエリ: 121 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180

サブジェクト: 121 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180

クエリ: 181 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240

サブジェクト: 181 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240

クエリ: 241 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300

サブジェクト: 241 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300

クエリ: 301 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360

サブジェクト: 301 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360

クエリ: 361 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420

サブジェクト: 361 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420

クエリ: 421 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480

サブジェクト: 421 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480

クエリ: 481 KEPLQDQWV 490
 KEPLQDQWV 490 (配列番号: 6)

サブジェクト: 481 KEPLQDQWV 490 (配列番号: 6)

配列結果	説明	スコア	E値	N
PF00083	糖(および糖能)の輸送体	3.7	2.4	2

ドメインの解析:

モデル	ドメイン	seq-f	seq-t	hm-f	hm-t	スコア	E値		
PF00083	1/2	95	129	..	63	116	..	6.6	0.35
PF00083	2/2	422	450	..	442	474	..	0.2	24

【図 2 D】

タンパク質配列に対する上位ヒットとのBLASTクエリメント:
 >gi|668022|ref|NP_032272.1| hippocampus abundant gene transcript 1
 (Mus musculus) >gi|751396|pir|JC5641 sugar transporter
 protein NIA1 - mouse >gi|1506078|dbj|BA22622.1|
 (DB8315) tetracycline transporter-like protein [Mus musculus]
 Length = 490

Score = 895 bits (2288), Expect = 0.0
 Identities = 454/490 (92%), Positives = 458/490 (92%), Gaps = 32/490 (6%)

クエリ: 1 MTQKKEKRAANSIMAAKLIIRKGGT----- 28
 MTQKKEKRAANSIMAAKLIIRKGGT----- 28
 MTQKKEKRAANSIMAAKLIIRKGGT----- 28

サブジェクト: 29 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 80
 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 80
 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 80

クエリ: 61 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120
 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120
 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120

サブジェクト: 61 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120
 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120
 VLHETFFRHTFLMGLIQGVKGLSFLSAPLIGALSVDWGRKSFLLVFTFCAPILMK 120

クエリ: 89 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 148
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 148
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 148

サブジェクト: 89 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 148
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 148
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 148

クエリ: 121 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180

サブジェクト: 121 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 ISPMWTFVIVSVGVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180

クエリ: 149 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 208
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 208
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 208

サブジェクト: 149 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 208
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 208
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 208

クエリ: 181 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240

サブジェクト: 181 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 LGRVYGDVSVVLATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240

クエリ: 209 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 268
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 268
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 268

サブジェクト: 209 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 268
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 268
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 268

クエリ: 241 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300

サブジェクト: 241 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 GQSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300

クエリ: 269 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 328
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 328
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 328

サブジェクト: 269 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 328
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 328
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 328

クエリ: 301 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360

サブジェクト: 301 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 LSLMRSIGKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360

クエリ: 329 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 388
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 388
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 388

サブジェクト: 329 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 388
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 388
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 388

クエリ: 361 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420

サブジェクト: 361 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420
 TADADQGVVQMTGTIGLNCGLGAPALVGFYIYTHVLEKLPITDGLGNTSPQHPF 420

クエリ: 389 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 448
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 448
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 448

サブジェクト: 389 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 448
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 448
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 448

クエリ: 421 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480

サブジェクト: 421 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 EQNSTIPGPFPLGACSVLLALVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480

クエリ: 449 KEPLQDQWV 490
 KEPLQDQWV 490 (配列番号: 7)

サブジェクト: 449 KEPLQDQWV 490 (配列番号: 7)

HMM結果:					
モデル	説明	スコア	E値	N	
PF00083	糖(および糖能)の輸送体	3.7	2.4	2	

ドメインの解析:

モデル	ドメイン	seq-f	seq-t	hm-f	hm-t	スコア	E値		
PF00083	1/2	53	97	..	63	116	..	6.6	0.35
PF00083	2/2	390	418	..	442	474	..	0.2	24

【図 2 C】

タンパク質配列 2:
 (配列番号: 14)
 1 MTQKKEKRAANSIMAAKLIIRKGGTPOGISPSVHIVIFLEFFWMLGTAFTLV 60
 51 LKSTFAGRLI GALSVDWGRKSFLLVFTFCAPILMK 120
 101 SOVFAVTFVSVFVTFVSVFVAVADITQKHSRMAVGLVSATFASLVTSPALGAY 180
 151 RYVIGDELVVV LATAIALDLCFLVAVPESLPERMPSAINGAPISWEQAPFASLKV 240
 201 PPSALAKVQVQ DSVILLCITVFLVSLPEAGQVSSFLVLRKINRFSPEVAATFVGLSILAGTIV 300
 251 AATVAVGLIL SILAQITVLS LMRISIGNKNTILLGLGQFLQLAWYGFSEFWMMAAGAVAMSSITFFAVSALVSR 360
 301 KPMNMWAGA VAMSSITFFAVSALVSR 360
 351 GSAFALVYFI FYTHVLEKLPITDGLGNTSPQHPF 420
 401 PMSCLVIAL LVALFPEHTNLSLRSSWRKCGSHRSHHTQAPGEA 480
 451 ELKQDQWV

特徴:
 機能的ドメインと主要領域:
 タンパク質配列中に存在するドメインの結果:
 [1] P000001 P000001 ASN_GLYCOSYLATION
 Nグリコシル化部位
 マッチ数: 2
 1 12-15 NRST
 2 421-424 LMSL

[2] P000005 P000005 PKC_PHOSPHO_SITE
 カドニウムイオン-リン酸化部位
 マッチ数: 3
 1 204-206 SLK
 2 423-425 SLR
 3 428-430 SMR

[3] P000006 P000006 CR2_PHOSPHO_SITE
 カドニウムイオン-リン酸化部位
 マッチ数: 2
 1 180-183 SLRP
 2 374-377 TQTD

[4] P000008 P000008 MYRISTYL
 N-ミリスチン酸部位
 マッチ数: 12
 1 43-48 GLIQTV
 2 129-134 GLVST
 3 146-151 GAYLGR
 4 191-196 GAPISW
 5 230-235 GQVSEF
 6 277-282 GQVSEF
 7 309-314 GAVVAM
 8 326-341 GQVQGN
 9 340-345 GQVSEF
 10 347-359 GLQVGL
 11 351-356 GLQVGL
 12 375-380 GQVSEF

[5] P000009 P000009 AMINATION
 アミノ化部位
 マッチ数: 2
 1 3-6 QSKK
 2 67-70 WSKK

[6] P000013 P000013 GRAM_POS_ANCHORING
 グラム陽性細胞膜アンカリング [144] ハンサムド
 371-376 LPITGT

[7] P000019 P000019 SUGAR_TRANSPORT_1
 糖輸送タンパク質ファミリー
 60-76 IGLASVWGRKSFLLT

【図 3 A】

(配列番号: 15)
 1 AGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 51
 51 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 51
 101 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 101
 151 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 151
 201 KCCACACAGC CTGATATCTG TGTATATTTT GTGTTTGT GGGATTAA 201
 251 TACTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 251
 301 TTTTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 301
 351 GTCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 351
 401 AGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 401
 451 GTCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 451
 501 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 501
 551 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 551
 601 AGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 601
 651 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 651
 701 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 701
 751 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 751
 801 KCCACACAGC CTGATATCTG TGTATATTTT GTGTTTGT GGGATTAA 801
 851 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 851
 901 TACTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 901
 951 GTCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 951
 1001 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1001
 1051 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1051
 1101 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1101
 1151 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1151
 1201 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1201
 1251 GTCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1251
 1301 AGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1301
 1351 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1351
 1401 TTTTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1401
 1451 CAATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1451
 1501 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1501
 1551 TGTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1551
 1601 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1601
 1651 TACTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1651
 1701 TACTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1701
 1751 ATCTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1751
 1801 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1801
 1851 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1851
 1901 TGTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1901
 1951 TGTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 1951
 2001 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2001
 2051 TTTTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2051
 2101 TTTTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2101
 2151 TTTTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2151
 2201 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2201
 2251 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2251
 2301 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2301
 2351 TTTTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2351
 2401 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2401
 2451 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2451
 2501 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2501
 2551 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2551
 2601 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2601
 2651 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2651
 2701 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2701
 2751 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2751
 2801 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2801
 2851 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2851
 2901 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2901
 2951 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 2951
 3001 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3001
 3051 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3051
 3101 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3101
 3151 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3151
 3201 GGCATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3201
 3251 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3251
 3301 TTTTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3301
 3351 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3351
 3401 CTGATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3401
 3451 TGTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3451
 3501 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3501
 3551 AATATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3551
 3601 TGTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3601
 3651 TGTATATCTG TATTTTTCAC TGTATATTTT GTGTTTGT GGGATTAA 3651

[illegible]

[illegible]

【 ㊦ 3 J 】

[illegible]

【 図 3 K 】

37501	AATTATGTC	CCCTGACGGG	TTGTGGCCCA	TTGCTGCTT	AGGCGCAAT
37502	TATCTGTTC	GGGCGGAGTA	TTGCTGCTT	TTCCTGAT	ATACGCGCA
37503	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37504	CGATCTTTC	CTATACGAGC	CTTCTGTGG	GGGTACTTC	CACATCTAT
37505	TTCTCGATCG	CGGCTGTGAG	CTGCTATAT	TTGCTGCTT	TTGCTGCTT
37506	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37507	CCGCTATCT	CTCTATCTAT	CTGATGTGCT	TCACGACGC	GGGCTGCTT
37508	CGGCTGAGG	CTGCTGCTT	ATTCGCTGCT	TTGCTGCTT	TTGCTGCTT
37509	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37510	TGATCATCA	GGGTTCTCA	TTCTGCCAT	TATGATAGAT	CTCATCTAG
37511	CGGCTGCTT	ATCTGCTGCT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37512	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37513	CGATCTTTC	CTATACGAGC	CTTCTGTGG	GGGTACTTC	CACATCTAT
37514	TTCTCGATCG	CGGCTGTGAG	CTGCTATAT	TTGCTGCTT	TTGCTGCTT
37515	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37516	CCGCTATCT	CTCTATCTAT	CTGATGTGCT	TCACGACGC	GGGCTGCTT
37517	CGGCTGAGG	CTGCTGCTT	ATTCGCTGCT	TTGCTGCTT	TTGCTGCTT
37518	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37519	TGATCATCA	GGGTTCTCA	TTCTGCCAT	TATGATAGAT	CTCATCTAG
37520	CGGCTGCTT	ATCTGCTGCT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37521	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37522	CGATCTTTC	CTATACGAGC	CTTCTGTGG	GGGTACTTC	CACATCTAT
37523	TTCTCGATCG	CGGCTGTGAG	CTGCTATAT	TTGCTGCTT	TTGCTGCTT
37524	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37525	CCGCTATCT	CTCTATCTAT	CTGATGTGCT	TCACGACGC	GGGCTGCTT
37526	CGGCTGAGG	CTGCTGCTT	ATTCGCTGCT	TTGCTGCTT	TTGCTGCTT
37527	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37528	TGATCATCA	GGGTTCTCA	TTCTGCCAT	TATGATAGAT	CTCATCTAG
37529	CGGCTGCTT	ATCTGCTGCT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37530	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37531	CGATCTTTC	CTATACGAGC	CTTCTGTGG	GGGTACTTC	CACATCTAT
37532	TTCTCGATCG	CGGCTGTGAG	CTGCTATAT	TTGCTGCTT	TTGCTGCTT
37533	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37534	CCGCTATCT	CTCTATCTAT	CTGATGTGCT	TCACGACGC	GGGCTGCTT
37535	CGGCTGAGG	CTGCTGCTT	ATTCGCTGCT	TTGCTGCTT	TTGCTGCTT
37536	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37537	TGATCATCA	GGGTTCTCA	TTCTGCCAT	TATGATAGAT	CTCATCTAG
37538	CGGCTGCTT	ATCTGCTGCT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37539	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37540	CGATCTTTC	CTATACGAGC	CTTCTGTGG	GGGTACTTC	CACATCTAT
37541	TTCTCGATCG	CGGCTGTGAG	CTGCTATAT	TTGCTGCTT	TTGCTGCTT
37542	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37543	CCGCTATCT	CTCTATCTAT	CTGATGTGCT	TCACGACGC	GGGCTGCTT
37544	CGGCTGAGG	CTGCTGCTT	ATTCGCTGCT	TTGCTGCTT	TTGCTGCTT
37545	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37546	TGATCATCA	GGGTTCTCA	TTCTGCCAT	TATGATAGAT	CTCATCTAG
37547	CGGCTGCTT	ATCTGCTGCT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37548	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37549	CGATCTTTC	CTATACGAGC	CTTCTGTGG	GGGTACTTC	CACATCTAT
37550	TTCTCGATCG	CGGCTGTGAG	CTGCTATAT	TTGCTGCTT	TTGCTGCTT
37551	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37552	CCGCTATCT	CTCTATCTAT	CTGATGTGCT	TCACGACGC	GGGCTGCTT
37553	CGGCTGAGG	CTGCTGCTT	ATTCGCTGCT	TTGCTGCTT	TTGCTGCTT
37554	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37555	TGATCATCA	GGGTTCTCA	TTCTGCCAT	TATGATAGAT	CTCATCTAG
37556	CGGCTGCTT	ATCTGCTGCT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37557	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37558	CGATCTTTC	CTATACGAGC	CTTCTGTGG	GGGTACTTC	CACATCTAT
37559	TTCTCGATCG	CGGCTGTGAG	CTGCTATAT	TTGCTGCTT	TTGCTGCTT
37560	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT	TTGCTGCTT
37561	CCGCTAT				

【 図 3 L 】

[illegible]

【 図 3 M 】

45901	AGGTGCTGAG	GTGCTGATCA	TTTTCACCA	GCGATGTTT	AGAGGTGCTA
45902	ATGCTACGGGA	ATTAAGTAGT	TTTGTGAGC	TTGCTTCATG	TGAGCATATG
45903	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45904	GTGATGCTG	GTGCTGCTAA	ATATATATAG	TTTTCATGCA	CATTTCTGCA
45905	ATGCTACGAT	GTGCTGATAT	AGCATATGAT	TGATGATATG	TGATGATATG
45906	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45907	ATGATTTGAT	TAAACATGAT	AGGCTGTGAT	CAGAGGATGA	TACAGAGGAT
45908	TCGATGAGTA	TGCTATGATG	TGCTATGATG	CTGCTATGAT	TGCTATTTCT
45909	TTGCTGCTG	TGCTATGATG	TGCTATGATG	TGCTATGATG	TGCTATTTCT
45910	ACATGATGCA	CAGGCTGCTA	GACCATCTAT	GACCATCTAT	TTCTCTATCT
45911	ACATGATGCA	TATGATATGAT	TATGATATGAT	ATGATTAAGG	CTGCTGCTCT
45912	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45913	TTTAAACAGC	TTGCTATTTG	TTGCTATTTG	TTGCTATTTG	TGAGCATATG
45914	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45915	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45916	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45917	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45918	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45919	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45920	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45921	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45922	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45923	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45924	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45925	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45926	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45927	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45928	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45929	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45930	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45931	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45932	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45933	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45934	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45935	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45936	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45937	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45938	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45939	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45940	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45941	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45942	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45943	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45944	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45945	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45946	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45947	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45948	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45949	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45950	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45951	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45952	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45953	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45954	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45955	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45956	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45957	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45958	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45959	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TTGCTGCTG	TGAGCATATG
45960	TTGCTGCTG	TTG			

【図 3 N】

48701 TTACANTAAA GGTCTTCCCA AATACACCAA GAGGCTCTT TCCGAATAG
48751 CAAGTGTCTT CAAGTTACTG TACCAGATAT TCTCACTAAT ACAGTAGTGT
48801 ATCTAAATGG GAGAGGTGGG GAGATATCAA CTGCTTAATA TTAGTCTAGT
48851 AATTTGGA TTAGTCTCTT AGTTTAAGA CATTGTGAT TGGCAATAGT
48901 TTCAAGATAC TGGCTTGGTA AAAAATAAAA AAAATATAGC AAGAATCTCT
48951 TCTTACGTG GAGCTCGATA ATGTATTTAC ATCAATATCT TGAATGTGTT
49001 CTCTCTCTGT GTATTAAGGT AGTTGAATAT AATTTAAAT TATCTACGGA
49051 GCATATGATC TTTATAAATA TTGATCTAGA AGACTCTTGA GAGATGCTGA
49101 CCAAGTATGG TTTTCAATA AGAAACAAA AGAGGCGCTT ATGATCTTAG
49151 GTGATTTTA AAAAATTTT TCCCGAGGA AGATTTCTTA AACCTTGGC
49201 TTACAGGGA GAGAAGAAA TTGTACTCTC TACAGATAT TTTTCTTGGC
49251 TAGTTTATTA AGATTTGAAT GACCCGAGGT TGTGATGATA TAAATAAATA
49301 TGAATATCA TGAATAGATA GATTTTAGAG TACTACAGG TTTTAAAGGT
49351 CCACTGATAC AGATTTAATC GATATAAATA ACTACAATAT GAGAAAGATC
49401 AGATGATAT AATCATTTAG GAGCAGCTGA AGTTTAAATA ACTCTGAGC
49451 TGAATATCA AATCATTTAT GATTTTAGAG TACTACAGG TTTTAAAGGT
49501 TTTAGTATA TGAATTTAT TTGATCTATC ACTAGCAAT TTAATGCTCT
49551 CAAGGAATAT AGCTATGGA GCTTAAATTC ATTTCAGTA TTTAAGGTGA
49601 ATTAGATAT TTTTCTTGA AATCATTTAT AATTAATC TCAAGAGGTA
49651 ATCATTTTTC CTATGATTT TCAAGGTGTA AATTTAAAG TTTTACCAAT
49701 GAAATGATA AATGATATA GTATTTCTTG CTTCATATCT ATGAGATTAT
49751 GAGAGGTTT TATTTAGAT AGATTTTATA TATGATGGA ACTCTGCTCA
49801 TAAGGATCAA CTCTTTCTCT AGACATGAAT TTCTAAGAAA TACCTTAATA
49851 AATTTCTAT TATTTTATA AGCCGATGT TCAACAGCA AAAATATCA
49901 TTTAGTATCA ATTTTATCA ATCCGATGT AGATTTTGT TGTATATTA
49951 TAACTGCGA TAAAGGATC ACTTAAAGTA TCCA

特徴:

転写率 1:

開始:

ヌクレオチド 3000

イントロン 3080-3083

ヌクレオチド 3084-14681

イントロン 14682-14777

ヌクレオチド 14778-23400

イントロン 23401-23503

ヌクレオチド 23504-24652

イントロン 24653-24769

ヌクレオチド 24770-26609

イントロン 26610-26724

ヌクレオチド 26725-32754

イントロン 32755-32973

ヌクレオチド 32974-33250

イントロン 33251-33364

ヌクレオチド 33365-34391

イントロン 34392-34463

ヌクレオチド 34464-41941

イントロン 41942-42054

ヌクレオチド 42055-43227

イントロン 43228-43326

ヌクレオチド 43326-45272

イントロン 45273-45436

ヌクレオチド 45437-46779

イントロン 46780-46983

停止: 46984

転写率 2:

開始:

ヌクレオチド 3000

イントロン 3084 3083

ヌクレオチド 23441 23503

イントロン 23504 24652

ヌクレオチド 24653 24769

イントロン 24774 46984

ヌクレオチド 24773 24780

イントロン 24781 26609

ヌクレオチド 26610 26724

イントロン 26725 32754

ヌクレオチド 32755 32973

イントロン 32974 33250

ヌクレオチド 33251 33364

イントロン 33365 34391

ヌクレオチド 34392 34463

【図 3 P】

同位体変異体 (SNP):

DNA 位置	上鎖塩基	下鎖塩基	ドメイン	タンパク質の 位置	上鎖 アミノ酸	下鎖 アミノ酸
126	G	T	C	087の前の配列(5')		
113	G	A	C	087の前の配列(4)		
1301	C	T	C	087の前の配列(3)		
2775	T	C	C	087の前の配列(2)		
4917	A	T		イントロン		
5406	T	G		イントロン		
5626	C	T	G	イントロン		
7291	G	A		イントロン		
10581	A	-		イントロン		
10609	T	-		イントロン		
12495	C	T	G	イントロン		
12571	C	T		イントロン		
14024	G	C		イントロン		
14493	T	A		イントロン		
15137	A	G	C	イントロン		
16174	G	A		イントロン		
17112	C	T	A	イントロン		
17204	-	T		イントロン		
18636	G	C	T	A	イントロン	
18639	-	T		イントロン		
18639	G	A		イントロン		
21578	G	A	C	イントロン		
22755	G	A		イントロン		
27705	-	A		イントロン		
28039	A	-		イントロン		
28045	C	-		イントロン		
28909	G	A		イントロン		
28912	C	G		イントロン		
28980	T	T		イントロン		
29187	G	T		イントロン		
31491	A	G		イントロン		
31589	G	A	C	イントロン		
32608	A	G		イントロン		
36093	-	G	A	イントロン		
37842	A	T		イントロン		
41941	G	A		イントロン		
41983	G	A		ヌクレオチド		
42047	G	A		ヌクレオチド		
43183	A	G		イントロン		
43785	C	G		イントロン		
45812	A	G	T	イントロン		
47215	G	A	T	087の前の配列(2')		

【図 3 O】

イントロン 34464 41941

ヌクレオチド 41942 42054

イントロン 42055 43227

ヌクレオチド 43228 43326

イントロン 43326 45272

ヌクレオチド 45273 45436

イントロン 45437 46779

ヌクレオチド 46780 46983

停止: 46984

染色体マップ位置:

第1染色体

【図 3 Q】

シグナルペプチド

DNA 位置	AGCTACATGATATTTTACTTGTATTTTGTGTTTGTGGGATTTAAATAATTTT ATCTGAGATATGTTGAATCCAGAGATCTAGGCGAGCTGATTCGACCCAAATC AGATA (5, 6)
113	AGCTACATGATATTTTACTTGTATTTTGTGTTTGTGGGATTTAAATAATTTT ATCTGAGATATGTTGAATCCAGAGATCTAGGCGAGCTGATTCGAC (5, 6)
1901	AGCTACATGATATTTTACTTGTATTTTGTGTTTGTGGGATTTAAATAATTTT ATCTGAGATATGTTGAATCCAGAGATCTAGGCGAGCTGATTCGAC (5, 6)
2775	AGCTACATGATATTTTACTTGTATTTTGTGTTTGTGGGATTTAAATAATTTT ATCTGAGATATGTTGAATCCAGAGATCTAGGCGAGCTGATTCGAC (5, 6)
4917	AGCTACATGATATTTTACTTGTATTTTGTGTTTGTGGGATTTAAATAATTTT ATCTGAGATATGTTGAATCCAGAGATCTAGGCGAGCTGATTCGAC (5, 6)
5406	AGCTACATGATATTTTACTTGTATTTTGTGTTTGTGGGATTTAAATAATTTT ATCTGAGATATGTTGAATCCAGAGATCTAGGCGAGCTGATTCGAC (5, 6)
5626	AGCTACATGATATTTTACTTGTATTTTGTGTTTGTGGGATTTAAATAATTTT ATCTGAGATATGTTGAATCCAGAGATCTAGGCGAGCTGATTCGAC (5, 6)

【 図 3 W 】

[illegible]

TAACTCTCAATGCTGTGTAGAGAAATATACCCAGCAATTTACAGTTTGGGACATGTAT
TCCCTATTTCATCATACACCCTCAGACAGCTGTGGAAATGATGATGATGATGATGAT
CAGGGCTCAGATGTTAGGACACTGCCCACTGCTACAGTATAGATGATTCAGCTTCT
TTTCTTCACAGCACTTCACAGATAGATTCAATATGATGTTTGCATCTGCCTCAT
[A,G]
GTGGGAGAAATTTACTGCTTTTGTGTTCTGCTCTCTCTATAGATATGTGGGCTGCT
GGGCGACATGACAGCACTGCTACAGTACACCTTCTCTGCTGACATGCTGCTTCTCAGGAGTAT
ACTGCTGATGCTGATCAGCAGGCGGCTGATGAGAACTATGCGATTAATTTAAAGATGCT
CAGAAATGCTTCAATGCTGTTCTGCTCTCTATCATGATGAACATTAAGTATGCTTCAGGAGTAT
TGAGACATTTTAAATTCGTTTGATATATACAGGAAAGACATCTAGGATATTAT

43785
 CTTAGGTTATCATATATTCCTTCTGCTGACATAGATATGTGGCTATGCTCTTCGTT
 CATAGAAAATTTATGCTGAGGTGGGGCATGCTCCAGCTCTGACACCTGCTGACCAACT
 GAGAGGCGGAGCGAGGACAGATGCTTGAAGTTGCTGGAGACAGCTGCGGCACAT
 GGGGAGACCCCATCATTAATGCAAAAATTAGCTGGGAGGAGTGGGCGATCTCTGAT
 CCGAGTTCTTGGGAGCGTAGGTTGGGAGGTACCTTAGGTTCTGAGGCGGAGGAGTGG
 [5,6]
 GCGGCAAAATGAGCGGCGCTGACGTCGCAACTGGGCGACGACGAGAGACCTGTGAGAG
 CTTCTCTCAAAAATAATTTAGTATGAATGTGGCCAGCAAGATTAATCAAGCTTTAAGT
 TTTTCTGCTGTGTAGACACCTCTGGGGATATCTATCAGAGCGTTTATGAAATGTT
 TCCGCAAGCTTGGGAGAGGAGGAGGAGGAGGATTTGAGCTCAGAGTTTGAAGCTGCT
 CTTAGGACACACACAGCATGCCCATCTCTATAAAAAAAGAGAAATCAATATGATTT

45812 AGCTTTTATGTGAGATTCTATTTGAGATTGGATAAAATGCTTAAAAACAGAGTTTAAAGG
GACTGGTGTCTCTCATACACCCTTGTGAAAATGGCGTGATATTTTTCCTCCAT
GAGATAGAAAAAATTTAAGTACTAGTTTGTCAATGGCTGATATTAACAAACCTTTCTCT
TTACTACTAGCTCTCTTTTGAACATTTCTCAATATCACAGCTGTGATAGTTAACTAACTTT
TCATCAAAAGGACGGCTTTATGGTAAATAACCTTGTTCACTGTTCGTATTTCCCTC
(A,G,T)
TTCACTCTGAAAATTAATGGCAACACATCTGATATTTTGGCGAAATATTAGCAGCTCTT

CAGTTTCCTGGCTTATTATGACAGAGTTCGTACTTATCTCCAAACCCACTAGAGACTACIT
TAAATAAAAGAAATGTTTGGGGTATACACAAAGCTGGCAGTGAACGCTTCTGCTAAGA
ATGAATGTAAATTTTGGGCTAAACAAAGGCTCTTTTCTGGTAACCTGFTTTTCTCT
ACTCCAGTCACCTTGAATTATAATTACCTTCTGAAACAAGAAATGGTAATTATT

47215 CACAGCCATCCTCATATACACAAAGCGCCAGGAGAGGCCAAAGAACCCTTTACTCCAGGAC
ACAAATGTGTGACGACTGAAATCAGGAAGATTTTCTATCAGCAGCCAGGCTTGATGTTT
CAGCTGTAGTTCTGGATGTACATTCATTTCCCTACCAAGTGTACTTTAAGATTGCTTAT
GGAAATCTATCTCGACTGAACCTCCGTGGGAATCCAGGAAGTGGAACTTAAGACCCAGACN

GTTTTCAAAAGATGTTACATTTCTTTTGA AAAACCTTTTGTTTATAGCACCATTTCT
 [G, A, T]
 GCCACTAAGCTATTGGTTTATATACATCTTTTAATAAAAACATATATGTAACTTCT
 TAGATATTAGCAAAAGCTCTGCTACCATTTCTTAAGTGTTGAGCTTTAACTCTATGG
 TGACTAGTGAGACACAGTAGTAGTATGGTTGTGGACCTATTGTTTAAACATTGTAA
 ATTATTTAGTCAGATATTGAATATTGTA AAAATCTTTGGGTCAAAATCTCAAAAGCTTAATGG
 AGATGCACTAAAACAAAGAAATGGTAAATGAATGTTTCTGTTTAAAAAAAACCACTT

【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 02/48366 A2(51) International Patent Classification: C12N 15/12,
15/11, 15/63, 1/21, 5/10, C07K 14/47, 14/705, 16/18,
16/28, C12Q 1/68, G01N 33/68, A01K 67/027

(21) International Application Number: PCT/US01/47559

(22) International Filing Date:
11 December 2001 (11.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/254,553 12 December 2000 (12.12.2000) US
09/739,457 19 December 2000 (19.12.2000) US(71) Applicant: PE CORPORATION (NY) [US/US]; 761
Main Avenue, Norwalk, CT 06859 (US).(72) Inventors: MERKULOV, Gennady; c/o Celera Ge-
nomics, 45 West Gude Drive C2-4#21, Rockville, MD
20850 (US); GUEGLER, Karl; c/o Celera Genomics,
45 West Gude Drive C2-4#21, Rockville, MD 20850
(US); WEY, Ming-Hui; c/o Celera Genomics, 45 West
Gude Drive C2-4#21, Rockville, MD 20850 (US); DI
FRANCESCO, Valentina; c/o Celera Genomics, 45
West Gude Drive C2-4#21, Rockville, MD 20850 (US);
BEASLEY, Eden, M.; c/o Celera Genomics, 45 West
Gude Drive C2-4#21, Rockville, MD 20850 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MY, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF

Transcript 1:
(SEQ ID NO:1)
1 AGGACGCGG GGAAGAGAA GAAAGGCGC GCGAACCGA GATACAGCT
51 GCGACAGAG ATCATATTA AGAGAGGCG GAGGCGTGA GCGATAGCT
101 GCGGAGGCT GATGATGCA GCGGCGGCG GCGGCGGCG GCGGCGGCG
151 GCGGAGGCT TCGAGAGAG CAGCTGCGT GATGCGGCG GCGGCGGCG
201 TCGAGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
251 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
301 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
351 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
401 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
451 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
501 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
551 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
601 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
651 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
701 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
751 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
801 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
851 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
901 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
951 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1001 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1051 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1101 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1151 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1201 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1251 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1301 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1351 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1401 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1451 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT
1501 TCGGAGAG TCGGAGAG CAGCTGCGT TCGGAGAG CAGCTGCGT

Transcript 1 Features:
Start Codon: 1
Stop Codon: 1471
Transcript 1 Top 10 BLAST Hits:
g1|6889221|celfp_02272.1| hippocampus abundant gene transcript... 377 0.0
g1|7283251|h1ANF1731.1| (A000547) cell1379 gene product (alt... 557 e=-157
g1|7504440|p1e1|73370 hypothetical protein T2503.1 - Caenorhab... 474 e=-132
g1|6889281|h1ANF1731.1| (A000547) cell1379 gene product (alt... 388 e=-136
g1|7283271|h1ANF1613.1| (A000547) cell1379 gene product (alt... 382 e=-105
g1|7283251|h1ANF1613.1| (A000547) cell1379 gene product (alt... 163 e=-45
g1|7283251|h1ANF1613.1| (A000547) cell1379 gene product (alt... 163 e=-45
g1|7773927|p1a1|175512 tetrahymena cell1379 gene product (alt... 139 e=-30
g1|7504440|p1e1|73370 hypothetical protein T2503.1 - Caenorhab... 132 e=-30
g1|7504440|p1e1|73370 hypothetical protein T2503.1 - Caenorhab... 132 e=-30

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the transporter peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the transporter peptides, and methods of identifying modulators of the transporter peptides.

WO 02/48366 A2

WO 02/48366 A2



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

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

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/48366

PCT/US01/47559

**ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES
ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF**

RELATED APPLICATIONS

The present application claims priority to provisional application U.S. Serial No. 60/254,553, filed December 12, 2000 (Atty. Docket CL001014-PROV) and U.S. Serial No. 09/739,457, filed December 19, 2000 (Atty. Docket CL001014).

FIELD OF THE INVENTION

The present invention is in the field of transporter proteins that are related to the sugar transporter subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel peptides and proteins that effect ligand transport and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

Transporters

Transporter proteins regulate many different functions of a cell, including cell proliferation, differentiation, and signaling processes, by regulating the flow of molecules such as ions and macromolecules, into and out of cells. Transporters are found in the plasma membranes of virtually every cell in eukaryotic organisms. Transporters mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of molecules and ion across cell membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, transporters, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) *Annu. Rev. Physiol.* 50:111-122.

Transporters are generally classified by structure and the type of mode of action. In addition, transporters are sometimes classified by the molecule type that is transported, for example, sugar transporters, chlorine channels, potassium channels, etc. There may be many classes of channels for transporting a single type of molecule (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters: *Receptor and transporter nomenclature*

WO 02/48366

PCT/US01/47559

supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 (1997) and <http://www-biology.ucsd.edu/~msaier/transport/titlepage2.html>.

The following general classification scheme is known in the art and is followed in the present discoveries.

Channel-type transporters. Transmembrane channel proteins of this class are ubiquitously found in the membranes of all types of organisms from bacteria to higher eukaryotes. Transport systems of this type catalyze facilitated diffusion (by an energy-independent process) by passage through a transmembrane aqueous pore or channel without evidence for a carrier-mediated mechanism. These channel proteins usually consist largely of α -helical spanners, although β -strands may also be present and may even comprise the channel. However, outer membrane porin-type channel proteins are excluded from this class and are instead included in class 9.

Carrier-type transporters. Transport systems are included in this class if they utilize a carrier-mediated process to catalyze uniport (a single species is transported by facilitated diffusion), antiport (two or more species are transported in opposite directions in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy) and/or symport (two or more species are transported together in the same direction in a tightly coupled process, not coupled to a direct form of energy other than chemiosmotic energy).

Pyrophosphate bond hydrolysis-driven active transporters. Transport systems are included in this class if they hydrolyze pyrophosphate or the terminal pyrophosphate bond in ATP or another nucleoside triphosphate to drive the active uptake and/or extrusion of a solute or solutes. The transport protein may or may not be transiently phosphorylated, but the substrate is not phosphorylated.

PEP-dependent, phosphoryl transfer-driven group translocators. Transport systems of the bacterial phosphoenolpyruvate:sugar phosphotransferase system are included in this class. The product of the reaction, derived from extracellular sugar, is a cytoplasmic sugar-phosphate.

Decarboxylation-driven active transporters. Transport systems that drive solute (e.g., ion) uptake or extrusion by decarboxylation of a cytoplasmic substrate are included in this class.

Oxidoreduction-driven active transporters. Transport systems that drive transport of a solute (e.g., an ion) energized by the flow of electrons from a reduced substrate to an oxidized substrate are included in this class.

Light-driven active transporters. Transport systems that utilize light energy to drive transport of a solute (e.g., an ion) are included in this class.

WO 02/48366

PCT/US01/47559

Mechanically-driven active transporters. Transport systems are included in this class if they drive movement of a cell or organelle by allowing the flow of ions (or other solutes) through the membrane down their electrochemical gradients.

Outer-membrane porins (of β -structure). These proteins form transmembrane pores or channels that usually allow the energy independent passage of solutes across a membrane. The transmembrane portions of these proteins consist exclusively of β -strands that form a β -barrel. These porin-type proteins are found in the outer membranes of Gram-negative bacteria, mitochondria and eukaryotic plastids.

Methyltransferase-driven active transporters. A single characterized protein currently falls into this category, the Na^+ -transporting methyltetrahydromethanopterin:coenzyme M methyltransferase.

Non-ribosome-synthesized channel-forming peptides or peptide-like molecules. These molecules, usually chains of L- and D-amino acids as well as other small molecular building blocks such as lactate, form oligomeric transmembrane ion channels. Voltage may induce channel formation by promoting assembly of the transmembrane channel. These peptides are often made by bacteria and fungi as agents of biological warfare.

Non-Proteinaceous Transport Complexes. Ion conducting substances in biological membranes that do not consist of or are not derived from proteins or peptides fall into this category.

Functionally characterized transporters for which sequence data are lacking. Transporters of particular physiological significance will be included in this category even though a family assignment cannot be made.

Putative transporters in which no family member is an established transporter. Putative transport protein families are grouped under this number and will either be classified elsewhere when the transport function of a member becomes established, or will be eliminated from the TC classification system if the proposed transport function is disproven. These families include a member or members for which a transport function has been suggested, but evidence for such a function is not yet compelling.

Auxiliary transport proteins. Proteins that in some way facilitate transport across one or more biological membranes but do not themselves participate directly in transport are included in this class. These proteins always function in conjunction with one or more transport proteins. They may provide a function connected with energy coupling to transport, play a structural role in complex formation or serve a regulatory function.

WO 02/48366

PCT/US01/47559

Transporters of unknown classification. Transport protein families of unknown classification are grouped under this number and will be classified elsewhere when the transport process and energy coupling mechanism are characterized. These families include at least one member for which a transport function has been established, but either the mode of transport or the energy coupling mechanism is not known.

Ion channels

An important type of transporter is the ion channel. Ion channels regulate many different cell proliferation, differentiation, and signaling processes by regulating the flow of ions into and out of cells. Ion channels are found in the plasma membranes of virtually every cell in eukaryotic organisms. Ion channels mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of ion across epithelial membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, ion channels, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) *Annu. Rev. Physiol.* 50:111-122.

Ion channels are generally classified by structure and the type of mode of action. For example, extracellular ligand gated channels (ELGs) are comprised of five polypeptide subunits, with each subunit having 4 membrane spanning domains, and are activated by the binding of an extracellular ligand to the channel. In addition, channels are sometimes classified by the ion type that is transported, for example, chlorine channels, potassium channels, etc. There may be many classes of channels for transporting a single type of ion (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters (1997). *Receptor and ion channel nomenclature supplement. Trends Pharmacol. Sci., Elsevier*, pp. 65-68 and <http://www-biology.ucsd.edu/~msaier/transport/toc.html>.

There are many types of ion channels based on structure. For example, many ion channels fall within one of the following groups: extracellular ligand-gated channels (ELG), intracellular ligand-gated channels (ILG), inward rectifying channels (INR), intercellular (gap junction) channels, and voltage gated channels (VIC). There are additionally recognized other channel families based on ion-type transported, cellular location and drug sensitivity. Detailed information on each of these, their activity, ligand type, ion type, disease association, drugability, and other information pertinent to the present invention, is well known in the art.

Extracellular ligand-gated channels, ELGs, are generally comprised of five polypeptide subunits, Unwin, N. (1993), *Cell* 72: 31-41; Unwin, N. (1995), *Nature* 373: 37-43; Hucho, F., et

WO 02/48366

PCT/US01/47559

al., (1996) J. Neurochem. 66: 1781-1792; Hucho, F., et al., (1996) Eur. J. Biochem. 239: 539-557; Alexander, S.P.H. and J.A. Peters (1997), Trends Pharmacol. Sci., Elsevier, pp. 4-6; 36-40; 42-44; and Xue, H. (1998) J. Mol. Evol. 47: 323-333. Each subunit has 4 membrane spanning regions: this serves as a means of identifying other members of the ELG family of proteins. ELG bind a ligand and in response modulate the flow of ions. Examples of ELG include most members of the neurotransmitter-receptor family of proteins, e.g., GABA_A receptors. Other members of this family of ion channels include glycine receptors, ryanodine receptors, and ligand gated calcium channels.

The Voltage-gated Ion Channel (VIC) Superfamily

Proteins of the VIC family are ion-selective channel proteins found in a wide range of bacteria, archaea and eukaryotes Hille, B. (1992), Chapter 9: Structure of channel proteins; Chapter 20: Evolution and diversity. In: Ionic Channels of Excitable Membranes, 2nd Ed., Sinaur Assoc. Inc., Pubs., Sunderland, Massachusetts; Sigworth, F.J. (1993), Quart. Rev. Biophys. 27: 1-40; Salkoff, L. and T. Jegla (1995), Neuron 15: 489-492; Alexander, S.P.H. et al., (1997), Trends Pharmacol. Sci., Elsevier, pp. 76-84; Jan, L.Y. et al., (1997), Annu. Rev. Neurosci. 20: 91-123; Doyle, D.A., et al., (1998) Science 280: 69-77; Terlau, H. and W. Stühmer (1998), Naturwissenschaften 85: 437-444. They are often homo- or heterooligomeric structures with several dissimilar subunits (e.g., $\alpha_1\alpha_2\gamma\delta$ Ca^{2+} channels, $\alpha_1\beta_2$ Na^+ channels or $(\alpha)_4\beta$ K^+ channels), but the channel and the primary receptor is usually associated with the α (or α_1) subunit. Functionally characterized members are specific for K^+ , Na^+ or Ca^{2+} . The K^+ channels usually consist of homotetrameric structures with each α -subunit possessing six transmembrane spanners (TMSs). The α_1 and α subunits of the Ca^{2+} and Na^+ channels, respectively, are about four times as large and possess 4 units, each with 6 TMSs separated by a hydrophilic loop, for a total of 24 TMSs. These large channel proteins form heterotetra-unit structures equivalent to the homotetrameric structures of most K^+ channels. All four units of the Ca^{2+} and Na^+ channels are homologous to the single unit in the homotetrameric K^+ channels. Ion flux via the eukaryotic channels is generally controlled by the transmembrane electrical potential (hence the designation, voltage-sensitive) although some are controlled by ligand or receptor binding.

Several putative K^+ -selective channel proteins of the VIC family have been identified in prokaryotes. The structure of one of them, the KcsA K^+ channel of *Streptomyces lividans*, has been solved to 3.2 Å resolution. The protein possesses four identical subunits, each with two transmembrane helices, arranged in the shape of an inverted teepee or cone. The cone cradles the "selectivity filter" P domain in its outer end. The narrow selectivity filter is only 12 Å long.

WO 02/48366

PCT/US01/47559

whereas the remainder of the channel is wider and lined with hydrophobic residues. A large water-filled cavity and helix dipoles stabilize K^+ in the pore. The selectivity filter has two bound K^+ ions about 7.5 Å apart from each other. Ion conduction is proposed to result from a balance of electrostatic attractive and repulsive forces.

In eukaryotes, each VIC family channel type has several subtypes based on pharmacological and electrophysiological data. Thus, there are five types of Ca^{2+} channels (L, N, P, Q and T). There are at least ten types of K^+ channels, each responding in different ways to different stimuli: voltage-sensitive [K_a , K_v , K_{vr} , K_{vs} and K_{sr}], Ca^{2+} -sensitive [BK_{Ca} , IK_{Ca} and SK_{Ca}] and receptor-coupled [K_M and K_{ACH}]. There are at least six types of Na^+ channels (I, II, III, $\mu 1$, H1 and PN3). Tetrameric channels from both prokaryotic and eukaryotic organisms are known in which each α -subunit possesses 2 TMSs rather than 6, and these two TMSs are homologous to TMSs 5 and 6 of the six TMS unit found in the voltage-sensitive channel proteins. $KcsA$ of *S. lividans* is an example of such a 2 TMS channel protein. These channels may include the K_{Na} (Na^+ -activated) and K_{Vol} (cell volume-sensitive) K^+ channels, as well as distantly related channels such as the Tok1 K^+ channel of yeast, the TWIK-1 inward rectifier K^+ channel of the mouse and the TREK-1 K^+ channel of the mouse. Because of insufficient sequence similarity with proteins of the VIC family, inward rectifier K^+ IRK channels (ATP-regulated; G-protein-activated) which possess a P domain and two flanking TMSs are placed in a distinct family. However, substantial sequence similarity in the P region suggests that they are homologous. The b, g and d subunits of VIC family members, when present, frequently play regulatory roles in channel activation/deactivation.

The Epithelial Na^+ Channel (ENaC) Family

The ENaC family consists of over twenty-four sequenced proteins (Canessa, C.M., et al., (1994), Nature 367: 463-467, Le, T. and M.H. Saier, Jr. (1996), Mol. Membr. Biol. 13: 149-157; Garty, H. and L.G. Palmer (1997), Physiol. Rev. 77: 359-396; Waldmann, R., et al., (1997), Nature 386: 173-177; Darboux, L., et al., (1998), J. Biol. Chem. 273: 9424-9429; Firsov, D., et al., (1998), EMBO J. 17: 344-352; Horisberger, J.-D. (1998), Curr. Opin. Struc. Biol. 10: 443-449). All are from animals with no recognizable homologues in other eukaryotes or bacteria. The vertebrate ENaC proteins from epithelial cells cluster tightly together on the phylogenetic tree: voltage-insensitive ENaC homologues are also found in the brain. Eleven sequenced *C. elegans* proteins, including the degenerins, are distantly related to the vertebrate proteins as well as to each other. At least some of these proteins form part of a mechano-transducing complex for

WO 02/48366

PCT/US01/47559

touch sensitivity. The homologous *Helix aspersa* (FMRF-amide)-activated Na^+ channel is the first peptide neurotransmitter-gated ionotropic receptor to be sequenced.

Protein members of this family all exhibit the same apparent topology, each with N- and C-termini on the inside of the cell, two amphipathic transmembrane spanning segments, and a large extracellular loop. The extracellular domains contain numerous highly conserved cysteine residues. They are proposed to serve a receptor function.

Mammalian ENaC is important for the maintenance of Na^+ balance and the regulation of blood pressure. Three homologous ENaC subunits, alpha, beta, and gamma, have been shown to assemble to form the highly Na^+ -selective channel. The stoichiometry of the three subunits is $\alpha_2\beta_1\gamma_1$ in a heterotetrameric architecture.

The Glutamate-gated Ion Channel (GIC) Family of Neurotransmitter Receptors

Members of the GIC family are heteropentameric complexes in which each of the 5 subunits is of 800-1000 amino acid residues in length (Nakanishi, N., et al, (1990), Neuron 5: 569-581; Unwin, N. (1993), Cell 72: 31-41; Alexander, S.P.H. and J.A. Peters (1997) Trends Pharmacol. Sci., Elsevier, pp. 36-40). These subunits may span the membrane three or five times as putative α -helices with the N-termini (the glutamate-binding domains) localized extracellularly and the C-termini localized cytoplasmically. They may be distantly related to the ligand-gated ion channels, and if so, they may possess substantial β -structure in their transmembrane regions. However, homology between these two families cannot be established on the basis of sequence comparisons alone. The subunits fall into six subfamilies: a, b, g, d, e and z.

The GIC channels are divided into three types: (1) α -amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA)-, (2) kainate- and (3) N-methyl-D-aspartate (NMDA)-selective glutamate receptors. Subunits of the AMPA and kainate classes exhibit 35-40% identity with each other while subunits of the NMDA receptors exhibit 22-24% identity with the former subunits. They possess large N-terminal, extracellular glutamate-binding domains that are homologous to the periplasmic glutamine and glutamate receptors of ABC-type uptake permeases of Gram-negative bacteria. All known members of the GIC family are from animals. The different channel (receptor) types exhibit distinct ion selectivities and conductance properties. The NMDA-selective large conductance channels are highly permeable to monovalent cations and Ca^{2+} . The AMPA- and kainate-selective ion channels are permeable primarily to monovalent cations with only low permeability to Ca^{2+} .

WO 02/48366

PCT/US01/47559

The Chloride Channel (ClC) Family

The ClC family is a large family consisting of dozens of sequenced proteins derived from Gram-negative and Gram-positive bacteria, cyanobacteria, archaea, yeast, plants and animals (Steinmeyer, K., et al., (1991), Nature 354: 301-304; Uchida, S., et al., (1993), J. Biol. Chem. 268: 3821-3824; Huang, M.-E., et al., (1994), J. Mol. Biol. 242: 595-598; Kawasaki, M., et al., (1994), Neuron 12: 597-604; Fisher, W.E., et al., (1995), Genomics. 29:598-606; and Foskett, J.K. (1998), Annu. Rev. Physiol. 60: 689-717). These proteins are essentially ubiquitous, although they are not encoded within genomes of *Haemophilus influenzae*, *Mycoplasma genitalium*, and *Mycoplasma pneumoniae*. Sequenced proteins vary in size from 395 amino acid residues (*M. jannaschii*) to 988 residues (man). Several organisms contain multiple ClC family paralogues. For example, *Synechocystis* has two paralogues, one of 451 residues in length and the other of 899 residues. *Arabidopsis thaliana* has at least four sequenced paralogues, (775-792 residues), humans also have at least five paralogues (820-988 residues), and *C. elegans* also has at least five (810-950 residues). There are nine known members in mammals, and mutations in three of the corresponding genes cause human diseases. *E. coli*, *Methanococcus jannaschii* and *Saccharomyces cerevisiae* only have one ClC family member each. With the exception of the larger *Synechocystis* paralogue, all bacterial proteins are small (395-492 residues) while all eukaryotic proteins are larger (687-988 residues). These proteins exhibit 10-12 putative transmembrane α -helical spanners (TMSs) and appear to be present in the membrane as homodimers. While one member of the family, *Torpedo* ClC-O, has been reported to have two channels, one per subunit, others are believed to have just one.

All functionally characterized members of the ClC family transport chloride, some in a voltage-regulated process. These channels serve a variety of physiological functions (cell volume regulation; membrane potential stabilization; signal transduction; transepithelial transport, etc.). Different homologues in humans exhibit differing anion selectivities, i.e., ClC4 and ClC5 share a $\text{NO}_3^- > \text{Cl}^- > \text{Br}^- > \text{I}^-$ conductance sequence, while ClC3 has an $\text{I}^- > \text{Cl}^-$ selectivity. The ClC4 and ClC5 channels and others exhibit outward rectifying currents with currents only at voltages more positive than +20mV.

Animal Inward Rectifier K^+ Channel (IRK-C) Family

IRK channels possess the "minimal channel-forming structure" with only a P domain, characteristic of the channel proteins of the VIC family, and two flanking transmembrane spanners (Shuck, M.E., et al., (1994), J. Biol. Chem. 269: 24261-24270; Ashen, M.D., et al.,

WO 02/48366

PCT/US01/47559

(1995), *Am. J. Physiol.* 268: H506-H511; Salkoff, L. and T. Jegla (1995), *Neuron* 15: 489-492; Aguilar-Bryan, L., et al., (1998), *Physiol. Rev.* 78: 227-245; Ruknudin, A., et al., (1998), *J. Biol. Chem.* 273: 14165-14171). They may exist in the membrane as homo- or heterooligomers. They have a greater tendency to let K^+ flow into the cell than out. Voltage-dependence may be regulated by external K^+ , by internal Mg^{2+} , by internal ATP and/or by G-proteins. The P domains of IRK channels exhibit limited sequence similarity to those of the VIC family, but this sequence similarity is insufficient to establish homology. Inward rectifiers play a role in setting cellular membrane potentials, and the closing of these channels upon depolarization permits the occurrence of long duration action potentials with a plateau phase. Inward rectifiers lack the intrinsic voltage sensing helices found in VIC family channels. In a few cases, those of Kir1.1a and Kir6.2, for example, direct interaction with a member of the ABC superfamily has been proposed to confer unique functional and regulatory properties to the heteromeric complex, including sensitivity to ATP. The SUR1 sulfonylurea receptor (spQ09428) is the ABC protein that regulates the Kir6.2 channel in response to ATP, and CFTR may regulate Kir1.1a. Mutations in SUR1 are the cause of familial persistent hyperinsulinemic hypoglycemia in infancy (PIHH), an autosomal recessive disorder characterized by unregulated insulin secretion in the pancreas.

ATP-gated Cation Channel (ACC) Family

Members of the ACC family (also called P2X receptors) respond to ATP, a functional neurotransmitter released by exocytosis from many types of neurons (North, R.A. (1996), *Curr. Opin. Cell Biol.* 8: 474-483; Soto, F., M. Garcia-Guzman and W. Stühmer (1997), *J. Membr. Biol.* 160: 91-100). They have been placed into seven groups ($P2X_1$ - $P2X_7$) based on their pharmacological properties. These channels, which function at neuron-neuron and neuron-smooth muscle junctions, may play roles in the control of blood pressure and pain sensation. They may also function in lymphocyte and platelet physiology. They are found only in animals.

The proteins of the ACC family are quite similar in sequence (>35% identity), but they possess 380-1000 amino acid residues per subunit with variability in length localized primarily to the C-terminal domains. They possess two transmembrane spanners, one about 30-50 residues from their N-termini, the other near residues 320-340. The extracellular receptor domains between these two spanners (of about 270 residues) are well conserved with numerous conserved glycyl and cysteyl residues. The hydrophilic C-termini vary in length from 25 to 240 residues. They resemble the topologically similar epithelial Na^+ channel (ENaC) proteins in possessing (a) N- and C-termini localized intracellularly, (b) two putative transmembrane spanners, (c) a large extracellular loop domain, and (d) many conserved extracellular cysteyl residues. ACC family

WO 02/48366

PCT/US01/47559

members are, however, not demonstrably homologous with them. ACC channels are probably hetero- or homomultimers and transport small monovalent cations (Me^+). Some also transport Ca^{2+} ; a few also transport small metabolites.

The Ryanodine-Inositol 1,4,5-triphosphate Receptor Ca^{2+} Channel (RIR-CaC) Family

Ryanodine (Ry)-sensitive and inositol 1,4,5-triphosphate (IP_3)-sensitive Ca^{2+} -release channels function in the release of Ca^{2+} from intracellular storage sites in animal cells and thereby regulate various Ca^{2+} -dependent physiological processes (Hasan, G. et al., (1992) Development 116: 967-975; Michikawa, T., et al., (1994), J. Biol. Chem. 269: 9184-9189; Tunwell, R.E.A., (1996), Biochem. J. 318: 477-487; Lee, A.G. (1996) *Biomembranes*, Vol. 6, Transmembrane Receptors and Channels (A.G. Lee, ed.), JAI Press, Denver, CO., pp 291-326; Mikoshiba, K., et al., (1996) J. Biochem. Biomem. 6: 273-289). Ry receptors occur primarily in muscle cell sarcoplasmic reticular (SR) membranes, and IP_3 receptors occur primarily in brain cell endoplasmic reticular (ER) membranes where they effect release of Ca^{2+} into the cytoplasm upon activation (opening) of the channel.

The Ry receptors are activated as a result of the activity of dihydropyridine-sensitive Ca^{2+} channels. The latter are members of the voltage-sensitive ion channel (VIC) family. Dihydropyridine-sensitive channels are present in the T-tubular systems of muscle tissues.

Ry receptors are homotetrameric complexes with each subunit exhibiting a molecular size of over 500,000 daltons (about 5,000 amino acyl residues). They possess C-terminal domains with six putative transmembrane α -helical spanners (TMSs). Putative pore-forming sequences occur between the fifth and sixth TMSs as suggested for members of the VIC family. The large N-terminal hydrophilic domains and the small C-terminal hydrophilic domains are localized to the cytoplasm. Low resolution 3-dimensional structural data are available. Mammals possess at least three isoforms that probably arose by gene duplication and divergence before divergence of the mammalian species. Homologues are present in humans and *Caenorabditis elegans*.

IP_3 receptors resemble Ry receptors in many respects. (1) They are homotetrameric complexes with each subunit exhibiting a molecular size of over 300,000 daltons (about 2,700 amino acyl residues). (2) They possess C-terminal channel domains that are homologous to those of the Ry receptors. (3) The channel domains possess six putative TMSs and a putative channel lining region between TMSs 5 and 6. (4) Both the large N-terminal domains and the smaller C-terminal tails face the cytoplasm. (5) They possess covalently linked carbohydrate on extracytoplasmic loops of the channel domains. (6) They have three currently recognized

WO 02/48366

PCT/US01/47559

isoforms (types 1, 2, and 3) in mammals which are subject to differential regulation and have different tissue distributions.

IP₃ receptors possess three domains: N-terminal IP₃-binding domains, central coupling or regulatory domains and C-terminal channel domains. Channels are activated by IP₃ binding, and like the Ry receptors, the activities of the IP₃ receptor channels are regulated by phosphorylation of the regulatory domains, catalyzed by various protein kinases. They predominate in the endoplasmic reticular membranes of various cell types in the brain but have also been found in the plasma membranes of some nerve cells derived from a variety of tissues.

The channel domains of the Ry and IP₃ receptors comprise a coherent family that in spite of apparent structural similarities, do not show appreciable sequence similarity of the proteins of the VIC family. The Ry receptors and the IP₃ receptors cluster separately on the RIR-CaC family tree. They both have homologues in *Drosophila*. Based on the phylogenetic tree for the family, the family probably evolved in the following sequence: (1) A gene duplication event occurred that gave rise to Ry and IP₃ receptors in invertebrates. (2) Vertebrates evolved from invertebrates. (3) The three isoforms of each receptor arose as a result of two distinct gene duplication events. (4) These isoforms were transmitted to mammals before divergence of the mammalian species.

The Organellar Chloride Channel (O-ClC) Family

Proteins of the O-ClC family are voltage-sensitive chloride channels found in intracellular membranes but not the plasma membranes of animal cells (Landry, D. et al., (1993), J. Biol. Chem. 268: 14948-14955; Valenzuela, Set al., (1997), J. Biol. Chem. 272: 12575-12582; and Duncan, R.R., et al., (1997), J. Biol. Chem. 272: 23880-23886).

They are found in human nuclear membranes, and the bovine protein targets to the microsomes, but not the plasma membrane, when expressed in *Xenopus laevis* oocytes. These proteins are thought to function in the regulation of the membrane potential and in transepithelial ion absorption and secretion in the kidney. They possess two putative transmembrane α -helical spanners (TMSs) with cytoplasmic N- and C-termini and a large luminal loop that may be glycosylated. The bovine protein is 437 amino acid residues in length and has the two putative TMSs at positions 223-239 and 367-385. The human nuclear protein is much smaller (241 residues). A *C. elegans* homologue is 260 residues long.

Sugar transporter

Organic substrates (sugars, amino acids, carboxylic acids and neurotransmitters) are actively transported into eukaryotic cells by Na⁺ co-transport. Some of the transport proteins

WO 02/48366

PCT/US01/47559

have been identified--for example, intestinal brush border Na⁺/glucose and Na⁺/proline transporters and the brain Na⁺/Cl⁻/GABA transporter--and progress has been made in locating their active sites and probing their conformational states. The archetypical Na⁺-driven transporter is the intestinal brush border Na⁺/glucose co-transporter, and a defect in the co-transporter is the origin of the congenital glucose-galactose malabsorption syndrome.

Cotransporters are a major class of membrane proteins -- typically with 13 membrane spanning helices. They cause the concentration of molecules across a membrane -- nutrients, neurotransmitters, osmolytes and ions. For example there are co transporters for amino acids, sugars, nucleosides and vitamins.

Na⁺/glucose co transporter (SGLT1) was reported in 1960 by Bob Crane. Sodium dependent glucose transport occurs in both the kidney and the intestine of animals. Both of these transporters show a close similarity to each other.

These transporters are reported to be multifunctional and have been shown to operate in 4 ways: 1) Uncoupled passive Na⁺ transport, 2) Downhill water transport, 3) Na⁺ and substrate transport, 4) Na⁺, water and substrate transport. For further information regarding to the present invention, see Matsuo et al., *Biochem Biophys Res Commun* 1997 Sep 8;238(1):126-9.

Hexose transport into mammalian cells is catalyzed by members of a small family of 44- to 55-kD membrane proteins that have specific functions and differ in their tissue distribution. Observed hexose transporters have 12 membrane-spanning helices and a number of critical conserved residues. By EST database searching for clones containing conserved GLUT sequences, followed by screening of rat tissues and 5-prime RACE, Ibberson et al. and Doege et al. identified rodent and human cDNAs encoding a novel glucose transporter. (Ibberson, M., et al., *J. Biol. Chem.* 275: 4607-4612, 2000, PubMed ID : 10671487; and Doege, H., et al., *J. Biol. Chem.* 275: 16275-16280, 2000, PubMed ID : 10821868) The human cDNA encodes a deduced 477-amino acid protein, designated GLUT8 or GLUTX1, that shares 85% sequence homology with the mouse sequence. Ibberson et al. found that the approximately 37-kD rat Glutx1 expressed in frog oocytes is unable to take up glucose unless the N-terminal dileucine motif, which may serve as an internalization signal, is mutated to alanines. Immunofluorescence analysis demonstrated that Glutx1 is expressed intracellularly, whereas Glutx1(LL-AA) is expressed on the plasma membrane. In apparent contrast, Doege et al. found that membrane preparations from cells expressing GLUT8 cannot bind cytochalasin B in the presence of glucose and, when reconstituted in liposomes, have increased D-glucose transport activity. By Western blot analysis, Doege et al. determined that human GLUT8 is expressed as a 42-kD protein. Northern blot analysis revealed expression of a 2.4-

WO 02/48366

PCT/US01/47559

kb transcript, with strongest expression in testis and moderate expression in other tissues except thyroid. In addition, Doege et al. found that GLUT8 was not detectable in 2 patients with testicular carcinoma or in testicular tissue of 4 patients treated with estrogen. They found that Glut8 mRNA was detectable in testis from pubertal and adult, but not prepubertal, rats

Glucose transport activity in early preimplantation mouse embryos had been attributed to the known facilitative glucose transporters GLUT1 (SLC2A1; 138140), GLUT2 (SLC2A2; 138160), and GLUT3 (SLC2A3; 138170). GLUT1 is present throughout the preimplantation period, which begins with the 1-cell embryo and ends with the blastocyst stage. GLUT2 and GLUT3 are first expressed at a late 8-cell stage and remain present for the rest of the preimplantation period. The simultaneous appearance of all 3 transporters corresponds to the critical time in mammalian development when an embryonic fuel metabolism switches from the oxidation of lactate and pyruvate via the Krebs cycle and oxidative phosphorylation to anaerobic metabolism of glucose via glycolysis. Mammalian preimplantation blastocysts exhibit insulin-stimulated glucose uptake despite the absence of the only known insulin-regulated transporter, GLUT4 (SLC2A4; 138190). Carayannopoulos et al. found that mouse Glut8 exhibits 20 to 25% amino acid sequence identity with Glut1, Glut3, and Glut4. (Carayannopoulos, M. O., et al., Proc. Nat. Acad. Sci. 97: 7313-7318, 2000, PubMed ID : 10860996) Insulin induced a change in the intracellular localization of this protein, which translated into increased glucose uptake into the blastocyst, a process that was inhibited by antisense oligoprobes. The presence of this transporter may be necessary for successful blastocyst development, fuel metabolism, and subsequent implantation. The existence of an alternative transporter may explain examples in other tissues of insulin-regulated glucose transport in the absence of Glut4

Doege et al. noted that the International Radiation Hybrid Mapping Consortium localized the GLUT8 gene to chromosome 9 (A005N15).

Transporter proteins, particularly members of the sugar transporter subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown transport proteins. The present invention advances the state of the art by providing previously unidentified human transport proteins.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human transporter peptides and proteins that are related to the sugar transporter subfamily, as

WO 02/48366

PCT/US01/47559

well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate transporter activity in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of cDNA molecules or transcript sequences that encode the transporter proteins of the present invention. In addition structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain.

FIGURE 2 provides the predicted amino acid sequence of the transporter of the present invention. In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences that span the gene encoding the transporter protein of the present invention. In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As illustrated in Figure 3, SNPs, including insertion/deletion variants ("indels"), were identified at 42 different nucleotide positions.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share

WO 02/48366

PCT/US01/47559

structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a transporter protein or part of a transporter protein and are related to the sugar transporter subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human transporter peptides and proteins that are related to the sugar transporter subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these transporter peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the transporter of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known transporter proteins of the sugar transporter subfamily and the expression pattern observed. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known sugar transporter family or subfamily of transporter proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the transporter family of proteins and are related to the sugar transporter subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figures 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be

WO 02/48366

PCT/US01/47559

referred herein as the transporter peptides of the present invention, transporter peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprising the amino acid sequences of the transporter peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the transporter peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated transporter peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. For example, a nucleic acid molecule encoding the transporter peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be

WO 02/48366

PCT/US01/47559

isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:3 and SEQ ID NO:4), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1 and SEQ ID NO:2) and the genomic sequences provided in Figure 3 (SEQ ID NO:5). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:3 and SEQ ID NO:4), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1 and SEQ ID NO:2) and the genomic sequences provided in Figure 3 (SEQ ID NO:5). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:3 and SEQ ID NO:4), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1 and SEQ ID NO:2) and the genomic sequences provided in Figure 3 (SEQ ID NO:5). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the transporter peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The transporter peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a transporter peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the transporter peptide. "Operatively linked" indicates that the transporter peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the transporter peptide.

WO 02/48366

PCT/US01/47559

In some uses, the fusion protein does not affect the activity of the transporter peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant transporter peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A transporter peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the transporter peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the transporter peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal

WO 02/48366

PCT/US01/47559

alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and

WO 02/48366

PCT/US01/47559

XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the transporter peptides of the present invention as well as being encoded by the same genetic locus as the transporter peptide provided herein. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1.

Allelic variants of a transporter peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by the same genetic locus as the transporter peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 42 different nucleotide positions in introns and regions 5' and 3' of the ORF. Such SNPs in introns and outside the ORF may affect control/regulatory elements. Two SNPs in exons, of which 1 of these cause changes in the amino acid sequence (i.e., nonsynonymous SNPs). The changes in the amino acid sequence that these SNPs cause is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference.

WO 02/48366

PCT/US01/47559

Paralogs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the transporter peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the transporter peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a transporter peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant transporter peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to transport ligand, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

WO 02/48366

PCT/US01/47559

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as transporter activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

The present invention further provides fragments of the transporter peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a transporter peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the transporter peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the transporter peptide, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in transporter peptides are described in basic texts, detailed

WO 02/48366

PCT/US01/47559

monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, 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 crosslinks, 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.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

Accordingly, the transporter peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature transporter peptide is fused with another compound, such as a compound to increase the half-life of the transporter peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature transporter peptide, such as a leader or secretory sequence or a sequence for purification of the mature transporter peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed

WO 02/48366

PCT/US01/47559

(either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a transporter-effector protein interaction or transporter-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

Substantial chemical and structural homology exists between the sugar transporter protein described herein and sugar transporter expressed in the neonatal mouse hippocampus (see Figure 1). As discussed in the background, sugar transporter expressed in the neonatal mouse hippocampus are known in the art to be involved in sugar absorption. Accordingly, the sugar transporter protein, and the encoding gene, provided by the present invention is useful for treating, preventing, and/or diagnosing neuropsychiatric disorders, sugar malabsorption and other disorders associated with hippocampus sugar transporter.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, transporters isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the transporter. Experimental data as provided in Figure 1 indicates that the transporter protein of the present invention is expressed in the ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain by a virtual northern blot. In addition, PCR-based tissue screening panels indicate expression in kidney. A large percentage of pharmaceutical agents are being developed that modulate the activity of transporter proteins, particularly members of the sugar transporter subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue),

WO 02/48366

PCT/US01/47559

uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. Such uses can readily be determined using the information provided herein, that known in the art and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to transporters that are related to members of the sugar transporter subfamily. Such assays involve any of the known transporter functions or activities or properties useful for diagnosis and treatment of transporter-related conditions that are specific for the subfamily of transporters that the one of the present invention belongs to, particularly in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates that the transporter protein of the present invention is expressed in the ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain by a virtual northern blot. In addition, PCR-based tissue screening panels indicate expression in kidney. The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems ((Hodgson, *Bio/technology*, 1992, Sept 10(9);973-80). Cell-based systems can be native, i.e., cells that normally express the transporter, as a biopsy or expanded in cell culture. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the transporter protein.

The polypeptides can be used to identify compounds that modulate transporter activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the transporter. Both the transporters of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the transporter. These compounds can be further screened against a functional transporter to determine the effect of the compound on the transporter activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the transporter to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the transporter protein and a molecule that normally interacts with the transporter protein, e.g. a substrate or a component of the signal pathway that the transporter protein normally interacts (for example, another transporter). Such assays

WO 02/48366

PCT/US01/47559

typically include the steps of combining the transporter protein with a candidate compound under conditions that allow the transporter protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the transporter protein and the target, such as any of the associated effects of signal transduction such as changes in membrane potential, protein phosphorylation, cAMP turnover, and adenylate cyclase activation, etc.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant transporters or appropriate fragments containing mutations that affect transporter function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) transporter activity. The assays typically involve an assay of events in the signal transduction pathway that indicate transporter activity. Thus, the transport of a ligand, change in cell membrane potential, activation of a protein, a change in the expression of genes that are up- or down-regulated in response to the transporter protein dependent signal cascade can be assayed.

Any of the biological or biochemical functions mediated by the transporter can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the transporter can be assayed. Experimental data as provided in Figure 1 indicates that the transporter protein of the present invention is expressed in the ovary

WO 02/48366

PCT/US01/47559

(adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain by a virtual northern blot. In addition, PCR-based tissue screening panels indicate expression in kidney.

Binding and/or activating compounds can also be screened by using chimeric transporter proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a ligand-binding region can be used that interacts with a different ligand than that which is recognized by the native transporter. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the transporter is derived.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the transporter (e.g. binding partners and/or ligands). Thus, a compound is exposed to a transporter polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble transporter polypeptide is also added to the mixture. If the test compound interacts with the soluble transporter polypeptide, it decreases the amount of complex formed or activity from the transporter target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the transporter. Thus, the soluble polypeptide that competes with the target transporter region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the transporter protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the

WO 02/48366

PCT/US01/47559

supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of transporter-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a transporter-binding protein and a candidate compound are incubated in the transporter protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the transporter protein target molecule, or which are reactive with transporter protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the transporters of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of transporter protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the transporter pathway, by treating cells or tissues that express the transporter. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. These methods of treatment include the steps of administering a modulator of transporter activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

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

WO 02/48366

PCT/US01/47559

example, downstream elements of a transporter-mediated signaling pathway. Alternatively, such transporter-binding proteins are likely to be transporter inhibitors.

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

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

The transporter proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. The method involves contacting a biological sample with a compound capable of interacting with the transporter

WO 02/48366

PCT/US01/47559

protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered transporter activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing

WO 02/48366

PCT/US01/47559

enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the transporter protein in which one or more of the transporter functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and transporter activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. Accordingly, methods for treatment include the use of the transporter protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal

WO 02/48366

PCT/US01/47559

antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, *Antibodies*, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the transporter proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or transporter/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ¹²⁵I, ¹³¹I, ³⁵S or ³H.

WO 02/48366

PCT/US01/47559

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in Figure 1 indicates that the transporter protein of the present invention is expressed in the ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain by a virtual northern blot. In addition, PCR-based tissue screening panels indicate expression in kidney. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression,

WO 02/48366

PCT/US01/47559

antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the transporter peptide to a binding partner such as a ligand or protein binding partner. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a transporter peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide

WO 02/48366

PCT/US01/47559

sequence that encodes one of the transporter peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1 and SEQ ID NO:2, transcript sequences and SEQ ID NO:5, genomic sequence), or any nucleic acid molecule that encodes the proteins provided in Figure 2, SEQ ID NO:3 and SEQ ID NO:4. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1 and SEQ ID NO:2, transcript sequences and SEQ ID NO:5, genomic sequence), or any nucleic acid molecule that encodes the proteins provided in Figure 2, SEQ ID NO:3 and SEQ ID NO:4. A nucleic acid molecule consists

WO 02/48366

PCT/US01/47559

essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1 and SEQ ID NO:2, transcript sequences and SEQ ID NO:5, genomic sequence), or any nucleic acid molecule that encodes the proteins provided in Figure 2, SEQ ID NO:3 and SEQ ID NO:4. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprise several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the transporter peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein

WO 02/48366

PCT/US01/47559

sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the transporter proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could be at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or

WO 02/48366

PCT/US01/47559

mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1.

Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 42 different nucleotide positions in introns and regions 5' and 3' of the ORF. Such SNPs in introns and outside the ORF may affect control/regulatory elements. Two SNPs in exons, of which 1 of these cause changes in the amino acid sequence (i.e., nonsynonymous SNPs). The changes in the amino acid sequence that these SNPs cause is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

WO 02/48366

PCT/US01/47559

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, SNPs, including insertion/deletion variants ("indels"), were identified at 42 different nucleotide positions.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

WO 02/48366

PCT/US01/47559

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in Figure 1 indicates that the transporter protein of the present invention is expressed in the ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain by a virtual northern blot.

Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in transporter protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a transporter protein, such as by measuring a level of a transporter-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a transporter gene has been mutated. Experimental data as provided in Figure 1 indicates that the transporter protein of the present invention is expressed in the ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain by a virtual northern blot. In addition, PCR-based tissue screening panels indicate expression in kidney.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate transporter nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the transporter gene, particularly biological and pathological processes that are mediated by the transporter in cells and tissues that express it. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain. The method typically includes assaying the ability of the compound to modulate the expression of the transporter nucleic acid and thus identifying a

WO 02/48366

PCT/US01/47559

compound that can be used to treat a disorder characterized by undesired transporter nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the transporter nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for transporter nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the transporter protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of transporter gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of transporter mRNA in the presence of the candidate compound is compared to the level of expression of transporter mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate transporter nucleic acid expression in cells and tissues that express the transporter. Experimental data as provided in Figure 1 indicates that the transporter protein of the present invention is expressed in the ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain by a virtual northern blot. In addition, PCR-based tissue screening panels indicate expression in kidney. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for transporter nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the transporter nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in Figure 1 indicates expression in ovary (adenocarcinoma tissue),

WO 02/48366

PCT/US01/47559

uterus (leiomyosarcoma tissue), cervix, kidney, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the transporter gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in transporter nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in transporter genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the transporter gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the transporter gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a transporter protein.

Individuals carrying mutations in the transporter gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 42 different nucleotide positions in introns and regions 5' and 3' of the ORF. Such SNPs in introns and outside the ORF may affect control/regulatory elements. Two SNPs in exons, of which 1 of these cause changes in the amino acid sequence (i.e., nonsynonymous SNPs). The changes in the amino acid sequence that these SNPs cause is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. As indicated by the data presented in Figure 3, the map position was determined to be on

WO 02/48366

PCT/US01/47559

chromosome 1. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abrahava *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a transporter gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant transporter gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and

WO 02/48366

PCT/US01/47559

Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the transporter gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 42 different nucleotide positions in introns and regions 5' and 3' of the ORF. Such SNPs in introns and outside the ORF may affect control/regulatory elements. Two SNPs in exons, of which 1 of these cause changes in the amino acid sequence (i.e., nonsynonymous SNPs). The changes in the amino acid sequence that these SNPs cause is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control transporter gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of transporter protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into transporter protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of transporter nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired transporter nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions

WO 02/48366

PCT/US01/47559

include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the transporter protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in transporter gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired transporter protein to treat the individual.

The invention also encompasses kits for detecting the presence of a transporter nucleic acid in a biological sample. Experimental data as provided in Figure 1 indicates that the transporter protein of the present invention is expressed in the ovary (adenocarcinoma tissue), uterus (leiomyosarcoma tissue), cervix, kidney cancer tissue (hypernephroma), germinal center B cell, colon, and infant brain by a virtual northern blot. In addition, PCR-based tissue screening panels indicate expression in kidney. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting transporter nucleic acid in a biological sample; means for determining the amount of transporter nucleic acid in the sample; and means for comparing the amount of transporter nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect transporter protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1, 2, and 5).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60

WO 02/48366

PCT/US01/47559

nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides that cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is

WO 02/48366

PCT/US01/47559

produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the transporter proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and/or alleles of the transporter gene of the present invention. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter protein of the present invention. SNPs were identified at 42 different nucleotide positions in introns and regions 5' and 3' of the ORF. Such SNPs in introns and outside the ORF may affect control/regulatory elements. Two SNPs in exons, of which 1 of these cause changes in the amino acid sequence (i.e., nonsynonymous SNPs). The changes in the amino acid sequence that these SNPs cause is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The

WO 02/48366

PCT/US01/47559

Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified transporter gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the

WO 02/48366

PCT/US01/47559

nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in

WO 02/48366

PCT/US01/47559

the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for

WO 02/48366

PCT/US01/47559

affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterotransporter. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A*

WO 02/48366

PCT/US01/47559

Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al. (Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).*

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and

WO 02/48366

PCT/US01/47559

dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as transporters, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with transporters, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a transporter protein or peptide that can be further purified to produce desired amounts of transporter protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the transporter protein or transporter protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native transporter protein is useful for assaying compounds that stimulate or inhibit transporter protein function.

WO 02/48366

PCT/US01/47559

Host cells are also useful for identifying transporter protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant transporter protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native transporter protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a transporter protein and identifying and evaluating modulators of transporter protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the transporter protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the transporter protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

WO 02/48366

PCT/US01/47559

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science* 251:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, transporter protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* transporter protein function, including ligand interaction, the effect of specific mutant transporter proteins on transporter protein function and ligand interaction, and the effect of chimeric transporter proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more transporter protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be

WO 02/48366

PCT/US01/47559

unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

WO 02/48366

PCT/US01/47559

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5;
 - (c) an amino acid sequence of an ortholog of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5; and
 - (d) a fragment of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said fragment comprises at least 10 contiguous amino acids.
2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:
 - (a) an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4;
 - (b) an amino acid sequence of an allelic variant of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5;
 - (c) an amino acid sequence of an ortholog of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions

WO 02/48366

PCT/US01/47559

- to the opposite strand of a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5; and
- (d) a fragment of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said fragment comprises at least 10 contiguous amino acids.
3. An isolated antibody that selectively binds to a peptide of claim 2.
4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:
- (a) a nucleotide sequence that encodes an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4;
 - (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5;
 - (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5;
 - (d) a nucleotide sequence that encodes a fragment of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said fragment comprises at least 10 contiguous amino acids; and
 - (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).
5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:
- (a) a nucleotide sequence that encodes an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4;
 - (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein

WO 02/48366

PCT/US01/47559

- said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5;
- (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5;
 - (d) a nucleotide sequence that encodes a fragment of an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4, wherein said fragment comprises at least 10 contiguous amino acids; and
 - (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).
6. A gene chip comprising a nucleic acid molecule of claim 5.
 7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.
 8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.
 9. A host cell containing the vector of claim 8.
 10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
 11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
 12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows

WO 02/48366

PCT/US01/47559

detection of the presence of the peptide in the sample and then detecting the presence of the peptide.

13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.
14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.
15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.
16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.
17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.
18. A method for treating a disease or condition mediated by a human transporter protein, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.
19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.
20. An isolated human transporter peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4.

WO 02/48366

PCT/US01/47559

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence selected from the group consisting of: SEQ ID NO:3 and SEQ ID NO:4.
22. An isolated nucleic acid molecule encoding a human transporter peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5.
23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule selected from the group consisting of: SEQ ID NO:1, SEQ ID NO:2 and SEQ ID NO:5.

WO 02/48366

1/30

PCT/US01/47559

Transcript 1:
 (582 ID NO:1)
 1 ATGACCCAGG GGAAGAAGAA GAAACGGGCG CGAACCAGCA GTATCATGCT
 51 GGCCACAGAG ATCATCATTA AGGACGGAGG CACGCCCTCA GGAATAGGTT
 101 CTCCTAGTGT CTATCATGCA GTATATGCTA TCTTTTGA GTTTTGTGCT
 151 TGGGGACTAT TGACAGACCC CACCTTGGTG GTATTACATG AATCCTTTCC
 201 TAAACATACA TTCTGTGTA AGGGTTTAT TCAGAGAGTA AAGGTTTCT
 251 TGTCATCTCT TAGTGCOCGG CTATATGGTG CTCCTTCTGA TGTTTGGGGC
 301 CGAATACTCT TCTTGTCTGT AACGGTGTCT TCACATGTG CCCAATTC
 351 TTTAATGAAG ATCAGCCCAT GGTGGTACTT TGTGTATC TCTGTCTCTG
 401 GGGTTTTCG AGTGACCTTT TCTGTGTTAT TGCATACGT AGCAGATATA
 451 ACCCAGAGGC ATGAAAGAGG TTGGCTTTAT GGTCTGTTT CACCAACCTT
 501 TGCTGCAAGT TAGTCACCA GTCTGCAAT TGGAGCTTAT CTGGACGAG
 551 TATATGGGGA CAGCTTGGTG GTGCTCTTAG CTACAGCAAT AGCTTTGCTA
 601 GATATTTGTT TTATCCTTGT TGGTGTGCA GAGTCTGTC CTGAGAAAT
 651 GGGGCAACA TCTTGGGAG CACCATCTC CTGGAGACA GCGACCTTT
 701 TTGCGTCTCT AAAAAGAGC GGCACAGATT CCATAGTCTT GCTATCTTGC
 751 ATTACAGTGT TTCTCTCTTA CCTACCGGAG CGAGGCCAAT ATTCCAGCTT
 801 TTTTATGATC CTCAGACAGA TAATGAATTT TTCACAGAA AGTGTTCAGG
 851 CGTTTATAGC AGTCCCTTGC TTCTTTTCCA TTATTCACA GAUATATGTC
 901 TTGGTTTTC TTATGAGGTC AATTGCAAT AAGAACCA TTTTACTGGG
 951 TCTAGGATTT CAATATATAC AGTTGGCATG GTATGGCTTT GGTTCAGAAC
 1001 CTTGGATGAT GTGGGCTGCT GGGGAGTAG CAGCATGTC TAGCATCACC
 1051 TTTCCTGCTG TCAGTGCACCT TGTTCACGA ACTGCTGATG CTGATCACA
 1101 GGGGTGCTT CAAGGATGA TACAGGAAAT TCAGAGTTA TGCATGCTC
 1151 TGGGACGCGC CCTCATGGA TTCTTTTCTT AATATTTCA TGTGGACTTT
 1201 AAGAGACTGC CAATACAGG AACAGACTTG GGAACAACA CAGCCCTCA
 1251 GCACACCTTT GACAGAAAT CCATCATCCC TGGCCCTCCC TTCTATTTTG
 1301 GAGCCCTGTC AGTATGCTG GCTCTGCTTG TTGCTTGT TTATCCGGA
 1351 CAATACATTT TACCTTAGG GTCCAGCAT TGGAGAGGC ACTGTGCGAG
 1401 TCACAGCATT CCTCATATA CACAGGCGC AGGAGAGGCC AAGAACCTT
 1451 TACTCCAGGA CACAAATGTG TGA

Transcript 1 Features:
 Start Codon: 1
 Stop Codon: 1471

Transcript 1 Top 10 BLAST Hits:

gi 6680221 ref NP_032272.1	hippocampus abundant gene transcrip...	977	0.0
gi 7292310 gb AAE47717.1	(AB003477) Cg11537 gene product [alt ...	557	e-157
gi 7508440 pir J133372	hypothetical protein T2503.1 - Caenorhab...	474	e-132
gi 6808288 emb CA870819.1	(AL137576) hypothetical protein [Hom...	388	e-106
gi 7296327 gb AAE51617.1	(AB003591) C65078 gene product [Droso...	382	e-105
gi 7296326 gb AAE51616.1	(AB003591) C61767 gene product [Dros...	183	3e-45
gi 7296325 gb AAE51615.1	(AB003591) C618281 gene product [Dros...	183	4e-45
gi 7473957 pir JH75515	tetracycline-efflux transporter - Deinoc...	139	8e-30
gi 3860032 gb AA099860.1	(AF090987) drug efflux protein TetA [...	132	6e-30
gi 207852 gb AA77664.1	(M20189) tet protein [synthetic const...	132	6e-30

FIGURE 1A

WO 02/48366

2/30

PCT/US01/47559

Transcript 2:

(SEQ ID NO:2)

```

1  atgacccagg ggaagaaga gaacgggccc gcgaaccgca gtatcatgct
51  ggccaagaag atcattcatt aggcaggagg cagctgatta catgaacact
101  ttcttaasca tacatttctg atgacgggct taattcaagg agtaaaaggt
151  ttgttgkcat tcatttagtc ccgccttatt ggtgctcttt ctgatgtttg
201  ggccgcaaaa tcctcttctg tgcatacgt gtctttcaaa tggccccaaa
251  ttctcttaat gaagatcgc ccactgggtgt accttgctgt tatctctgtt
301  tctgggggttt ttgcagtgc tttttctgtg gtatttgcat acgtagcaga
351  tataacccaa gagcatgaaa gaagtatggc ttatggactg gtttcagcaa
401  catttgctgc aagtttagtc accagtcctg caattggagg ttatcttgga
451  cgaatatatg gggaacgctt ggtggtggtc ttactatcag caatagcttt
501  gctagatatt tgttttatcc ttgttctgtg gccagagtcg ttgactgaga
551  aaatgcggcc agcatcctgg ggagcaacca ttctctggga acagctgac
601  ccttttgctg ccttaaaaaa agtcggccaa gattccatag tgcgtctgat
651  ctgcattaca gtgttctctc cctactatcc ggaaggcagg caatattcca
701  gctttttttt atacctcaga cagatatga antttctccc ggaaggttt
751  gcagcgttta tagacgtctc tggcattctt tccattatg ccacagccat
801  agtcttgagt ttacttatga ggtcaattgg aaataagaac accattttac
851  tgggtctagg atttcaaaaa ttacagttgg catggtatg ctttggttca
901  gaaccttggg tgttggggc tgcgtgggca gtagcagcca tgcctagcat
951  caactttctc gctgtcagtg cacttttttc acgaactgct gatgctgac
1001  aacagggtgt cgttcaagg atgataacag gaattcgagg attatgcaat
1051  ggtctggggc cggccctcta tggattcatt ttctacatat tccatgtgga
1101  acttaagaaa ctgcaataaa caggaaacaga cttgggaaca aacnaagcc
1151  ctccggccca ctttgaaag aattccatba tccctggccc tcccttcta
1201  ttgtggacct gtctagatct cctggtctct cttgttgctt tgtttattcc
1251  ggaacatacc aatttagctc taaggtccag cagttggaga aagcactgtg
1301  gcagtcacag ccattctcat aatacacaag ccgccaggaga ggcacaaaga
1351  cctttactcc agaacacaa tgtgtga

```

Transcript 2 Features:

Start Codon: 1

Stop Codon: 1375

Transcript 2 Top 10 blast hits:

gi16680221 ref NP_032272.1	hippocampus abundant gene transcrip...	895	0.0
gi17292310 gb AAAF7717.1	(AB003477) CG11537 gene product [alt ...	509	e-143
gi17508440 pir J23372	hypothetical protein T2503.1 - Caenorhab...	457	e-121
gi11360185 pir J246272	hypothetical protein DFP2p564E0864.1 - ...	388	e-107
gi17296327 gb AAFS1617.1	(AB003591) CG5078 gene product [Droso...	381	e-105
gi17296326 gb AAFS1616.1	(AB003591) CG17637 gene product [Dros...	145	1e-33
gi17296325 gb AAFS1615.1	(AB003591) CG18261 gene product [Dros...	145	1e-33
gi17473957 pir W75515	tetracycline-efflux transporter - Deinoc...	139	8e-32
gi13840032 gb AAD09860.1	(AF090987) drug efflux protein TetA [...	132	6e-30
gi1207952 gb AAA77664.1	(M20189) tet protein [synthetic constr...	132	6e-30
gi1984918 gb AAC53625.1	(U35134) tetracycline resistance prote...	132	8e-30
gi1236519 gb AA92917.1	(U46855) tetracycline resistance prot...	131	1e-29
gi19507411 ref NP_052131.1	protein responsible for tetracycline...	131	1e-29
gi12274944 emb CA04021.1	(AJ000346) NapC protein [Enterococcu...	127	2e-28
gi110956601 ref NP_052571.1	tetracycline resistance protein Te...	126	4e-28

FIGURE 1B

WO 02/48366

3/30

PCT/US01/47559

Best hits for Transcripts 1 and 2:

gb BF036183 BF036183 601450251F1 NIH_MGC_66 Homo sapiens cDNA c...	1197	0.0
gb BE884890 BE884890 601506923F1 NIH_MGC_71 Homo sapiens cDNA c...	1150	0.0
gb BE539469 BE539469 601060133F1 NIH_MGC_10 Homo sapiens cDNA c...	1013	0.0
gb BF029950 BF029950 601557355F1 NIH_MGC_58 Homo sapiens cDNA c...	912	0.0
gb AA223840 AA223840 xrl2d09.r1 Stratagene hNT neuron (#937233)...	876	0.0
gb AA641767 AA641767 nsl9d08.r1 WCI_OGAP_OCB1 Homo sapiens cDNA...	837	0.0
gb RE7165 RE7165 yb08c08.v2 Soares InfanE brain INIB Homo sapie...	753	0.0
gb AA099754 AA099754 xl79a03.r1 Stratagene colon (#937204) Homo...	735	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

Library Source:

gb|BF036183 Ovary, adenocarcinoma
gb|BE884890 Uterus Leiomyocarcinoma
gb|BE539469 Cervix
gb|BF029950 Kidney hypernephroma
gb|AA223840 Neuron
gb|AA641767 germinal center B Cell
gb|RE7165 Infant brain
gb|AA099754 Colon

5

Expression information from PCR-based tissue screening panels:
Human Kidney

FIGURE 1C

WO 02/48366

4/30

PCT/US01/47559

Protein Sequence 1:
(SEQ ID NO:3)
1 MTQSKKKRA ANRSIMLAKK IIKDGGTFQ GIGSPSVYHA VIVIFLEFFA
51 NGLLTAPTLV VLEETFFKHT FLMNGLIQGV KGLLSFLSAP LIGALSVDVWG
101 KKSLILLTVF FTCAPIPIAK ISPWNYFAVI SVSGVFATF SVVPAIVADI
151 TQCHESWAY GLYSKTERAS LVTSFAIGAK LGRVYGDSIV VVLAFAIAL
201 DICFLLVAVF ESLPERKREFA SNGRAFISWEE ADPFASLKV GDSIVLLIC
251 ITVPLSYLFE AGQYSSEFLY LRQIMKFSFE SVRAFIAVLG ILSIAQTIV
301 LSLIMRSIGN KNTILLGLGF QILQLAWYGF GSEFWMMNNA GAVRAMSSIT
351 FFAVSNLVSN TADADQGVV QSMITGINGL CHGLGPAIUG FIFTIFHVEE
401 KELPTCTOL GTWTSPOHHF RQNSIIFGPF FLFGACSVLL ALLVALTIFE
451 HTNLSIRSSS WRKHCGSHH ENNIQAPGEA KEVLLQDTRV

Membrane spanning structure and domains:
Helix Begin End Score Certainty
1 45 65 1.706 Certain
2 85 105 1.209 Certain
3 108 128 1.726 Certain
4 131 151 2.108 Certain
5 163 183 1.682 Certain
6 195 215 1.732 Certain
7 248 268 1.518 Certain
8 287 307 1.974 Certain
9 317 337 1.176 Certain
10 340 360 1.861 Certain
11 385 405 1.393 Certain
12 435 455 2.247 Certain

FIGURE 2A

WO 02/48366 5/30 PCT/US01/47559

BLAST Alignment to Top Hit for Protein Sequence 1:
>gi|6680221|ref|NP_032272.1| hippocampus abundant gene transcript 1
[Mus musculus] >gi|7513836|pir||JC5641 sugar transporter
protein H1AT1 - mouse >gi|2506078|dbj|BAZ2622.1|
(D88315) tetracycline transporter-like protein [Mus
musculus]
length = 490

Score = 977 bits (2498), Expect = 0.0
Identities = 486/490 (99%), Positives = 490/490 (99%)

Query: 1 MTQGGKKKRAANRSLMAKKIIKDGSTPQGIQSPSVYHAVIVIFLEFFAWGLLTAPTLV 60
MTQGGKKKRAANRSLMAKKIIKDGSTPQGIQSPSVYHAVIVIFLEFFAWGLLTAPTLV
Sbjct: 1 MTQGGKKKRAANRSLMAKKIIKDGSTPQGIQSPSVYHAVIVIFLEFFAWGLLTAPTLV 60

Query: 61 VLHETFPKHTFLMNGLIQGVKGLLSFLSAFLIGALSDVWGRKSFLLTVFTTCAPILMK 120
VLHETFPKHTFLMNGLIQGVKGLLSFLSAFLIGALSDVWGRKSFLLTVFTTCAPILMK
Sbjct: 61 VLHETFPKHTFLMNGLIQGVKGLLSFLSAFLIGALSDVWGRKSFLLTVFTTCAPILMK 120

Query: 121 ISFWWYFAVISVSGVFVTFVVFVAVADITQHERSMAYGLVSAFPAASIVTSPAIGNY 180
ISFWWYFAVISVSGVFVTFVVFVAVADITQHERSMAYGLVSAFPAASIVTSPAIGNY
Sbjct: 121 ISFWWYFAVISVSGVFVTFVVFVAVADITQHERSMAYGLVSAFPAASIVTSPAIGNY 180

Query: 181 LGRVYGSLSVVVLTATAIALDICTILVAVFESLFEKMRASNGAPISWEQADFFASLKKV 240
LGRVYGSLSVVVLTATAIALDICTILVAVFESLFEKMRASNGAPISWEQADFFASLKKV
Sbjct: 181 LGRVYGSLSVVVLTATAIALDICTILVAVFESLFEKMRASNGAPISWEQADFFASLKKV 240

Query: 241 GQDSIVLLICITVFSLYLPBAGQYSSFFLYLQIMKFSFESVAAFIATVIGLISIIAQIV 300
GQDSIVLLICITVFSLYLPBAGQYSSFFLYLQIMKFSFESVAAFIATVIGLISIIAQIV
Sbjct: 241 GQDSIVLLICITVFSLYLPBAGQYSSFFLYLQIMKFSFESVAAFIATVIGLISIIAQIV 300

Query: 301 LSLMRSIGNKNTILLGLGQILQAWYFGSSEPMWAGAVAMSSITTFVAVSALVSR 360
LSLMRSIGNKNTILLGLGQILQAWYFGSSEPMWAGAVAMSSITTFVAVSALVSR
Sbjct: 301 LSLMRSIGNKNTILLGLGQILQAWYFGSSEPMWAGAVAMSSITTFVAVSALVSR 360

Query: 361 TADADQGVVQGMITGIRLCNGLGALYGFIFYIFHVELKELPITGDLGTNTSPQHNF 420
TADADQGVVQGMITGIRLCNGLGALYGFIFYIFHVELKELPITGDLGTNTSPQHNF
Sbjct: 361 TADADQGVVQGMITGIRLCNGLGALYGFIFYIFHVELKELPITGDLGTNTSPQHNF 420

Query: 421 PQNSIIPGPPFLFGACSVLLALLVALEIPEHNLRLSSSWRRKCGSHSHPHNTQAPGEA 480
PQNSIIPGPPFLFGACSVLLALLVALEIPEHNLRLSSSWRRKCGSHSHPHNTQAPGEA
Sbjct: 421 PQNSIIPGPPFLFGACSVLLALLVALEIPEHNLRLSSSWRRKCGSHSHPHNTQAPGEA 480

Query: 481 KEPLQDITNV 490
KEPLQDITNV
Sbjct: 481 KEPLQDITNV 490 (SEQ ID NO:6)

HMM results:

Model	Description	Score	E-value	N
PF00083	Sugar (and other) transporter	3.7	2.4	2

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value		
PF00083	1/2	85	129	..	63	116	..	6.6	0.35
PF00083	2/2	422	450	..	442	474	..	0.2	24

FIGURE 2B

WO 02/48366

6/30

PCT/US01/47559

Protein Sequence 2:
(Seq ID NO:4)
1 MTQGGKKKRA ANRSIMIAKK IIKDGGTVL HETFFKHTFL MNGLIQGVKG
51 LLSFLSAPLI GALSVDWGRK SFLLLTVEFT CAPFLMKIS PMWYFAVISV
101 SGVFAVTFEV VFAYVADITQ ERSRMAVGL VSATFANSLV TSPAIGAILG
151 NVYGDSDVVV LATAIALDI CFIIVAVES LPEKRRPASW GAPISWEGAD
201 PFASLKKVGO DSITLLICIT VFLVLEPAG QVSSFFVLR QIMKFSPEV
251 AAFIAVLGIL SITAQTVLS LLMRSIGNHN TILGLGFQI LQLAWYGFQS
301 EPWMMWAAGA VAMSSITFF AVSALVSRKA DADQGVVQG MITGIRGLCN
351 GLGPAHYGI FYIHFVELKE LPITGTDLGT WTSPQHRFEG NSIIPGPFPL
401 FRCQSVLLAL LVALTPEHT NLSLASSWR KEGGSRSHPR WTQAPGEAKL
451 PLLQDTRV

FEATURES:
Functional domains and key regions:
Prosite results for Protein Sequence 2:
[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site
Number of matches: 2
1 12-15 NRSI
2 421-424 NLSL

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site
Number of matches: 3
1 204-206 SLR
2 423-425 SLR
3 428-430 SWR

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site
Number of matches: 2
1 180-183 SLPE
2 374-377 TGTD

[4] PDOC00008 PS00008 MYRISTYL
N-myristoylation site
Number of matches: 12
1 43-48 GLTGGV
2 129-134 GLVSAT
3 146-151 GAYLGR
4 191-196 GAFISW
5 230-235 GQVSSF
6 277-282 GNKNII
7 309-314 GAVVAM
8 336-341 GVVGGM
9 340-345 GWITGI
10 347-352 GLCKGL
11 351-356 GLGPAI
12 375-380 GTDLGT

[5] PDOC00009 PS00009 AMIDATION
Amidation site
Number of matches: 2
1 3-6 QGKK
2 67-70 WGRK

[6] PDOC00373 PS00343 GRAM_POS_ANCHORING
Gram-positive cocci surface proteins 'anchoring' hexapeptide
371-376 LPITGT

[7] PDOC00190 PS00216 SUGAR_TRANSPORT_1
Sugar transport proteins signature 1
60-76 IGALSDVWGRKSFLLLT

FIGURE 2C

WO 02/48366

7/30

PCT/US01/47559

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	47	67	1.209	Certain
2	70	90	1.726	Certain
3	93	113	2.108	Certain
4	125	145	1.682	Certain
5	157	177	1.732	Certain
6	210	230	1.518	Certain
7	249	269	1.974	Certain
8	279	299	1.176	Certain
9	302	322	1.861	Certain
10	347	367	1.393	Certain
11	397	417	2.247	Certain

BLAST Alignment to Top Hit for Protein Sequence 2:

>gi|6680221|ref|NP_032272.1| hippocampus abundant gene transcript 1
[Mus musculus]
pir||JC5641 sugar transporter protein Hmt1 - mouse
dbj|BA022622.1| (D88315) tetracycline transporter-like protein [Mus musculus]
Length = 490

Score = 895 bits (2288), Expect = 0.0
Identities = 454/490 (92%), Positives = 458/490 (92%), Gaps = 32/490 (6%)

Query: 1 NTQGRKKRRAANSIMLAKKIIIKDQGT----- 28
NTQGRKKRRAANSIMLAKKIIIKDQGT
Sbjct: 1 NTQGRKKRRAANSIMLAKKIIIKDQGTQQIGSPVHVAVIVILEFFANGLLTATPLV 60

Query: 29 VLHETFPKHTFMNGLIQGVKGLLSFSLAFLIGALSVDVGRKSFLLTVFTCAPIPLMK 88
VLHETFPKHTFMNGLIQGVKGLLSFSLAFLIGALSVDVGRKSFLLTVFTCAPIPLMK 120
Sbjct: 61 VLHETFPKHTFMNGLIQGVKGLLSFSLAFLIGALSVDVGRKSFLLTVFTCAPIPLMK 120

Query: 89 ISPWNYFAVISVSGVFAVTFVVFAYVADITQHERSMAYGLVSATFAASLVTSFAIGAY 148
ISPWNYFAVISVSGVFAVTFVVFAYVADITQHERSMAYGLVSATFAASLVTSFAIGAY 180
Sbjct: 121 ISPWNYFAVISVSGVFAVTFVVFAYVADITQHERSMAYGLVSATFAASLVTSFAIGAY 180

Query: 149 LGRVYGDLSVVVLTATIALLDICFILVAVPESELPEKMRASWGAPISWEQADFFASLKKV 208
LG++YGDLSVVVLTATIALLDICFILVAVPESELPEKMRASWGAPISWEQADFFASLKKV 240
Sbjct: 181 LGRVYGDLSVVVLTATIALLDICFILVAVPESELPEKMRASWGAPISWEQADFFASLKKV 240

Query: 209 QQDSIVLLICITVFLSYLPEAGQYSFFLYLRQIMKFSFESVAFTAVLGLLSIIAQTIV 268
QQDSIVLLICITVFLSYLPEAGQYSFFLYLRQIMKFSFESVAFTAVLGLLSIIAQTIV 300
Sbjct: 241 QQDSIVLLICITVFLSYLPEAGQYSFFLYLRQIMKFSFESVAFTAVLGLLSIIAQTIV 300

Query: 269 LSLMRSIGNKNTILLGLGQILQLAWYFGSGEPWMMNAGAVAMSSITFPAVSALVSR 328
LSLMRSIGNKNTILLGLGQILQLAWYFGSGEPWMMNAGAVAMSSITFPAVSALVSR 360
Sbjct: 301 LSLMRSIGNKNTILLGLGQILQLAWYFGSGEPWMMNAGAVAMSSITFPAVSALVSR 360

Query: 329 TADADQQGVVQGMITGIRGLCNGLSPALYGTIFYIFHVELKELPTGTDLGTNTSPQHFF 388
TADADQQGVVQGMITGIRGLCNGLSPALYGTIFYIFHVELKELPTGTDLGTNTSPQHFF 420
Sbjct: 361 TADADQQGVVQGMITGIRGLCNGLSPALYGTIFYIFHVELKELPTGTDLGTNTSPQHFF 420

Query: 389 EQNSIIPGPFPLGACSVLLALLVALEIPEHTNLSLRSSWRKHCQSHSHHTQAPGEA 448
EQNSIIPGPFPLGACSVLLALLVALEIPEHTNLSLRSSWRKHCQSHSHHTQAPGEA 480
Sbjct: 421 EQNSIIPGPFPLGACSVLLALLVALEIPEHTNLSLRSSWRKHCQSHSHHTQAPGEA 480

Query: 449 KEPLLQDTNV 458
KEPLLQDTNV
Sbjct: 481 KEPLLQDTNV 490 (SEQ ID NO:7)

RMS results:

Model	Description	Score	E-value	N
PF00083	Sugar (and other) transporter	3.7	2.4	2

Parse for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00083	1/2	53	97	63	116	6.6	0.35
PF00083	2/2	390	418	442	474	0.2	24

FIGURE 2D

WO 02/48366

PCT/US01/47559

```

(SEQ ID NO:5)
1  AGTCATCTGCT  TATTTTITAC  TTGTAATTTT  GTTGTITGT  GGGATTFAAA
51  AATATATTTT  ATTCTGAGGA  TAGTTGAATC  CACAGGATAC  TGAGGGCCAG
101  CTGTATTTCAC  AATCCAAATC  ACATACAAAG  CGACAGTTTC  ATACACATA
151  GGCCATATTAG  AACAGGACTG  TTCTCTCTTG  TTTATCATTG  CAGCCCTTCT
201  AGCACAAGGC  CTGGGCAATT  CTGGCAATT  AGTATGTGTT  AATTTCTCT
251  TACTACATTA  TTTCCAAACG  TATTTACTGC  AATCTGCAAT  TACCTTCCTT
301  TTGTTTTGTA  ACTGTGTCCC  CCACTAGAA  GTAAGCTCTG  TGACAGATAGT
351  GTCTCATTTA  TTGATGTATC  COTGGCATCT  AATAAAGCAC  TGACAAACCA
401  AGACCCAGAT  AATATATTTT  TGAATGACTG  AACATATACC  AGTTCTAAG
451  CTGTATATA  TTGATATGAC  TGAATGAAAT  ATGATTTTGA  GGAATCTGA
501  AATATCAAGT  AATCCAAACG  TAAATATACA  GCCTCTACT  TCCATTCGAA
551  ACAGCTCCT  GGAATTTGAA  TATATGAATC  TATCATATTA  CTCTCTCTAA
601  AGACATCTCA  GTTGGGTCTC  AACTAATAAG  TTGCAAGCGA  GCGCGGCTGG
651  CTCAAACCTG  TATTTCCAGC  ACTTTGCTAG  AACCTAGAT  CTTTGAAGGC
701  CAGTTTGA  ACCAGCTGG  TCAACATAC  CAGCTCTCT  CTTCACAAA
751  GAAATATTA  AATTAGCCAG  CCACTGGTGT  GTGCAATTTG  AGTCCAGAT
801  ACTTGAAGCG  CTAGGCGAGA  AGATCTGTT  CGAAGAGGC  TGACAGAGGC
851  CATGATGCA  CCACTGCACT  CCAAGCTGG  TGACAGAGTG  AGACCTTCC
901  TCAAAATA  ATATATGCT  ACCATGCTG  GCTCAACCT  GATATCCAG
951  CACTTTGGA  GCTTGGGCG  GGAATACAC  CTCAGTAG  GATTCAGC
1001  CTGGCCACA  CGGTGAAT  CCACTCTAC  TAAATAACA  AAAAAATTA
1051  GCGAGCATG  GTAGTGGGA  CCTGTATCC  CAGCTACTCG  GAGGCTGAG
1101  CAGAGAGAT  CACTTGAC  TGGGAGCTGA  AAGTTGCACT  GAGCCAGAT
1151  GCGCTTTTG  CACTGAGCG  TGGGAGCAAC  AGGCAATCT  CTCTCTCAA
1201  AATATATA  TAAAAAAGT  TTCAAAATGA  GAATATATGT  TTCAAAACA
1251  GTATATGAA  TATACTTAT  GATTGGAATA  TATAATAGA  AGTATCTATC
1301  AGGCTATATA  TTGCTTTCT  TCTCCCTCC  ATGGAATTA  GTTTTTTTT
1351  CATTTTTTA  TCAATATGA  ATATACAGA  AAGGAATAT  CATCAATAT
1401  TACTTTTTTA  CATTTTAT  GGAATATAT  TCAACTTAC  AGAATAGTG
1451  CAAAGACAT  ACAAGAACT  CATACATTA  TTTACTGTT  TTCTTTTAC
1501  CCTATATAT  AGTATTTAT  AGCTGTGTA  CAATATACC  CAAAGCTTAG
1551  TGGCTTAC  CAGTACTTT  TCACTTACA  CTCTTCTGA  GTACAGATC
1601  CAGGATGG  TACTATGT  GGTCTTACT  CTGGGTCTCT  CAGTATGTT
1651  TAGTACCA  AAGCTTGC  CAGGTCTGA  GATCTACT  CCAAGTGC
1701  TCACTCGTG  GCAITTTGTA  GAGGCTACA  AACATCTCT  GAGCACTGG
1751  ATCTCTCAT  AGGCTGCTG  AGTGTCTGA  AACACGGA  CAGGCTTCC
1801  CAGGCTCA  AGCCCAAA  TGAATGA  AAGGAGCCG  AAGGAGAT
1851  GCACTGCTT  TTATGACCTA  GCTCTGAG  TCAATCTCT  CACTCTGTT
1901  TTGATCTAT  AAGATCACT  AAGCTATG  TACATCAAG  GGGAGGGA
1951  TTAGATCCA  CCTCTCCAG  GAGGAAAT  CATTAATCT  GTAGCATAT
2001  CTAGACAT  CCACTCCAG  TTATGATCT  TTAAACAT  GGCATTTTG
2051  CTGTCTCT  CTCTTTCC  CACTATAT  TACTATACA  TACTATAT
2101  TCCCTAAC  TCTGAGATA  GAGGAGCTG  GGTATCTCC  TATCTCTGT
2151  GTTTTCTCT  TAAGGAAAG  GATATCTAT  TATACAGAC  GATAGTTATC
2201  ATATCTAAT  TTATACATG  TGAATCTTA  AAGTCCACT  CCACTGTGT
2251  AATTTGCTT  TTATGCAAT  GTTCCACCA  AATTTATTT  TAAACATAT
2301  AATTAAGAG  TAAAGGCTA  CAAAGGCTG  CATGCTGAG  GTATGATAT
2351  TTTTGTGCA  GGAATAAAA  AAGTCTGAG  GTTAGGAG  GTAAACAGC
2401  GTAGCCAGC  TGAGTCTAG  CCGGCTCC  TGAACACCA  AAGGACAGC
2451  CTGTTCTGA  AACATACAT  TATCTCTTA  GTTACCAT  ACTTACGGA
2501  AATGATTTT  TATGCTTCC  TTGCTCTCT  GTCTGATTA  GAGCAAGCT
2551  GAGGTTGTG  TAAATAGCT  AGGAGCTCT  TTCTGAGAG  CAGTATACA
2601  CTTTTTAAA  CAAATTTCT  CATCTGCTA  AAGAGGAGA  CAAATACGC
2651  CCACTCTGC  CAGACCTGT  GCGAGATTT  AGTATGCTG  TGAAGCCGA
2701  GAGGAGAT  GTAGAGTAG  CAGAGGAC  CAGGAGGAG  CAGGAGGAG
2751  CTTGCTGCT  CAGCTACTG  AGATGCTCT  GCGGAGGAG  GCGGAGGAG
2801  GTATCAGCG  GAGGAGGAG  GCGGCTCTG  TCGGCTGAT  GCACTGATG
2851  GAGGAGGAG  CAGGAGGAG  CCGGCTCTG  AAGATGCTG  GCGGAGGAG
2901  GAGGAGGAG  GCGGAGGAG  AGTATGCTG  GCGGAGGAG  TCGTGGGAG
2951  TCGGAGGAG  GAGGAGGAG  AAGGAGGAG  AAGGAGGAG  GAGGAGGAG
3001  TGAAGGAGG  GAGGAGGAG  AAGGAGGAG  CAGGAGGAG  TATCTGCTG
3051  GCAAGGAGG  TCAATATTA  GAGGAGGAG  CAGGAGGAG  GAGTCTGCTG
3101  CCGGAGGAG  TCGTGGGAG  CCGGAGGAG  GTCTCTCTT  CAGGAGGAG
3151  CACTGCTCT  TCTAGGAGG  CCGGAGGAG  CCGGAGGAG  CCGGAGGAG
3201  CAGGAGGAG  GCTCTCTCT  GCTCTCTCT  CCGGAGGAG  CCGGAGGAG
3251  AATGCTCTCT  GCTCTCTCT  GCTCTCTCT  CCGGAGGAG  CCGGAGGAG
3301  TTTTCCGAG  CTGCTCTCT  CTGCTCTCT  CCGGAGGAG  CCGGAGGAG
3351  CCTGAGGAG  CAGGAGGAG  CCGGAGGAG  GAGGAGGAG  GAGGAGGAG
3401  CCGGAGGAG  CCGGAGGAG  GCGGAGGAG  GCGGAGGAG  GCGGAGGAG
3451  TCGGAGGAG  GCGGAGGAG  GCGGAGGAG  GCGGAGGAG  GCGGAGGAG
3501  ACGGAGGAG  CTCTCTCTA  TGTGCTTAC  GAGGAGGAG  CTGAGGAGG
3551  GATGAGGAG  CAGGAGGAG  GTATGCTCT  TTTCTGTTT  CAGGAGGAG
3601  TCGGAGGAG  CCGGAGGAG  GCGGAGGAG  CAGGAGGAG  TGAAGGAG
3651  TGGGAGGAG  TCACTCTCT  GATGAGGAG  CAGGAGGAG  GAGGAGGAG

```

FIGURE 3A

WO 02/48366

PCT/US01/47559

3701 GCCTACATTC GAACCTGCTA GTTTTCAAAG AAAAGTCATC ACTGTGTGTC
 3751 TTAGAGATCRA AAGTATTAGA ATCAGTCATG GCCTAAGAT CGAGAGAGSA
 3801 CACTTGTAGG GGAAGAGAGG TTCTCTTTT AGAAGACGTT GCATCACA
 3851 TAAACTTTAT GCAGTGTGTA GTTACCCAGC TGGGAGAGTA GGATTTTAA
 3901 TTGAAAACA AACAAACA AACTGTGTTT ATGCTAACAT TTCTCGGTG
 3951 CTACACTGTG TGGCTTTTGT TGCATCCGCT GATACCGGCT TGTGAATAG
 4001 AATGSAAGG TGTATATAT GTTTCACATA CTTGAGTGT CGAGAAATG
 4051 TACCAATAT TCCATTTCG TTATATCTT ATTGAGGCGG GATCAACTG
 4101 CTAGCACAGT AGTAAATGTG TAAGTAGGCG ACCATTGAGG ATTTGCTGAA
 4151 TTCAGTTGAA AAGCGTGACA AATTTTATG ACATTTCAGA ACACGGCCCA
 4201 GTCAATTTGC CAAAGTTTAG AAAACTGAG ACATATGTAA TGACTTTGGA
 4251 AATATTTTT ACTTATACCT TTATATATG TTAAGCTTT GACATTAAT
 4301 GAAAAAAGA AACAAATCC TCAAGTCTT TTTATTGAC TTGATTAAAT
 4351 AAAACATTAC TTTGATTAGA TCGGTTATGA AGAGTCATAG CTCCTTTGAC
 4401 CAGTAGAGTA AGACTATGT GGGAGAGAAA ATACTGTTGC CTTTGTCTAC
 4451 CTTTAGAGAG AGACAAATTT TTACATCTT CATAAATCT ACAAAACAGT
 4501 GCACAGGAA GTTGTATTT TGTAAAGGCG AGTCAATTT AGATCTAGTA
 4551 TTTTITACAA AATGTAGAA TGAAGCTGAT TAAGAAACAC AGCTTCCTTT
 4601 TCTTGGAAA GTTCAGTTT ATTACTTTC TTGGGGTIT TGTTTGATTT
 4651 GCTTTACAGC AGATGCTTTC TTCCAAATC CTGTAGATTT TGGAAAAGAT
 4701 GCTTTTAA CTTCTCTGTC CTATATATA GCTGTAAAT AATCTTAGC
 4751 CTGCTTTGGG ACACAAATA AATGTTTGC ACCAGCAATA GGTTTACAT
 4801 AGACCAATG AAGACTTTTC TTGAGGGGTC TGAACATGGG GGCATTATC
 4851 ATTTCATCAT TTTATACACT TAATATTTCA TTCTCTATTC TAAGAGCACT
 4901 GGGCACTCTT TTAGAAAAGG GGCCTTGTTT TGTATGTTG GATCCACAG
 4951 GGCCTAGAT GGAATTTTA AATCAAAA AAGCTCTCTA TTTTATCTA
 5001 GCACCTTCTT AGATGAATTT TTATGTAAAT TTGTGTTAAT CTATACGTA
 5051 ATCAGAGGAT ATATTTCAAT AATATCAGG GGAATATTTT GCAATTATTTG
 5101 TATTTTATTC CATCCAGCT TTAATTTTAA AAGGTATAC TATTCAGATC
 5151 ATAGAAATA TTGTAAATG GTATGTGCTT ATCTACCTCT CACTATACA
 5201 TAGTTCAAGC TACTATTATG AACITTTTTT GTTCTTTTGT TGTAGATGGA
 5251 GTCTCACTCT GTTGCCTCAG CTGAGGAGGT GCAGTGGCAG GATCTGGCT
 5301 CACGTATACC ACCGCTCTCT GGGTTCAAGT GATTCTCTG CTCACGCTC
 5351 CCGAGTAGCT GCGACTACAG GCAGCTGCCA CCACTGCTGG CTAATTTTTT
 5401 ATATTTTAGT TGAAGAGAA GTTCTACAT ATTGGCAGG CTGCTGTGA
 5451 ACTCTGACC TCATGATCTA CCACTCTGG CCTCCCAAG TCAAGGATTT
 5501 ACAGGTGTGA GCAACCGTGC CCACTCTGAA CATTTTTAA GAAAGGGGAA
 5551 AARATTGCGA TTGATATC TCATGTTGTC TGTTTTAA TCACTGTAT
 5601 CAGAGAAAT TTCACTGCTA TTCTGTGTA CTTCAAGTT TCGAGATTTT
 5651 TATAAGCTTC CCGCAATGGG TAGATTCATT CAGTGAATG TCTGAGGCTC
 5701 TATGCTCGGT ACATGACAGT CGTGAACAGT ATTTCACATA CTTGCTCAAT
 5751 GGTACTGATT TGATCCCCCT TCTGATTTCT TCTTTTAA CATTGTTAAT
 5801 AATCTTTTC CCGTTGTCTT GCTAATGACA TATATGTAG CTAATTTGGC
 5851 CMTTAAAT ATTTATAC AATCACTATA AGATGTGTA AGATTTTTT
 5901 TGAAGATTAG AGCAGATTAG AGCAGATTG TAGTTTTCAA CGGCTGAGA
 5951 AATAAATCTT TCTAAATGAG CCAATTAAT CGTAGTTAC TGATTTTTT
 6001 ATTGAATTTG TATTTCAATG AATTTGATTT CATTACGCTG AATGAAAAC
 6051 AGGCAAGAT AAGCTACACA AGTAGGCTAC CTATGTAGT AGGCAATTT
 6101 AAGTAAATTT ACATTAAGCT GTCTATTTT AATATAGGTC TTTTAACTT
 6151 GGGTCTGTCT ACCCTGAAGT TGTTTGCAA ATTTCTTTT GCTATACAT
 6201 GTTCTTGGGA AGAGTCCCAA AAGTCCATA CTTCCAAA GTTTAAGAGC
 6251 AATGTTCTG TTGAAAACA GCAATAGTAA CTAAAGATA AGTTCCACAT
 6301 ATTATATCA GTAAATTTT AATATATAG TGAATATC TTACTGATG
 6351 AAGTAAACA TATTAATGAA TTGCTTTTAA AAGCATCAT ATATAGAAAT
 6401 AGTTTTTAGG CCAAGGGCAG TGGCTCAGC CTGTAACTCC AGCACTTTGG
 6451 GAGGCCAGTG TGGGAGATC ACTTGAAGCC AGGAGTCGA GACTAGGCTG
 6501 GCAACATG TGAAGCCA TCTTACAA AATACAAA ATAGSCTTAGG
 6551 TGTGCTGTA TGTGCTGTA ATCTAGTA CCGGGGAG CGAGGAGCA
 6601 GATCACTTG AACTGGGAG GCAAGATTG CAGTGAAGCT AGATCAGGC
 6651 ACTGCAATCC AGCTGGGTC ATGAGTGAA ACTGTCTAA AAAAAAATA
 6701 AAGAGATTT TTAGTTACAG GTTTTATGT ATGTAACTT CAGTGTAGGT
 6751 ATTAAAGCA GCTGAATTA AATAGCTT TGAATTTT AATATCAGA
 6801 ATCTGTTT GTTTTATTA AGCATACCA CTTGTTTTTA AGCATCTCTG
 6851 TTAGAGGCTA AGAGTAAAG AGTTATTTAC AGTATTAATA TTGAATTTT
 6901 CTATCTTTT AATATGCTCA TCTTAAATA TGATCTTAT TGTTTGGCC
 6951 ATCAATTTG GGAATACAT CTTGACAGT GGAATATTA TAGTCTTTC
 7001 AAGATTTTGT GTTAAATCA CTTTAAAGT TGTATAGAA GTAGSCTTA
 7051 TGTGAATTC ATGTGCGAA GTACCTGAAA AAGTAAAT TGATATATCA
 7101 TTTGAAATG AATGTCAAT CCGTGATTA GCAGACAC AGATTTTTT
 7151 TTCTTGGCA TCCCTGATTT TAATTAATA GGAGTCAGCA ACCGTTTCAA
 7201 GAGAGAGCC CAGCTCTGA CCGTTGCA TTTCTACTG CAGAGTGGC
 7251 TGAAGTAGTG CAGGAGAGC TCTCTGGAG GTAGGAGCTT TGTAGACCA
 7301 GAGAGCTGTT AATTAACCTT TGGTTGCTAG CATGACAGA ATAGAAAGG
 7351 CCGTGTGTC TTCTCTTTT CTCTCTTTT TTTTCTTT TGTAGCAGAG
 7401 TTTGCTCTT GTTCTCAG CTGGGTCGA ATGGCTGAT CTTGCTTAC

FIGURE 3B

WO 02/48366

PCT/US01/47559

7451 AGCAACCTCT GCCTCCCTGG TTCAGGAAT TCTCTACCT TAGCCTCCTG
 7501 AATAGCTGGG ATTACAGGCA TGTGCAATCA TCCCAGCTA ATTTTGTAT
 7551 TTTTGTATGA GAGACGGAT TTTCACATG TGGCAGGCT GGTCTGAGC
 7601 TCTTGACTCT GGGTGATCCA CCGCGCTCAG TCTTCCAAAG TGGGATTACA
 7651 GGTGTGAGCC ACTGCGCCTG GCCCGCTGGT GCTTTTCAAA AAGCCTAGAA
 7701 ACATCAGGGT GTTTATATAG TCTTTGGCAG GTGTGTGGCT GGCAGCATCA
 7751 TTAATTAAGT AGCTCCCTK CYCAGTGGT CAGTGTGGG TTTGAGTGG
 7801 TGTGTTTGGG GATAAATTA TATCCAGTIT TTTTTCAGA TGCTATATG
 7851 CATCAGTCT ATCCTCATGT AGTATGAAGA CTGCATACC AACATATCA
 7901 GACCGTCTGC AGAATPTCT CTACAGTTGA AATGTAACT CTTTGAGCT
 7951 ACTGAAAGTT TAAGTTTBA GTAAATAAAT GAATAGCTTT CTTCAGTAA
 8001 CATCTAGCA AGCTGTGTG ATTAATAAGT TCAATATA AGGACACTG
 8051 ATTGTCTTT AGCATTATTT TAATTTGGA GTGTGAAGT ACAGTTGAC
 8101 AGTTTCAAGC AGGTATCTTC AAGAAAGAT GAATATGGT ATATATAGC
 8151 TCTCAAAAT TCCATATCT TATAACAAG AATATCCCT CCACTGCCC
 8201 TGTTTTIRA AAAATHAATA ATACATGTT TCTTCCAGT CCGTGGAAAG
 8251 TATTAAGAT GTTTCAGGAG GAGCAGATTT ATGCACATA CTGTGATTT
 8301 TCCATTCAAG TTTTACTTTA TTTTAAAT TATATCTTT TAAATATTT
 8351 TCAATATGTT TGGGGTACAG GTGGGTTTTT GGTACATAG ATGTTTTTA
 8401 GTGATGATTT CTGAGATTTT AGTGCACCTG TCACCTGACC AGTGTATCT
 8451 GTACCCARAT TATAGTCTTT TATCCTCTC AGCTTCCCC CCATCTCTA
 8501 AAGTCCCTCT TATAGTCTTT AGCCTTTTGC GTCCCATAG CTGAATCTC
 8551 ACTGTAAAGT GAGACATAC GACATTTGGT TTTCTATCC TGAATTAATC
 8601 ACTTAGATA ATGCGCTTCA ATTCCATCCA AGTTCTGCA AAGACATTA
 8651 TTTCTATCTT TTTTATGCT AGATATCCA TGGTATATAT ACACCATTT
 8701 TCTTATATCT ACTTTTGGT CATTGGGCA TGGGCTGGT TCCATATCT
 8751 TGCAATGTG AATTGTGCTG CTATAACNT GCATGTACAT GTGTCTTTT
 8801 CAATATATGA CTCTTTTAC TTTGGTGGG TACCAATAG TGGGATGCT
 8851 GGATCAATCA GTAGTCTTAT TTTTATGCT TTAAGGATC GCCATATCT
 8901 TTTCTATATCT GTTGTATCA GTTATATCT CCAACAGAG TGTCAATGT
 8951 TCTATTTGCT ACCACATCCA CACATCTAT TATTTTTGA TTTTAAAT
 9001 ATGGCCATCT TTGAGGAGT AATGATATC TCAATGTGGT TTTAATTTG
 9051 ATTTCCCTGA TATTTGGTGA TGTGAGCAT CTTTTCATAT GTTGTGTTG
 9101 TATTTATGCT CTTTGTAAA AATGTCTAT CATGTCTTT GCCTACTTT
 9151 GATGGGATG ATTTGTTTT TTTTGTGCA TTTGATCTG TTTGATATC
 9201 TGGGTACTAG TCTTTTCTA GATGCAAGT TCAATAATAT TTTTCAAC
 9251 TGTATGGGTT GTCTGTTTAC TCTGCTGAT TTTTTTTT TTTTTTTT
 9301 AGATGGAAT TTTCTCTGT TTTCCAGGCT GAGTGCAT TGCATATCT
 9351 TGGCTCCCTG CAACTCTAG CTCTCAGGTT CAGCATTCT TCCAGCTCA
 9401 GCTTCCAGAG TACGTGGAT TACAGCACT CACCAAGAG CCGTGTATC
 9451 TTTTTGTAT TTTTATAGA GACAGTTTT CTCTATGTT GCGAGGCTG
 9501 ACTCAACTA CAGACCTAG GTGATCCAC CGCTTGGCC TCCAGGATG
 9551 CTGGATATC AGCATGCTT AGGCGCTAT AAGTATTTG CTTTATTTCT
 9601 GGTATATCT TTTGTCTCA TTTGCTTCA TGGCATTTT TATACAGTA
 9651 CATGCGGTT TTTGTATCT TACGCTTCT TATATTTTA AGTGGGTA
 9701 TGTGATGCT CAGATTTGT TTTTGTCTA GTCTGCTTT GGTATGTTG
 9751 GCTCTTTTT GTTCCATAT GAATTTTAG ATGTTTTT CTGTCTGT
 9801 GATGATATCT GTTGTATAT TTTTGGGAT TCAATGAT CTATAGATG
 9851 TTTTGTCTAG TATGTCTAT TTTCAAGT TGTCTCTC CTCTCTGAA
 9901 CATGGGATG GTTCTCTTT GTGTCTAT TATTTCTTT TACAGTGT
 9951 TTTTATCTT CTTGTATAG ATCTTCTAT TCTTTGTTA AGTGTATTC
 10001 TATGTTTTT GTTTTTTTT CAGTATCTT AAGAGGAT TGTGTCTGA
 10051 TTTGATCTC AGCTTTGCTG TTTCTGATAT TATACATCT TATGATTTG
 10101 TGTCTATGAT TTTGTATCT GAGATTTAT TGAATGTTT ATCAGATCTC
 10151 GAGCTTTTT GAGGCTTCA TTAGGTTTT CTAGATATC AGTATATCA
 10201 TTTGCAATCA GTTGCAGTTT GATTTCTCT TTTCAATTT GATGCTCTG
 10251 TATCTCTCT TCTGTCTGA TTTCTCTGT TATGATCTT AATTTTTTA
 10301 ATTTATATCT GTACATGAT GATCAGATA TTTATGAT CTATCTGAT
 10351 TTTTATACA AGCATATGT GATCAGGTA TACAGTAT AATTAATCAG
 10401 GGATATCTGG GTATCCATTA CTTCAACTT TATCATTTT TTTGTGTAG
 10451 GATCATGCT ATTTCTCTT TATTTTAT TATTTTTT TTTTGTAGC
 10501 AGATCTCTG TCTGTCTCT AGGCGATTA CATGATAG TATGTATCT
 10551 CAGCTCTGCT GATGCGGAGA CTCTCTCTA AATATATTA AATTAATAT
 10601 AATTAATCT GTACAGATA ATGCTCTGA GTGATTTCC ATGTTTTCT
 10651 TCTAGTCTGT TGTGTCTCA GGTTTATAT TTAAGTCTTT AATCATTTT
 10701 GATTTATGT TTTTATATG CAGAGATAG GGTATATTT TATTTCTCT
 10751 GTATAGGAT ATTCATTTT CTATGATCA TTTAGAGAT TATCTTTCT
 10801 CCAATGTATA CTTCGGTCTG GTTCTAAT ATGATGTAG TGTAAATGA
 10851 TGGATTTAT TCTGGTCTC CTATGCTCT CTATGCTCT TGTATCTGT
 10901 TTTATATCT TATATGCTG TTTTGTAT TATCATTGT TATATATTT
 10951 TGAAGGATG TATGATTT CTTCTAGCT TCGTTTTT TTTTTTTT
 11001 TTTTTTTTT TTTTATAG ACAGATCTA GCTCTGCTG CTAGGCTGA
 11051 GTGAGTGGC GCAATCTTGG CTACATGCA CTTCTGCTC CTTGTTCAA
 11101 GTGATTTCT TCTCTAGAT TCCGAGGAT CTTGATATC AGTCTCTCT
 11151 ACCGCGCTG CTTATTTTT GTATTTTAT TATGAGGCTG GTTCTCTTT

FIGURE 3C

WO 02/48366

PCT/US01/47559

11201 GATGGCCAGG CTGGCTTTGA ACTGCTGACC TCAGGTGATC CGCCGSCCTT
 11251 GGCCTCCAAA AGTGGCTGGA TTACAGGCGT GAGTCACTGT GCUCAGCTC
 11301 CAGCCTTGTT CTFTTCTCTC AGCAATGCTT TGGCTCTTCT GGCCTCTTGG
 11351 GTTCCATATA AGTTTAAAGA TTAAAGAA AAAATTCGTG GAGGAATGTC
 11401 ATTGTAGATT TGAATAGAAC TGCAATGAAT CTGTAGATTG CTTTGTGATG
 11451 TATTAAATTT TTACAGATAT TGAATCTTCC AATTATTGAA CATGAATATAT
 11501 CTTCCTCATTT GTSSTGTGTT CTTCTPCAAI TGGGTGCTTC AATGTTTGTG
 11551 AGCTGTGAGA CATCTTTCAC TTCTCTAAGT TTATCTTAGT GTATTACATC
 11601 TGTAGCTATT GTAAGTGGGA TTATTTCTCT GGTTCCTTTT TCAGATATTT
 11651 GCTGTTGGCA TATAGAAATG GTACTGATTT TTGTATCCTG CAACCTCAGT
 11701 GAATTTGCTT CAATTCTGAT AGTTTFTTGG TGGAGTATTT AGGCTCTCTT
 11751 CTATTAAGG TCAATGCTAT TGTAAAGAGG GACAGTTTGT ACTCTCTGTT
 11801 TTCTAAATTG CANGCCCTTT ATTTCTTACT CATGCTTAAT TGCTCTAGTT
 11851 GGTACTTTCC AGTACTTTGT TGAATAGAGT TGGCAGAAAT GGGCATCTTT
 11901 GTCTGTGCC AGATCTTTGA GGAAGGCTTT TCAGGTTTTC CCTGTACAGC
 11951 ATGATAGCTC TGTGCTGTCT ATATATGGCT TTATATCAAT TGAAGTATGT
 12001 TCTTCTCTTA CATTTTTCTA GACTTTTCAI GAGCAGCTGT TGAATTTAG
 12051 TAAATGCTTT TTCAATCATTA ATTGAATAGA TCAATTTATT TTCTCTCAAT
 12101 CTTTGTAAAT GATGATATCAC CTATGATAGT TTAATATAGT TGGACTATCC
 12151 TTCTATACCT GGAATGAATCC CACTTGACAA TGAATGAATGA TTTTATTGTT
 12201 TTTATTTTTT TTAGAGGGA GTTTTGTCTT TGTGCTCCAG CTGTGAATGC
 12251 AATGGGCGAA TCTTGGCTCA CCGCAACTTC CGCTTCCGCG GTTCAGCGGA
 12301 TTCTCTGCCC TCAGCTTCCCT GAGTACTGCG GATTACAGGC ATGGCGCACCC
 12351 AGCCTCTGCT AATTTTCTAT TTTTATGCTGA GACGGGGTTT CTTCATGTTG
 12401 GTCAGGCTGG TCTTGAATCT CTGACTCTAG GTGATCCACC CCTTGTGGCC
 12451 TCCCAAGATG CTGGATATAC AGGTGAGGCT CACTGCGGCG GGCCTGAGAA
 12501 TGAATCTTTT TARGACCTCC TTCTTGAGAG AGGTTTGCTA GTATTTTGTG
 12551 GAGGATTTTT GCATCAATGT TCAATCAGAGA TATTGTCCCA TAGTTTATTT
 12601 TGTTTTCTCT CATGCTAGTT TTAGGTATTT TTCTCTTAAA ATAAACAAGG
 12651 CATTTCTCTC CTAAATGCTA AGCATGCTTT TTAGGAAGAA TATGGAATAT
 12701 TCGAATATAG ATAGTAACAA ATGAGATATC ACTTAABATC ATTACCTAAT
 12751 ATAAATTTTA TTACATTTGA GGTCAATATT TATTGTTTTT CAGAGTTGAA
 12801 ATTACCTCTC CTATCAATGT TATATCTTAC TTGATTTTTT AAAAAATTA
 12851 CATGCTTTTA AGCCTCTAGA AGTTGTACCA AGCTTTCTGC TAGGGGCTGG
 12901 GTATAGTGG TGGTGAATAT GGTGAGAGAA ACAGCTTTTA AGTTATATCA
 12951 AATGTGCGCC AATTTTTTTT TTTTCTTACG AGTCTGTGTC GCCCAGGCTG
 13001 GGTGTCAATG GTGGGATCTC AACTCACTGC AACCTCTGCC TCCCAGGTTT
 13051 AAGCGATTTCT CTTGCTCTAG CCTTCTGAG TGGCTGGGAT TCACAGGACA
 13101 GGCACGCTGG CCACCTTAAT TTTTGTATTT TTATGAGAA CAGGGTTTTG
 13151 CCAATGTGGC AAGGCTGCTC TCGAATCTTT GACCTCAAGT GATCTGCCCA
 13201 CCTTGGCTTC CCAAGTGGT GAGATTACAG GCATAGAGCA CCATGCGCTG
 13251 CTTCTGCTGG TCTTTTTTGG ACCTTATATT ACTATGCTTG TTCCATTTGA
 13301 TACTAGGCTAT CACTCTCTGC TTCTTGAAAC TGCTCCATTG TCATATGTTA
 13351 CAGCTATGTC TCTTCACTCT CTGATATATT TCTTACTGTT CTTTATATCT
 13401 TCTCTTTTTT TTATCAAAAG AAGAAATACC TTCACTATTT TCTCTGTGCT
 13451 CTTACTTAAA ATGTGGGCGC TTCTTGGGGT TCTGTCTCAG CCACTGCTG
 13501 TTTTCACTCT TGACACTTCT GATATATCTA TCTACTCTGG TGGTTTCAGA
 13551 TATCAATTTT CTTGTATATT GTTTATATCA TGCTCTCAGG ACATATCAGA
 13601 TTTTCTATCT AAGAGGCGCA GTTTTGTAGA TAGACCTGTA AGTAATCTG
 13651 AGCTCTCTGG GGGGAAGTGC TGGGAAGCCA CCACCTTTAA CTGTAGACA
 13701 ACCAAGCTAA ACATTTGGAA GTTGTCTTGG ATTCTCCCTT CCGTATTTTA
 13751 TCAAGCTCTT CCAATTTTCA AATCTTGATG CCTTAATCAG TTCCCTCCCC
 13801 TCCAGATCTT ATACCTGCTC ATCTTTTATT GCGCTCATTT CTGTCTGAG
 13851 CTGATTCAGA TAGCTTCTCT TTAGATGCGC TTTGCTCTCT CATTTATACC
 13901 TTAGGAATAT CACCAAGATG ATAAATCTGC AGTGAGTCTT AAGACATCTC
 13951 TGGCAGCGGT ATAACTTAAA TTTTGTATTT TCTTCTCAT GTATATCAGA
 14001 TTCAATATCT CTATCACTCT TTCCGTGGGG AATGTTCTGC TTTGAGAGCA
 14051 TGTGTATATC GTCTTTATAT TTTTGCACCT TGGTCTATTT ATGCTTTTTT
 14101 TTTTTTTTTT TGGTTACATC TTGCGCAGAA TAACTTAAA ACTTCAATCT
 14151 GATGTGTGCA GCTTAAATAT CTTTTATGAG CTCCCAAGGG CTTTCAGAAAT
 14201 TAAATATAGA CTCTTTAGCA TGGAGCTGG TCTTTGAGTA CTTGTAGCTT
 14251 GTTTTCAAT ACACCGACG TGGAGCCCAT GCACTGGTGT TACTGACATC
 14301 GATATATGAG ACCCAATATG CCGCAAGCCT TGGAGCTTTT GTACAGGCTG
 14351 AGCATCTTTT TCCAGCCTAT ACCTCGGCTT GTTAACCTCT GTTGTGTCTT
 14401 TCCAGCCTTC CTCTTGGAG TCGATATATT CCGACCTCCC AAGCTCCCTT
 14451 GGAATCTGTA GTTTCACCT CCAATCTGTC CTATATCTAA TGAATTTGTT
 14501 TGTGTCTGTC TCTGTCTCTC TGACTTTTGA GAGCAGGACA GTGATATAGG
 14551 ATGTTTGA CAATGCCCAG TACATATATG ATCTTAATAT TTGTTGACTA
 14601 TTATCATGCT TGTATTTGTT ATTAATATTA GAATGACTGT TTAACTTTTT
 14651 TTAACTTTTT AAAAATCTTT GTTTTATATA GCCTCAGAGA ATAGGTTCTC
 14701 CTAGCTCTTA CATGAGAT ATGCTCATCT TTTTGGGATT TTTTCTGTGA
 14751 GGAATATTTA CAGCAGCCAC CTGAGGCTGA AGTAATCTTT TAAATTTATT
 14801 AACATGATCT CCAAAATCTC TTCTCTTCTA GTTTTGGAGG AAAATGTGGG
 14851 CTTTTCCTCT TTGACCGGTT AATTTCTCCA CCAATATATT TCAATATTTA
 14901 CCAATATTTT ACTGCTGTGC TTTTCCAACT GTTAATCTCT CTTTACTTTT

FIGURE 3D

WO 02/48366

PCT/US01/47559

14951 TTTTGGGAGG GGGGTGGCGT GGGAGTGTTT GAATTTGGAC TTGTCACTGG
 15001 GCATGTTCTA GCGAGAGCTC GCGACTACT CTGAGTAAAA TGGAGAGAT
 15051 TTTTAAACCG ACAGGTTTAG AAAAGATGAT GCTCTGAGCC TGCATGACTC
 15101 GGCATATATA CTTTGAGGTT CATTTATGCA GCTGTACTTT CCAAAAACAG
 15151 GTTCTCTGTC ATTTGGGCTA AGTACCTAGA AGGGCTATTC TTTATATAGT
 15201 CTAGAGCTGAT TTATCCCAAA TCTCCCAAGG TTGGAACITT TAGRAAAGAC
 15251 CTCCCTGGCC GACCAACACA CTCAGAGAT AGCGATTTT GTTATATGCG
 15301 TGTAGATAG GGGATGGA GGAAGGAGAG ACTATCTATG GTAAATATCT
 15351 ATACCATCTT GAAAGGAGTA ATTATGATTA ATGTACAGTT TACCAGATCC
 15401 TAGAGGATTA GAGTTTAA GATATATACT ATGTTTCAT GARGGTTTT
 15451 ATAAAAAGT TATTTATAG AAAATATAG TATGTAGTT GACACAGCCT
 15501 AAGACATTT ACAGTACATA TTTCTGATA TATTTATGCA CAGCTGTGTA
 15551 ATTGTACTA TCTATACATA AAAATATGAT GTTTAGCAGT TGCCTATGCC
 15601 TGTATCCCA GATTTTGGG AGGCTGGGTG GGCAGATCGC TTGAGCTCTG
 15651 GAGTTGAGAC CAGCCTGGCC AACATGGTAA AACCTGTCT CTACAAAAAA
 15701 TGCARAAAT AGTTGTGAT GTTGGCATAT GTTTGAGTC CAGGCTACTC
 15751 GGGAGGCTAA GGCAGGAA TACTTGGAG CCGGCGGCGC GAGGTTGTAG
 15801 TGACCCGATA TCGTGCCACC ACATCCAGC CTGGGCGAGC GAGGTGAAC
 15851 CTCTCTCTAA AAAAABAAAA CAAAAAABA AACAGCGGG CGGGTGGCT
 15901 CACACCTGTA ATCCCAATAC TTGGGAGGC CAGGCGGGT GATCAGAG
 15951 CTCAGAGAT TGAGACATC CTGACACACA TGGTGAACCT CTCTCTGAC
 16001 TAAABATACA AAAATAGCT GTGCTGGTG GTACGACCT GTAATCCAG
 16051 CTACTTGGGA GCTTGGGA GAGATCTC TTGAACCGG GAGTGGAG
 16101 TTGCACTGAG CCGAGACTGC ACCACTACCC TCCAGCCTGG ATACAGGGTG
 16151 AAGCTCTGTC TCAAAATATA AAGTCTTT TGAATATATA GAGCATGTC
 16201 ATGAGTATG CTAAABAAA ATATCAGGG GTTTTITTT TTTTITAGT
 16251 TTACTGATT CAGATAGAAA TCTTAAABA ATTAAATTAC ACATTTCTTG
 16301 GCTTCATANT CCAAGTACAA CAAATTTGAA CTCTCTCAGA TGATGCAGT
 16351 TGATTATGAC ATTCATACAT TCAATGATTT GATATACCT GTTTTGTGTA
 16401 AGGTTACTG AAGTCTGTA ATACATTTT GGGCTCATGA CTTCATCTA
 16451 GCTTCTCTAA TCGCCAGCG TCGTTTTAT AATCTGTCAG TTTAATATAC
 16501 AATCTGTCT GGTAGACCAT CACTCTACCC ATTAAAGTA CTTGACCTTT
 16551 GTATATAGTA ACAGCCACG TGTATTTATA TTTATAGTGA TCTAGGCACA
 16601 GTCTTTTAT GTATCAAG TGGACTATGT TAGCACAGT CCTATGTAT
 16651 TCCCTTTGA ATGATATAG ATTTTTCAT GCTTACAGC AGCTTATCAG
 16701 TGAAGATATA GCTCATATAT ACAGATGAGG AAACCGGGGC TTGAAAGAT
 16751 TAGCAATTT GATGCTACT CTGACAGTAA GCTGCACTGT TGTATTTGC
 16801 ACCCAGGCTT CTTGACTCC TCCAGTCTCT AGCTTTTGG GGAATGAGG
 16851 TAGTACTGT TTTGACCA TGTTTAGAT AGTTAGGTT GACAGGAGC
 16901 CCAACCACTA GCTAGTAGG GTATCAAAA TGTGATGAG CTGTAGCAA
 16951 GCTATGAAA AAGCATTT GTGATGTT CATATTTGT TATCAGTAT
 17001 TCAAGTGTG ATAGCTATTT TTAATTTTG CTTCTGTTT AATTTTTTT
 17051 AGGTATGTA TCTTCTGTT TATTTGTA CATTTTTC CAGTTGAGC
 17101 AAGGGAGCG TCTCTTTT AAGACAGG AGGAGTCC CTATATAGA
 17151 AAGGCTGTT TATCATTT TATAGACTA ATGTGCTTAA TATATCTCT
 17201 TTTTTTTTT TTTTTTTTT TGAAGCGAG TCTGCTCTG TCTGCCCA
 17251 GGTGAGAGT CAGAGGACG ATCTGGCTC ACTGACTCCC CACCTCCCA
 17301 GTTCAAGTG ATTTCTGTC CTGACCTCC CAGATAGCT GACATAGG
 17351 CCAATGCGAC CAGCCAGC GATTTTTGT ACITTAGTA GAGATGGGT
 17401 TTCAAGTGT TGACAGAA GGTCTGATC TCTTACCTC GFGATCGCC
 17451 CGCCTTGGCC TCCCAAGTG CTGGGATAC AGGTGTGAGC CACTGTGCT
 17501 GGLCAATATA TCTTATAT CTTAATTTT TGTTTCTTT TCTTTTTTT
 17551 TTTTATTTG TGTGTTGCT TTTTATGGA TGGAGCTCA CCGCTTGGC
 17601 CAGGCTGAG TGCACTGCT CACCTTGGC TCACTGCAAC CTCACCTCC
 17651 CAGGCTCAG CAATCTCCT CCGTACGCT CCGGATAGC TGGGACTATA
 17701 GGCAGCTGCC ACATACAG GCTAATTTT GTATTTTAA TGAAGAGGG
 17751 GTTCAACAC ATGCGCAG CTGCTGTA ACTCTGACC TGAAGTATC
 17801 CGCTTGGCT GSCCTCCAG AGTGTGGA TTAGAGCAT AAGCAGCA
 17851 GCGCAGCTG GCAATCTAC TTTTATAG TGAACCTGT ATATATTCAG
 17901 TATATGTCAC TAATATAGA CAATATATA ATCAATATG GCGTTTTAA
 17951 AAGGTTTTA CTATTTCAA GCGTTTTG TACAAATTT TGTTTTTTG
 18001 TGTGTGCTT TGTGGCAG CTGTTTTG GGGCCAGTA CTCTCTCAC
 18051 TCTTCCCGC CATGTGTG GAGAGGTCT TAAGAGGTG GGTTCCTGG
 18101 TAAAGAGAG CTTCTCTCT AAAAGTCTGT TAATCTTAA ACATTTCTAT
 18151 TCTGTTTTAT GTGTTTTGA ACATGATATA ATGCTGCAT GCCACAGAG
 18201 TCAAGCTTT TACATATAT TTTAAAGA AATGAGGAT GAGACATAA
 18251 TTTAATTTA ATTTATAGT AGTACTTC ATTTCCCTCT CCGCTCTCA
 18301 ACAGTATAT CATAGTTGT TTAGCATTT CAGATTCAA AGATATTCCT
 18351 TGAAGTATTT TTTAATCAG ATACCACTT TTAGACATAT TAATTTTGA
 18401 TGTCTGTTT GGAATTTAT ATAGCTTAA TTCTTAATT TTTAAACTA
 18451 ATGTGCAAT TTAAGCAA AAAAGCTGT TGTTCATAT ATCTTCTAC
 18501 TTTAGCCCT CATAGACAG TCAAAAACA AAGCTGTGT TTTGTGGAG
 18551 ATCTGCCAG GGAAGATGG TGAGCCTCTA CCAACAGGG GATTTAGCTA
 18601 AAAAGAGGA TTTGTACTG ACATATATT TTAAGATTG AGGCTACACA
 18651 CTTTGGAGG GTATGATA TATGTTGT CATATAGAT AGTCACTCA

FIGURE 3A

WO 02/48366

PCT/US01/47559

18701 GATTCAGTGA TTATTCCTTG ATTATGTAAC ATATTAGCTA AGTATGAGA
 18751 ATACAGATAG GTATAGGAT CTGTATGCC AGAGGTGGA ATTACCGGT
 18801 TTTTITTTTT CTTCCTTTT TTTTITTTTT CATAGAGAG GAGCTTAAAT
 18851 CTGGCATCCA GTTGGAGTG CATGGGCTG ATCTGGCTC ACTGCCAAGCT
 18901 CCACGGCCAA GTTCAAGAG ATCTCTCTGC CTCAGCTCC CAGTAGCTG
 18951 GAATACAGG TGCATGTCAC CAGCCGAGC TAATTTTIT TTTTATGAT
 19001 TTTTITGAGA CAGAGTTTCA CTCTCTCTGC TAGGCTGGAA TTCACTGGCA
 19051 CTATCTGGGC TCACTGCAAC CTTCGCTCC CAGGTCAAG CAGTTCTCTG
 19101 CCTCAGCTCC CCAAGTATGT GGGATTACAG GCACCTGCCA CCATGCTCTG
 19151 CTAAITTTTT TATTCTTAGT AGAGGCGGGG TTTTACCATC TTGGTCAGGC
 19201 TGGCTGTGTA CTTCTTACTT TGTGATCCAC CACCTTAGGC CTCCCAAGT
 19251 GCTGGGATTA CAGGCTGAG CCACTGTCTC TGGGCGGCTT TACCTTTTG
 19301 ACAGACCATT GGTCTCTGAA ATAAATAGGC AGTGTGTGAT GGTTCAGCT
 19351 CCTATATACA CAGTCCATCT TACGGRACAC TCARAATCCA CTAGCATCTC
 19401 TTCTACCTAG ATGGTTTCTG GTCTTTGGCT ACAGAACAG CCLAAAGCG
 19451 TTATAGCTTT TAGAGATAT TTACTTTTCA CATTTTARA GTTAAARA
 19501 AGTTAAGATC CATAAATTT TTTGGAAGG TGTTGATTT TCTCTGTCA
 19551 CCTCTAAGA CAGTGTCTAA AGGATCTGTA CATCAAAAT CTTTACARCA
 19601 TCGGAATTA TTGTATATAT TGCTGTGTTA AATTTTGTTA GAATCTGAT
 19651 GCGCCAGAG GAGAAATCTC TTGTGGAARA AAGTTTAGCT AGCAAGGAA
 19701 CAGTTTGGAA GGGTTGGGG TTGGCTCAGT AAGGAAGGG AAGAACATT
 19751 CAAATTTGAA AGATGCGGT GTATAAATA TGAATATTGG AAGCATAGA
 19801 ATATTTTCAA AACAGTGAAG OGAAACAGAT GATTGGATG GAATACAGCT
 19851 TGGCAAGTG ATCATTAGT GATTAAGTCT AGCATAGTAT AAAGCTTCT
 19901 ATAGAGCTTC ATATGTTTT TCAATCTTTC TACACAAA CTGTCTTTC
 19951 ATACCTAGAA AGATTTGGCT TGCAGTAGGC CCTATGTGAT TATTGAAGA
 20001 AAGCATAAAT ACATTTGAGT CCGTAAAAA GTTTTGATAT ACTAGTTTAA
 20051 GAGGTITAAA TCTTATCCTT TACACAAAG ACTGGGAAA TATGGCTAGA
 20101 GAGCTAGATC TTTTITGAAA TTTTITTTTT TTTTITGAT TTGCTCTGG
 20151 CAGCTCTCTC TTGTGTGTG GTTTTATCTT CTTTCTTACA AGTATGTGA
 20201 ATTGACCTTT TTCTTATTA TCACTTAGCT ACTTAGTTAA TGCATCGAT
 20251 AGAATCTTAA AAGAACTTTC TAGGAGTTTC TCARAAGCTT CCAATTAAT
 20301 CTITTCATTT AGAGAGGGCA TGGCATTTT CTTTTTTT TTTTAAATTA
 20351 TATTTTATTT TTTATTTTGG TGGGTACATA GGTATATATA TTATGGGGG
 20401 TACATGAGAT ATTTGTATAC AGGCATACAG TATGTAAATA TTACSTCAGG
 20451 GTAAATGGGG TATCCATCAT TTCAAGATT TATCTTTCT TTGTGTATA
 20501 AACATCCAA TTTTAGGTTT TGTTTTGT TGTTTTGT TTTTITGAGA
 20551 CGAGTCTTG CTCTCTCACC AGGCTGAGAT GCASTGGCG GATCTCAAGT
 20601 TATCTGATTC TCTCTCTCTC GAAATTCAGG ATTTCTCTG CTTGAGCTC
 20651 CCAAGTAGCT GGGACTACAG GCACCTGCCA CCATACCCAG CTAATTTTG
 20701 TATTTTAGT AGAGATGGGG TTTACCGTA TTGGCCAGGA TGGTCTCAAT
 20751 CTCTGAGCT TGTCTCTGC CCGCTCAGC CTCCCAAGT GCTGGATTA
 20801 CAGGCTGAG CACACAGCC TGGCCATAT TATTTATTT TCAATGTAG
 20851 AATTAATGTT GCTGTAGTC AACCTGTGT GCCTATCAG TACTAGATCT
 20901 TATTAATGTT ATTTTGTGT GCGCACTAAC CATCTCTCT CCACTAGCTC
 20951 TTCCCGGCTT CTGTAACCA TCAATCTGCT CTGTCTCAT GAGTTCAGTT
 21001 GTTGTATTT TTAGCTCTCA CAATTAATG AGAAGATG AGTTTCTCT
 21051 TTGTGTCTT GGTATTTTC ACTTAACATA AGGCTCTCA GTTCACTA
 21101 TGTGTGTGCA AATGACAGGA TCTCATCTT TTTCTGTGT TATAATACA
 21151 TTTTCTTAT CTTTTCATTC ATCTGTGTAT GGACATTAG GTTGTCTCA
 21201 AATCTTAGCT ATTAAGATA GGTCTGATA CAAATATAG CAGGCAATG
 21251 TGTGTACAC CPTAATCCA GCACTCAGG AGGCTGAGC AGAGATTTG
 21301 CTGGAAGCTG GGGGCGGAG GTTGCAGTGA GCGCAATG CACCATGCA
 21351 CTCCAGCTG GCGGACAGA GCGCACTCC GTCTCAACA ACAAARAAA
 21401 GGAATGTGTC TGCAGTAAT GTAGGATAC AGCTATCTCT TCAATATAT
 21451 GATTTCTTT TTTGGAGGG GTATATACCT AGTAGTACA TTGCTGATC
 21501 AATGAGAGC TCAATTTTA GTTTTGTGA GCGGCTCTC ACTGTCTTC
 21551 CTATGTGAT GTACTATTT ACATTCACC CACACATGA TGAGTGTTC
 21601 TTTTCTTCCA CATCTTCTC TATCTTTTG ATAAAGCTG TTTTAACTG
 21651 GGTGTAGAT ATATTCATC GTAGTTTGA TTGCAATTC CCGATGATC
 21701 AGATGTGAT GAGATTTT TCAATATCT ATGTCTACT TTGAGATAG
 21751 TCTATTCAGA TCTTTTSCC GTTTTITAA AATCAGATTA TGAATCTCT
 21801 TCTTACAGAG ATTTTGTAG CCGCTTATC ATTTTGTTA TTAATCCCT
 21851 GTCAGATGGA TAGTTTGCAG ATATTTTCT CTAATCTCTG GTTGTCTCT
 21901 TCACTTTGTT GTTGTCTTGG CTGTGAGCT TAAACATGA TGTGATCTA
 21951 TTGTGATCT CTACTTGG CTTTGGTGG CTGTGTCTT GAGATTTTA
 22001 CAGAAATCT TTGCCAGTC CAGTGTCTG GAGATCATC ACTATTTTA
 22051 TAAATTAAT TTTATGGAA CACAGTCRAA CCAATTCAT TACATACAGT
 22101 CAGGAGCTT TTTTTTTTT CTTTCTTCT TCTTTTCT TTTTITTAG
 22151 AAGGCTATC ACTGTCTGC TGAAGATGA GTGTGTGAG ACAGTCTAG
 22201 CTCACTGCA CCGCTGCTC CCGGCTTAA GCGATCTCT TGCCTAGCT
 22251 TCGTGTGAG CTGTGATTC AGGCGGCTG CACACAGCT GCGTAATTT
 22301 TGTATTTTGA GTAGAGATGA GTTTTGTGA TGTGTGAG CCGTGTCTG
 22351 AACTCCACAC TGAAGTATC CATCCGCTC TTCTTCTAA AGTGTGGA
 22401 TTACAGGTT GAGCAGCTC ACTGTGCTC TAAATGAT TTACTTGA

FIGURE 3F

WO 02/48366

PCT/US01/47559

22451 ATGAGCACAT TTTTGTAAA TTCTGCGCTA TTGGCAGGAG AAGATAACT
 22501 GAGAAAGGG GAGCATCTCT GATCCTCTTA AAGGTTCTC TTGCAACATC
 22551 TCAGRAAGTA TATTAGCAT AATGTTCTCT CTTAAGGGA AGACCTTCC
 22601 TACCTTCCTT ATTACCCACA TTCCATCTCT CTGTGTTAT TACTGAGCGA
 22651 TAGCATTTGA TARTAGAAGC ATAGTTTCTT AAGTCAACA GGAACCTCAGT
 22701 TGCTCATAT GTAAAGTGAT AATATTACT AATTCAAGT GTTGGAGTTA
 22751 AAGAGGAGTA CATAGAGCT GTAAATATG TACGTGCTGT TTAATTTTC
 22801 AGTTGCTGGG AATTGCTTT TCATTGTATG CATTCAGCG GTCTCTTGC
 22851 TGCCCACTGC AAAAAATTGA TACCACATGA TTGAGAACCA AGCCTTGGAA
 22901 AGGATAGAA TACTKATAT ACATTTTCAAT AGGTGGGAT TTCTTTCTT
 22951 TATAGAACTT TTCTAGACT ACTCTGTGGC AATTAAGAT TACTTATTA
 23001 TTTTCCAT CTCCATCTT AGATAATTA TCCATCTGAA AAGAAATAT
 23051 AAGTCAGTTA TTTTGGGGA GAGCATAGC ATAGTGAGTA AAAACATAGG
 23101 CTTTCAAGTC TGATCTCTTA GGTTCAGCTT CAGCTTTGCC ATTTACTTAC
 23151 TGACTGTAT CTAGACAGG AYTGTTAACU CTGCAATC AGTTTCTGA
 23201 TAGGCGCTT ATAGAGGAT AATAGAGTA ATATATGTA AGTGCCTAT
 23251 GTAGTGCCCT GTGACATATT AAGTACATA TAAATATTG GTTATTAATG
 23301 GTCATATGCA TGTCATACAA ATCTGAATAT GTAAATATA TCAGATTGTA
 23351 GTATGAATGG ATGTTCAAAA AGGTAAATGT AGAAATTTTA TTAGACTGA
 23401 AATATAGCAT GTGATTTTAA TTTTGGTTT TATCTTTAG GATATCATG
 23451 AATCTCTCC TAAATKCA TTCTGATGA AGGCTTAT TCAGGAGTA
 23501 AAGTCAGAT CAGTCATACA TATATATGT TATATATACA TACATACACA
 23551 TACAGCGACA CGGATGAACA TACATAAATA CATATATATA ACTATGCGAA
 23601 TATATGTTTG TTTACTGATT AATACCTTAA TTTTATATAT TAGMTGGAGT
 23651 ATATTGTGAG AATATACCTA ATACTCTTGA GCTGTAGGT GTATATATG
 23701 TCTATATACA ATATAGCCAA TCAATGATTA TAAATACCA CATCATACA
 23751 GATTCCTGT TTTATGTTG GCAATTTAAT TCCAGAGATC TCAATGATT
 23801 TGTAAAGACTA CAGATTGAAG AGGAAATAC TATACTAATA AACCCCATC
 23851 AATATAAAC TGATTCAGTA GGTAAAGAG TTGGTTATTA TATGAATCTT
 23901 TGCTCTTTC TTTTCTGAT TTTTCTTCT TTTTCTTCT TTTTCTGAG
 23951 CAGAGTCTTG CACTGTACC CAGGCTGGAG TGCAGTGGC ATGATCTCAG
 24001 CTGACTGCTG AGACCTCTG CTTCCAGAT TCAAGCATT CTCTGCTC
 24051 AGCTCTCTTA GTAGCTGGGA CCACAGACAT GTGCCACAC ACCCGGCTAA
 24101 TTTTGTGAG TTTTAGAGA GACAGGTTT TCCATGTTG GCGAGGCTG
 24151 TCTGAACTC CTGACTTAAG TGATCTGCC TCGACCTCC AAGTGTGTA
 24201 GATTATTGCT GCAAGTACTG TGCCAGCTC AATCTTTTA CTCTGATTC
 24251 AGTTTGACAA ACAAGTTTC GAATAGGTA ATTAGCTGT TTTATATATA
 24301 TATAAATATA GTTTTAIGT TTTATATATA TATATACACA CACACATACA
 24351 TATACACAA CATACACAA CAGGTATATA GCATATGGA TTGCTATGT
 24401 GGATTTATGT TGATTGAGAA TTTATTAGG CAGTTCATAT TTAATACCTA
 24451 CTTGGAGGCC CACAGATGAT TCTAAACTAT CTTGGGAAAT TTTAAATTTA
 24501 TATATAGATA AGCATTTGT TATTTAAAG TTTGTGATAT ATTACTTTA
 24551 GAAACAAAG TGTGAAAT TTTCTATAG GAGTAATGT ACTTATTTT
 24601 GATTCATCT CATAGAGTG TGTATATCA TTTTTCATT TAAATTTT
 24651 AAGGTTGCT GTATCTCTT AGTCCCGCC TTTTGGTGC TCTTCTGAT
 24701 GTTGGGCCC GAAATCTTT CTGCTGCTA ACGGTGTTT TCACATGTC
 24751 CCAATTTCT TTAATGAGA TCAAGCCATG GTATGTGAC ATTTAGATTA
 24801 TACACACAA ATCTATCTT CAGCTATGT TTCTATAGT TTCTTTTCT
 24851 TCTTCACTT GCGCATCTT GGTATGAGG TTTTAAATTT ATTTTCTTA
 24901 TACATTTAT CTTTACACA GATAAGAGG CTTCTAATA TACAGTAGT
 24951 AATAGTTCTT AAGATTTTA GAATAGTGT ACATCTTGG AGCATCTAG
 25001 ACTCTGCTT AACTTATTT TGGATCCAG AACCAATTT TCAAAAGAT
 25051 ATTTGGGGA CTGATAGGA TTGCTATCT TTTTCTTGT ATGAAATGT
 25101 AAAAAAGCA ACAAATTTG TCATTATCTA TATTAGTTAG GATAGGTTT
 25151 TGCTATATAT CAAAGCCAA AATTAAGATT GCTTAAATAT CARATTTCT
 25201 TTTTCAATG AAAAAAGGTC TGCAAGTTGA GCGTCTGGA GCTGTAATG
 25251 CAGTTCAAGT GAGCAGACT TTTCTATCT TCGCTCTCT CACATCTCA
 25301 CTCTCAGCT CAGTACAAA GTGGCTCTT GAGCTCTAGC CATCAAGTCA
 25351 TACTACTGAT AGCAAGTGG AAGGAGTTG TAAAGAGAGC ACATCACTT
 25401 CTTTAAAGA AACTTTACAG AATCCCATC CCACCTCTGC TTGCTTTCT
 25451 TTGCTCAGT TCAAGGATA TGCTGCTCTT AGTTCAGGA GTCTGATCT
 25501 AAGCCACCT TGAATTAAT CATTATTTT TTTTCTTCA CTGAGGAGG
 25551 AAGGGAAGA AGCTAACAA AAGGGAATA ATTTGCTTA AGTCAGAGC
 25601 CAATTTGCT GTAGTTTAT AGATTATTC TGACTAATGT TCTTTTACT
 25651 CCACTGAGTA TTAATGTTAG TTCTACATCA CGGATGGCT TTAATTCAT
 25701 GGTATTTCT ACAGAGGCC AAGGTTTAT ATGGGAGG ATCTCTCT
 25751 TGTAGTAGC ATCATCTCT AATAGACAA GCTTTGATG AAGGCACTC
 25801 ATGACATGA ATCTTACCA AAAATGTAC ATGGAATGAA CAAGAAATTT
 25851 GCTGTGTTA GCAATGCTT AGAAAATTT TAGTAGTAT AGTCTTGTG
 25901 GAATGATAT TTGAAGCTC ATATGATTA TATCTGATT TCGAACTA
 25951 AATATCAGTA TCAATAGGA AACTCTTAA TCCACATAT TTTGGAATTA
 26001 TTTATGCAA AATATAGTT ACTCTTAAA TATTTTCTC ACTGTTATA
 26051 GTATCTGCT ATCTCAGTT TTAATGATC TGAATCTGC CATATAGGTT
 26101 TTTTGTGCT AATCTGAGT TATTGTGTT GCGAGCATT GAAAGCTAA
 26151 AGATACTTA TGAATCAT TTAGTACTA TATAGGCTT CAATATAT

FIGURE 3G

WO 02/48366

PCT/US01/47559

26201 ATGAAGAANC TTGTCCTCA GTCCTGTTA CTGGATTGG TTTTATCT
 26251 AGTGTGTGG TGAAGAACG TAATGGATG ATATATATG ACTTAGTTT
 26301 GTGCTGTTT TTGGCAGAA AGCAGAAAT ACTTACAGT TCACAGTTG
 26351 GTCAAAAGAT TTGCCCTGAT ACATGATCA ATTTAAGTA TTCCGTATA
 26401 GTAGCTGAGC TGCTATTGAA GGGCTCCTTC TAGCCCTATG TTTTCATAAT
 26451 ATCTGTGGCT TATATTATG GATAGTTAA TCCATGAAT ATCCTAGTAA
 26501 GCTGTGAGCT GAATGAGCT GCTCTGAGC TGAATTAAT TATPAAATG
 26551 AAGAAAGATT AAACCAATGG TAATTCCTT AACCTTCTT TGTATCTTT
 26601 CATTCTAGG TGGTACTTG CTGTTATCTC TGTTCCTGGG GTTTTTCAG
 26651 TGACTTTTTC TGTGGTATT GCATAGTAG CAGATATAAC CCAGAGCAT
 26701 GAAGAAGATA TGGCTATAG ACTGGTATGT ATGTTATTC TATACCTTT
 26751 GTATCTGCTG AGAATGCTT TGTTTTAGG ATAAATATA TTATAGGAG
 26801 TGCAGACCTT TGCATACAA GATTTTGGG TAAATATAT TTGTAAATBA
 26851 AATCATTCAT TAGACTACAT TTAATMTTT TTTGCGGTAT GAGGCTATGT
 26901 AAGTTTIGAT TCTTTTCATT TAGTAGATAT TCATAAGTCA CAGTCAAGAA
 26951 TTGAATATAT AGTATATTTT ACTTGTAGA GTTCTTTTA ACAGATCCT
 27001 ABAATPAGA ATXTTTTGT ATCTCAGAG TTAABAAA TCTCTATCA
 27051 TTTAATGTCT AATCTPAAAT ACACAGGCT AACATCAGC TCAGGGATCA
 27101 GCAACCTTTT TTTGTGGCTT TGTGGGCCAC GACAGTCTC TGCTCATTC
 27151 TTTTGTTTT GCATGTGTAT TTATGTTAT AAACCTTTA AAAATGTAG
 27201 AACAGCCAG ATTTGAGCA TGTGTATAG TCCCACTIC CTGAGACTG
 27251 TTTTGTAAA CTAATATAG GCAGTAAAT GTGCTACTA TCAAACTAG
 27301 GAATTAAGG AAAAAGGCT AGTAATAGA TGACTACTAT AGGCATAAT
 27351 ATAGTCACT ACTGATAGC CAGAAATAGG ACAGTATATC ATTTUGTAA
 27401 ATGTGAGACA AATACCTTGT GATAAATAG GATTAATAT TGTGTGGGC
 27451 TGAATATGT TTAAGAGGA CTGATCTTG ATCTAAGGA GGTATATAG
 27501 AAGATCATA AGATTTTTG GAGGAAACA CCTATAGGA GAATTTAGA
 27551 AAAAGAACIA AIAATTTCTG GCCTGTTGAG TGGCTCACAC CTGTATCTC
 27601 AGCACTTTGG GAGGTGAGG CAGGCGGATC ACTGAGACT AGGATTCAC
 27651 GACAGGCTG GCAACATG TGAACCTTT TCTCTATTA AAAAABAAA
 27701 AAAAABAAA GTGAAGAAA AAAGAACIA TCAATTCAG TGTAACTTC
 27751 GACACTTAAA TGATACAGG CTGATGATAG CCAGGATAT TAAAAATAG
 27801 TCAATTAAG CTATAGTTA CATACCATTA AATTATCCT TTTATAGAT
 27851 ATAGTTGAT GATTTTAGT AATTTATAG TGTATGCAA ACACACCAT
 27901 AACCATTTT GGGGTGTCT GGTGTGTG TGGTGTCTT GGTTTGGTT
 27951 TTTTGAGTA ATTTATTTT CCAATAGCCA AGTTTGAAT TTAACAATTT
 28001 TCAATCTGGA GTTCTGTGT ATTAAGCCAT GTTCTGCCAA AAAACAAAC
 28051 AAACAAAC AAACCAAAAC AAACAAAAA ACACGAAAT CTTCAGAAA
 28101 TAATATGAT GCAGAAAAA GTTGGGAGG TGGCAGGCA CAGTGCAATG
 28151 TGCTGTATAT ACTACAGTT TGGGAGCCA AGGAGAGGG ATTCCTGAG
 28201 GCGAGGAGTT TGAGGCTGCA GCTATGATCA TGCCACTACA CTCGACTCA
 28251 GGGTACAGAG TGAGACCTGT TCTCTTAAA AAAAAGTTG GAGGGGCCAG
 28301 GTGAGTGGC TTATATCTC AGCACTTTGG GAGGCTGAGG CAGAGGATT
 28351 GTTTGAGCC AGGATTTTGA GACTAGCTG GCAACATAG TGAAGCCCA
 28401 TCTATACAGG AACTTAAAA ATTAGCCAGG TGTGTGTGT TGTGCTGTA
 28451 GTCCAGCTA CCTGGGAGGCT TAGGTGAGA GGAATCACTT GGCCTGAGG
 28501 GTTGGAGGCG CAGTGAGCCG TGATCCGACC ACTGCGCTCC AACATGAGCC
 28551 ACAGGCGAG AGCTGCTCC AAAAAGGGG GTTGGGGCT CCAGGCTGAC
 28601 CCCTGTGAT TTTTCTTGA CAGAGAAA ATGCTTCCA AGTGGAGAG
 28651 ACTGAGGAGA AGGAAATGA CAGTGAACAA CTGTACTGAC TTGCTCATG
 28701 TGTCAAAAT GTGATCTCTG TAAATGCCCT CAATGTCTT CAGTGACCT
 28751 CATAGTGAAG ACCATTTTCC CTFTCCUCC ACTGTGACCA GAGCCCTGCT
 28801 GAGGTCTGG TCCCTCTGA ACCACCTTA GGGTGTGAT AACAAATTA
 28851 CCACATCAAT TGAATATATA TATTTATATG TATGTGTGTG TGTATATGA
 28901 TGTGTGTGA TATATATATA GTTTGTTTT TGTGAGAGG GAGTCTGCT
 28951 CTGTACCCA GGTGTGAGTG CAGAGGTGTG ATCTGGCTT ACTGCAACCT
 29001 CGGCTCTCG GTTCAABAG ATTCCTCTG CAGGCTCTCC CCAATAGCTA
 29051 TGCCACCCAC CATGCCGAG TAAATTTTGT ATTTTTCCTA GAGGCGGCT
 29101 TTCAACCAT TGGCCAGTCT TGTCTTGAAC TCTGACCTT TGGTCCGCT
 29151 GCTTGGGCT CCCAAGTGT TGGGATTATA GGCCTGAGCC ATGGCGGCTG
 29201 GCGCCATGT GATATATATA AATACATTT AAAAAGCAG CACACCCAG
 29251 TTAAGAGCA TTTCCTTGA CCAAAATGT TCCCTAGCC CTGTTTCCCA
 29301 TGTGTCCCA TGTCTCACT GTGACCCCAA GCACCAACA ATTTAGCTTC
 29351 TGTCAACATG GTTTGCTT TCTAGAAAC TTATAGAAA TTAATATATA
 29401 CAAACATCT TTTGTGCTA ACTTCTTCA CTGCGATAA TCTTTGAGA
 29451 TTGATCCAG TTGATACAT AGATCAATAG GTTCTATTT TGTCTCTTT
 29501 CTTTCTTTT TTTTGTGAG CAGGCTCTG CTATATCCC CAGGTGAGG
 29551 TGCATGGGCA TGATCATGSC TCACTGAGGC CTGGCTTCC TGGGCTAAG
 29601 CGATCTTCT GCTCAGGCT CCAAGCAGT TGGACCCACA GGCATGATCC
 29651 ACCATGCCCA GCTATTTT TCTTTTGA GACAGGCTCT CACTCTATTG
 29701 CCGAGGCTG AGTGAGTGG TGCCTTACA GCTCACTGA CCTCTGTCT
 29751 CCGCTTACC TCCCTGCTC AAGTGATCT TCCACTCAG CCTCCGAGT
 29801 AGCTGGGACT ACTATTTTT GTATGTTTG TAGTATGGA GTTTCACAT
 29851 GTTGGCCAG ATGGCTCAA ACTCTTAAC TCAATGATC TGGCTGTCTC
 29901 AGCTCCCAA AGTGTGGA TTACAGGCA TGAACATGC ACTTGGCCAG

FIGURE 3H

WO 02/48366

PCT/US01/47559

29951 TAGTTTITTT TGAITGCTGT GTAGTGTAT CTATCCATC AGTGTATGGA
 30001 CATTCTGAT ATAGCTAAT GTTGAAT ACTAGATTT TATGATCTG
 30051 TTAANTATN TGACCTTGA AATGTGANT GCTTGCCTTA AGAATAGAG
 30101 TTGCAAGTAA GCATCTTGT GAAGTTTAA TCTTCCATCC AAAAGTCAA
 30151 AATGGCATAG AACAGAAATA AATTCACAC ATTAATCTCT ATGCTTTGAA
 30201 AGAATATGCT CTTTTCCTT TCTTCCCTT CCCCCTTCTT TCCCTTTGCG
 30251 CTTCTCTTCC CCTCCCTTC CCTCCCTTC TCCACTTTC ACTTTCCTT
 30301 TCCCTCTTCC TTTTCCCTT CAGCTTCTC CTTTCCCTTC CTTTCTCTT
 30351 TCCCTCTTCC TTTCCCTTCC TTTCCCTTTC CCTTTCCTT TCCCTCTCC
 30401 TTTTCCCTT CCTTCTTCT TCTTTTCTC TTTCTTTTC TTTCCCTTCC
 30451 CTTTCCCTT TTTCTTTCG CTTTTCCTT AAGCTTTTC CTTTCTTTC
 30501 CCTTTCCTT CTTTCTTCT CTTTCCCTT CCTTTCCTT TTTTCCCTT
 30551 TCTTCTTCC TTTCTCTTC TCTTCTTTC TCTTCTTTC TTTTCCCTT
 30601 TCTTCTTTC TTTTCCCTT CTTTCTTTC CTTTCTTTC TCCAGACAG
 30651 GTGTTGACCA GACTGGACT ACCTTTGGA TCAAGTGATC TCCACCTCA
 30701 GCTTCTGAG TATGCTGAG TATGCTGAG TCCACCTCA CCAAGCTAG
 30751 AGTCTATTT TTTGAGAGT TTTCTCTTC TCAATCTTC TAATTTCTAG
 30801 GCTGTGCTC AGTAGGCTAG AATGTCTGC GGTAAACAG TTTGCTTAC
 30851 TCGTGTCTT TCAACTTTC ATCTTCTCT TCCACAGCT TAAAGAAATA
 30901 GACTCTCTCA GACTTTTGA GATGTCTTA ATTATCTGT GATTTCATC
 30951 TCTTCTTCC TCCATATA AAGATATAC AGCTCTATC TGTCTCAGC
 31001 GCTTATCTA TATCTATCA GAACAGGAC TCCATATCT TTAAGCTTA
 31051 TTTTCTCTC CTTTCTCTA TGTATCAAGA TAAAGCTTCA TTAATTTTA
 31101 ATATTGAAA AGTCAAGAG AATTTGATC TCTTATGAG AGTGGATTT
 31151 AGCTCTCTT CACAGGAGC ACTTATAGGT TATGAGAGT GCCTTATAC
 31201 AGCTTCTTC ACCCAAGT CCAAGAGGT CACCAAGGT CCAAGAGGT
 31251 TTTTAAAGT TTTTAAAGT ATATGAGGT TTTTCTTCT TTAAGCTAT
 31301 CAGCTATGT TATGTATTT TATGTGCTC GGAAGAGGT TCTTCCAGC
 31351 TGGCCAGGG AAGCCAAAG ATTAGACAC CTTCCCTTAA GAGCCAGCT
 31401 GACTGCTAT CAGAGAGGG TGTGTCTAC GTTCCAGGT TGTCTCAGT
 31451 TCTTCTCTA AAGAGTTTA AATTTTAA TAGGCGGCT AGTGTGCTC
 31501 AGGCTCTTA TCCAGACCT TTTGAGAGG GAGGCGGCT GATCAGAGG
 31551 TCAAGAGATC GAGACATCT TGGTATCAC GGGGAAAC CCATCTTAC
 31601 TAAATATCA AATATATC CCGCAGGT GCGGAGGT TGTGCTTCA
 31651 GATTTGAGG AGGCGAGG AGGAGATC CTTTAAAGT GGGGCGGAG
 31701 CTTGAGAGT CCGAGAGCT CCGACTGCA CTTGAGCTG GCGGAGGAG
 31751 CAGAGCTCT TCTTAAATA AAAAAAATA TTTAATAGA
 31801 GCGAGGCTC TGTGTGTTG CCGAGGCTG TCCCAATCT CCGGCTTAC
 31851 CAGTCTCTC CACTTCTCC TCCCAAGT TGTGATTTA AGGATATAC
 31901 CACTCAGCT AGTCTCTTA AAAAAATC TTTTACCTAT GATTTGCTA
 31951 GTAGGTTTA TTTGAGAGT GGGCTTCA GCTTATAGT TTTGAGGAT
 32001 GTTCTCTAG AGGCTATCT TATTTTAT TTTTCTTGA GATGAGGCT
 32051 CACTCTCTA CTTTGGCTG AGTCTATCT TGTCTCTAG GCTTCTCTA
 32101 ACTTCTCTT TTTTGGCTA AAGATATCT CCGACTTGA CTTTCTAGT
 32151 AGCTTATCT ACAGGCTCT ACCAGCTCT CAGCTATCT TTTCTTTT
 32201 CTTTCTTGA ACAGAGAGG TTTCTCTAG TTTTCTAGG TTTTCTTGA
 32251 CAGAGATCT ATTTCTTAT TTTTCTCTG AAGGAGAGA GTATAAGAA
 32301 ATGAGGAGA GCTTCTCTT CTTTCTCTG TACTCTCTG CTTTCTCTG
 32351 GCTTCTCTG GAGGATCTT TGTCTCTAG AGTCTCTAG TACTCTCTG
 32401 AATATAGCT GACTTCTCT TCTTAAATA ATAAAGAA TTAGCTCTG
 32451 GTGCTCTGT GTGCTCTGT CCGATCTAG AGATCTCTG CAGGAGAT
 32501 CTTTCTCTG CTTGAGGCT AAGGATCTG TACTCTCTG TACTCTCTG
 32551 CTTTCTCTG CTTGAGGCT AAGGATCTG TACTCTCTG TACTCTCTG
 32601 CTTTCTCTG CTTGAGGCT AAGGATCTG TACTCTCTG TACTCTCTG
 32651 CTTTCTCTG CTTGAGGCT AAGGATCTG TACTCTCTG TACTCTCTG
 32701 TGTGAGGAA AAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 32751 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 32801 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 32851 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 32901 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 32951 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33001 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33051 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33101 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33151 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33201 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33251 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33301 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33351 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33401 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33451 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33501 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33551 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33601 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG
 33651 CAGGATCTG CAGGATCTG TACTCTCTG TACTCTCTG TACTCTCTG

FIGURE 31

WO 02/48366

PCT/US01/47559

33701 CAGTGGGCCA AGATTGGGCC ATTCACCTCA GCGTGGGTAA CAGAGCGGAA
 33751 ACTCCGCTCT AARATTAAAT AATAAATAA AATAAATAA AGAATATATT
 33801 TTGGCTCTACT AGACCTTACC TCCATCTCAG TGTGGCTGAT GTTCTTATAG
 33851 CATTCTTAAT GGGGGTTATT TGGTATATTA AGATTATTGG CTTAGGGAGA
 33901 ARGATAGTAT TCCCTGTCCA TATAGGTGGT TTGAGTTTGT TGGGTATAGA
 33951 ATTGAGGGGA TGATTTAACC CATTACACTG CTCAGACTTC TTGTGATATT
 34001 AGAGCTATGT TAGGTGAGGC AGCTAGCCCA ATGAGAGGCA AARACAGTA
 34051 TTTTCTCTCT TGACACTAGT CTCACATGAG GAGGATCAAG CCTTAAACCA
 34101 ATACATGTAA AATGCCAARA ATACTGCATAT ATTTATTTGT AAAATGATTC
 34151 TGAGTCTCTG ATAGATATCT CCAATTATAG ATATCCACAT TGAGGGAGCC
 34201 ACCATGGATA AGTAGGCATT TTATATATGT TAGAATATGT TGTATTTCCT
 34251 CTGGGGATTT CTGTGTTAT AARAGAAATA CATTTATTTT TTTCGTGGCC
 34301 ATTAATGTGT AATATACAG AGCTGCAAGT GGTTTCTTAT CTGGTATACA
 34351 TTTGTATTCA TTAATGTTAC CTTTCTGGA AACCTTCTTA GATAATGAAA
 34401 TTTTACACAG AAGGTGTTCG AGCGTTTATA GCAGTCTGT GATTCTTTTC
 34451 CATATATGCA CAGGTGAGTT TCTTTTATTA GTTAGAGTGA TGTCAGTAGC
 34501 CTTGGCTGGC CATTCAAACT GGGGCTCAAT CTAGCTATAG TATCTTGGTC
 34551 GAATCAACAA AGTCAGGAGT CTGAGGTGTA TAGGTTCAGA AATTCATTTT
 34601 GTTTGTGTAG CTAGTGTGTA TCAGATTTTT GGGACTGACA CATTATATAT
 34651 AGGATATATA AAGCTAAGAG CTAGTTTATT GATGTATATC AGGATATATG
 34701 TCTGGAGTAA GTTAGAGGT TATGTCTGT TAGTCTCTCT CCTATTATAT
 34751 CACTCATCTA ACAATATATT ATTGAATAC TACCGTGTGT CAGGCACTGG
 34801 GGATACAGCA GTGAACCTAA CTAAATCTCT GCTTTTGTAG AGTTTCAATT
 34851 CTGGTGGGAT AGACACAGAA ATAAACACAC CAATGTAAAC ATATCAGATG
 34901 TATTGAGAC TACACAGAAA AACTATAGGG TCATATGGAG AGTCACTTGG
 34951 GAGAACTGCT TTAAGGCAAC AGGGGCAATG GTTAGAGAAA CTTCTCTAAT
 35001 CGAAGGTGAC ATTTACGAC AGAGCCAAAG GTATAGGAAA CAAGCTGTAG
 35051 GATTAATCTGG AGAAGAGAAA TTTTAAACAG TACAACAAA ATTTCTAGGA
 35101 GAGACTGTAG CAGGCTTGTG TAGAAGACAA CAGGGGCACT GTGGGCTCC
 35151 AGTGTAGTGA AAGATTGAGA GCATCTTAAA AATCTAGTCT CAGAGGTTAG
 35201 TGAGGAGACA GATCTGGGG GCGTTACAGG TCTCTATAAG ACTTGATATA
 35251 GATAGGAAGT CATTTGGTGG TTTTGAGCAG CAGAAATGACA TGAAGTGTG
 35301 TATCTTTGCA GCCAGTTGG TAAACAGTGT TTGTGTGATG AATAGATATG
 35351 TAATCAGAC AGCAATTTGG ATTTCTGAG GAAATATTTT ATCTTGGCTT
 35401 CCAATCTGTT TCTGTTTAC AATAGGATA CATCTTTT TCCAGTGGC
 35451 ACCATATATC TCTCATACTA CAGACATTTG TCTTCTTCC TGTGAAGAG
 35501 CCCCACCCCC TTCTTGCTTA TTTGGCTTTC TGTCTTAAAG ACATATGAA
 35551 GACATCTAG TCTAAGTTAA TACTTTTCT CACCTTAGAT GTAAATCTACT
 35601 AATTAACAC TTATGTTTAT TCTAGGGTAT TTTTAATGGA CTCTGGATG
 35651 GTTCACCATT TCTCGAGTAG GATTCTCTCT CCAGGTCTCA TGATTCATTC
 35701 TGTTTTGGTG AGTTTAGCAA CAATTTCTTA ATTTAAATCC TATATGCTC
 35751 CCCAGGCTTC TGACATATCA TATATCTGTT GTTGAATATT AGTACTATTT
 35801 CTAGCTTCAG GTTCTGCTTA AATATCATCA GTCTGGAAAA ACAATGCAAT
 35851 TAATATCTCA TTCTTAGCAA TGAAGAAAGT GCTTTTATC TTGAGAGCA
 35901 AATATATCTG AGCCTTTATA TAATAATGGC TTTAAAAAAA GTTTTGGAG
 35951 CCAGGTGGGG TGGTTCATGT CTGTATTTCC AGCACTTTGG GTGGCAAGG
 36001 TCAGGAGCCG CTGAGGCCCA GAGGTTCAGG ACGAGTCAA GCACCTTGGC
 36051 ABAACCCCTT CTCTCCAAA ABAABAAAA ABAABAAAA ABAABAAAA
 36101 AGCCCGGTGT GGTGGTGTGT GCGCTGATC CCAAGCTACT AGGAAGCTGA
 36151 GATGGAGGGA TTGCTTGATC CTGGGAGGTC GAGGCTGCAG TGAGCCACAG
 36201 TTGTCCCATG GACTCCAGT ATGGGCAACA GAATGAGACC CTGTCTCAA
 36251 AAAAAAGTGT AAGGAAAAAT CACAGTTAGT ATGTGTAGAA CTGTATGAT
 36301 TATCAGAGT TACCTGACCT TTGCATAGA TGTGATCAG CTGGGCTAGA
 36351 GATTCTTTTT CATCATCTGC CACGATTAGA TTGCTAACA TTTTCTTTAA
 36401 GATTATGATG AAGAGGCGAG ACTAATATCT TAAAGACACA TAAAGATTTG
 36451 ATAGCTATTA GGTCTTATGT TCTGTTTCT ACTTAATGA CCAATGGAGC
 36501 TGATGCTATC TATTAATGT AGTCAAGTGC ACTCAAGCC CAGTGAAGAG
 36551 TGGAGGTGCT TTTTGAAGGT ATTTTCTCT TGTTTAGCAG ATGGCATCTG
 36601 GCATAGAGTT ACHAGATGAA AARCACTCTC TGAGAGTGCA GAATCTGCA
 36651 GGGCAGGAGC GTAAAGGAGA GGAATGTATG TGAATTTGTA TTGGAAGAT
 36701 CAGTCAAGCA CAGAGGCCCT GAGGCACTA AGACCCAGAG GTGCAAGCAA
 36751 CCCCAGAC ACATTTGCTG AAGCAAGCTG ACACACAG TCAATGTAGG
 36801 AACAAGTCTC TTAAGTGACC ATTTAGGTTT TGGTGTCTAT GTAGAGTAA
 36851 GAATTAACAC TTTATATCAT ATATGAGCTA CTACATATT ACATGTATAG
 36901 CTTCAAGGTG AGAGAGTGA TCTGTATCT TCAATTATGC AGAATATATC
 36951 AATTAAGATG GACTGTCTCC CTGTGAGCT TCAATATGCA TGAAGAGCAA
 37001 GTGATATTTCT CCCCCTCTCT CAGGCAAGG GAGCTGCTAA TAGCTTCCA
 37051 TTAATGCTAT TTTGCTTTAT TTTTACCCA CAGTCTTTTG TTTCTGCTAT
 37101 TCTTTCTGCT CCCCCTTCT CCAACCTAAG TAGCCCTTTT CTGGGCTCCC
 37151 ATGCTCATGT GCGCATATCT CCACTATTGT ACGAATACA GTGTATGAGA
 37201 GATTATATCT CTGTATTGT CCTTATGTTG TGGCTCTAG CGGACTCTG
 37251 AGTAGAGTAT GAGGTCAAGG TTAATATATA TTCAATTTTG TCACCTTAGC
 37301 ACCCTGAACT GCCAGGAGGT GCTAACTATA AGGTATTGG TTTTCTTCCA
 37351 TATGTATATA AAAAAAATA ACTTAAAAA TTAGTAGTAA GCACTTTAAA
 37401 ATTGGGAGT GTTACGTGAA ATTTATGATA TGTATCTCT CCCCAGAAAT

FIGURE 3J

WO 02/48366

PCT/US01/47559

37451 GATTAGATCT GGTACCGGG TGTGGGCCAA CTTTCACGG AGGCGCAGT
 37501 AGTTGCTGCT CCTTTGGAA CCTTTCTTT TGGTTTATC ATGAGGCCA
 37551 TTACTTCCCG GGCACCGGTA CTTATAGAG TTCTTACAC TTGCCTAAG
 37601 TAGTGTTTTA CATGAGACA TCAATATAGT TCTACATT TCTTTTCT
 37651 CAGTGTTTTC CTATCCAGCC CTCTCTGGG GGGTACATCC CACACTTACT
 37701 CTTCAGTCC AGTCCCTGAG CTCTATATG ACACATTGGC TGGGTATCTC
 37751 AGCTTTACA TGGCAAAAT TAAATCTGG GTCCATCCG TTCCCGGCA
 37801 CCCCATGCT CCRATCTCA TTACAGTGGT TCACCAACGC CAGTITGG
 37851 GGGCAGAGG CCTTAGCGAC ATTCATGAT CTTTCTCCG TAACCTTACA
 37901 TTGAGCCCAT CAATGATAC TTCTATCAC CTCTCTCCG AATATATATC
 37951 TTAATCAGA GGTTTCTGA TATTCTCAT TAGTAGAAC CCAATCTGAG
 38001 CGTGGCCT ATCTCCATC TGTCTCTCT GCTTCCACT TGCCTCCTA
 38051 TAGTTCATC TGCATCTGC AGAGTATTT TGTAAATG GAAATTTAAT
 38101 CGCATCTTAC CTATACTCA CTTTCTTGG CAACAGAAAT AATATCTAGA
 38151 CTCTTATTA TGTCAATTTG CTCCACTCTC CAGATGGTTG AGTATGTTCT
 38201 GTTTCATCC AGGCTCTTG CACTCTCTGT CAGCTTCTGT TGGGIGTCTC
 38251 TCTTCTTCA GTCCTTTGCA TAACTAGGTC TCTCTCTTA TGGGCTCTC
 38301 ACTCAGAGG GCTTCTCTG GATCTCTGT TAAAGAGCAT CTCTATCA
 38351 CTGCTCTGTT TATTTTGTG CATAGCAGCT ACACCTTGA ATATCTGTT
 38401 AATATAGGAG GCTCATATG ATCTTCTCTC CTAGAGAGG TACTCTGTT
 38451 GTTAGAGAT ACTTTGACTT ACTCCTCAT GCTTCTGAG CCAAGATA
 38501 AGCTATGCT TATTAGAA TTGGCAAT ATTTGTTAG AGCATACCA
 38551 AAGCTGATG TAGTCTTAG TTAGCTGTC ATCAGCTATG CCACTGCT
 38601 TAAGTATTT AACCTTTGG ACTCAGTCT GTATCTTCA AATAGAGAT
 38651 TAAATAGG GACACAAA ATTTTCTCG TTCAAGACA CTATATATT
 38701 AGATTTTGG TTCTCTTAT ATTAATATA TCGAATAT TAATACATT
 38751 AGAATATG ATGCTTTCA TAAACACTA TTAAGAACT ACTGGTGT
 38801 GGGGATATA ATGTATATA GGAAGAGT CCGCTTCA AGAAGAAAT
 38851 CAGTGTCTCC AGTTACATG TTGTATATG CGGTCTTTT TGGGCTTTG
 38901 GTCACAGA TAGCTGTGG GGAAGAGAT CCAAGAGAG GTTCTCACG
 38951 CAGTGGCTC GCTGAGTAA TTCTGCATC AGACCTTAC CCGAGAGAG
 39001 GGGGATGTT GTATCTCAT GTTCTCAAT TTAATGACT ATCTCATGGC
 39051 CTTAAGAAA DATATCTTG GCGGGCAGC GTGGCCCACT CCTATACCT
 39101 TAGCATTTG GAGGCCAAG GTGGGTGAT CACTTGAGT CAGAGTTCTG
 39151 AGACAGCTT GGCACATAT GTAAACCTT ATCTTCTTA AATATCAAA
 39201 AAAAAATTG CTGGGCTAG TGGCAGTGC CTGTAATCC AGCTACTCG
 39251 GAGCTGAGG CAGGAGATC ACTTGAACCT GGGAGGTGA GGTTCAGTG
 39301 AGCCAGATT GCGCCACAG ACTCCAGCT GGGACACAG AGGGAATCTC
 39351 CATCTGGGG GGAATAGAA AGGCTGGCA CGTGTGGCT CAGGCTGAG
 39401 ATACAGCAC TTTGGAGGC CGAGCAGCT GGAACACAG CTCAGAGTT
 39451 CGAACCAGC CTGGCAATA TGTGGAGCC CTGCTTTAC TAAATATCA
 39501 AAAAAATGA GGCATGCTG TGGGCACTG AGTCCAGCT ACTCGGAGG
 39551 CTAGAGCAGA AAAAACTCT GAACCCGGA GCGAGAGTG CAGTGGAGC
 39601 AAGATTTGA CATCTCAT TACCTTGGG ACACAGAGA GACTCGCT
 39651 CAAAGAAAA AAAAAACCA AAAAAAGA AATATATTC TCGTATTAAT
 39701 CTATCTATA AAAAAAAT GAGGCTATA ATGATTTTA AGCTTTTTT
 39751 ATTGATGTT GATTATTTCT TTGAGAACAG CTTTCCAAA AGGGGAAGAT
 39801 AGCTTTTCT CGAGTATAT ACTTCTGGG TAGATGGTT GGTGAGAGG
 39851 CTAGAGAGG ACCGCTATT ATAGATCTT TCTGTAATA TTAATGATCT
 39901 CTGCTGAGA AGCTGAGCT TTCTTGAAT AGCATCTTG AAGAGACAG
 39951 TGCTCAGGG TACATGCTG GCCATGCCA TCTTGTCCC ACTGAGCTGG
 40001 GAGTGTATC TCAGCTTCT TACTGTCTT GTCTCTTGG TGAAGTCCA
 40051 GGGATTTCT CATATGCTA ATCTTATTT TTAACAGAT TTCTTCTTG
 40101 TAATGCTTAT CATCATCTA ATGATGATG AGACAAATG TTTTCTGCTC
 40151 TGTGATCTC TAGGCCATA TAAAGATCA CAGCTTTTA GATGAAGTG
 40201 CTACCTTCCG GATGTTCTT AAAGTAGTT TCCAGAGAT TCTCAACTT
 40251 TCAATATAG TTCTCTGTC TTCTACAGA GTTGGAGGC ACAGGCTTT
 40301 GGGATACTT ACAGTTTCT TTTTCAATAT TATTTGAAG CAGTCTCTG
 40351 ATATAGTCCA CAAATAAAT CATTTTAAAT TATTCTAAC CTAATTAGA
 40401 ATCTTATCA TTCTTAATCT TTCTGATTT TTTTATAAT ATTGGCCAC
 40451 AGTTCCACC CAGAGTCTT GGTGACTTG AGATGCTTC TGTACCCAG
 40501 GCTGAGTGC AGTGGTGA TATGCTCTA CCGAGCTTC GACTCTGAG
 40551 GCTCAGTGA TCTCTGAC TCAGCTCTT GATGAGTGG GACACAGGT
 40601 GTGTGCCACC AGCTTGTCT ATTTTCAAT TTTTGTAGA GATGCTGCT
 40651 CTTTATGTTG TCCAGGCTG TCTCAACTC CTGGGTTCAA GCACTCTTC
 40701 CACTTATGTT CCGCAAGTG CTGGATATC AGTATGAGT CAGTGTGCT
 40751 GGCCTAATG TTTCTTTTCA TGAAGTTT TCTGATATA CTTCTGCTC
 40801 TGGGCAAGTT TCTTTTCTG GTTGGTTGG CGAAGATAT AATCTACAT
 40851 GGAATCAGT CTATATCTA CGTATTTTA GTAGATGCT TAGACAGGA
 40901 AGCTGTGAGT TCTAATGCT CAAATATCT TGAATGATC ATTTGCTTT
 40951 CACTGCTATA TTTGTGTTA ATCTCTTAG TGGCATCTC AATTTAAGA
 41001 ATTCARACT TATATTTAT GAGTTTANA GTGTAGGCC ATCTATAAT
 41051 AGCTATTTCC TGAATTTAT TACTTGTCT ATGAGTATC TAGCTATCT
 41101 ATGTAGTATA TAAAAACGT TAGTTAGAT AGCAAAAAA AATTTCTTA
 41151 TGTTCCTAAA TCATCTGTA ACTTTGTTA CTTTGAAGA AAGGAGGCG

FIGURE 3K

WO 02/48366

PCT/US01/47559

41201 ACAGAGAAAC CCACCCACTG ATATATTGTT TTATTACCGG CTGATTATTG
 41251 CAGGCGACCA TTCTAGAAC TTATATGAAA TATATTCTAT TTATTTCCCA
 41301 TGACAAGGTT TTGAAATAGG TCGTGTATC OCCATTTGCA AAGAGACAAA
 41351 AGTGAGGCTC AAGTGAAGT GACTTGCTGA GAGCCACACA AAGCCAGAT
 41401 TATGATCCAG GCTGTTTTC TAAAGTCTGC TGTATAGTAT ACTGCATTTA
 41451 TCCCAATCA AGCTTATTAA TTATGTTT ATAAAAGSCA TCATGGTTTC
 41501 AACACAGAT TACTTAGT AATAACATA TGGGTATACA TTGTCGTGT
 41551 AGTTTTCTCT TAGTTGTGGA AATAAGCMT TTAGACTAAC TTGACCTAA
 41601 ACAAGCTTTA AGCTATTAT GTAAATGGGA TCTCCAAATC ATTAGTTAGA
 41651 ACTTTTGACC CTTCCTTTT CAACTACTGA TTAAAGTGGT COTCAGTAGA
 41701 AATGACTGA AAGGAGATT TTATCTTCA GTTTCTTAC TCTCAGCTTG
 41751 GATCTAGCTC AGCGAGGGA CTTTTCACCT GCTTGTCTGA COTGACCTAG
 41801 TGATCTCAGA CATATTCAG GCAATTATG CTGAAAATCA GOCAAATTGT
 41851 AGAABAATGC CATATGCTCT TTATAGTGT AGATTGAAG AAGTCACTTT
 41901 TTAAACCTT ATCTGATPA ATCTTTT TTTTTTTCA GACCTAGCTC
 41951 TGGGTTCAC TTGAGGCTC AATGGAAT AGAGACACA TTTCCTGGG
 42001 TCTAGGATT CAATATTTAC AGTTGGCTG STATGCTTT GGTTCAGAAC
 42051 CTGGTAAGT ATAAATATT TAATGTTAAT ATTTTAAT TTGGTGTAG
 42101 CCGTTGTGT TTATTTGCT TCTCAACTGA GGGGTAGACT GTATCTGTCT
 42151 TCAATGATAG CTTTTATCT TCCAAATGT GTCTAATATA AGTCAGCCAC
 42201 TTGATATTT AATGTTCTC CTAGATGCG TTGAGGATTA AABAGGTGAC
 42251 CTTTATAGC TAAATGACAG GCTGATTTT TGAATGAGT TATACAGCTT
 42301 TTGAATCTTT AAGGAGCMT TATCTAAT CAGTCCGTTA CTAGGAAAAA
 42351 GTATGATATC CCATAGCAAC AAGGCCCTGA AAGTATATTA CATTTTGTG
 42401 TTTTCTGCT AGAABAAAG AGCTGTATA ACATAGTATC ATTACTGATA
 42451 GCACACAGA AGTGAAGA ATCCAGTTAT TTGGTGCCA GCTTCATGCT
 42501 GTGCTCAGC TTTTCTGACC AGTGGGTTT CCTGACGGG ACTTCAGTAG
 42551 TGACGCTGTG TGTAGGCTG CCCAGCTAGA TAGTATTATT GTGCTCAGGC
 42601 ATATAGGAC TGAATCTCAT GGTGTGCTCT AATATTCAA ACAGCTGACT
 42651 GCTCTGCTC TGTGTGACT TGAATTAGC CAGCCATCA GAGGGGTGAG
 42701 ATTTGCTCTC CAATCTAGC TTTCGAAGA GGTGCTCACA ATTGAATTA
 42751 CCTAAACAT TTCTTTTAAA CATCAAGCCA GGACATCCA AGTTACTTGT
 42801 TCTTACAT TTAAAGATTA GATCAATCA GCGATATCTT CACAAATCCA
 42851 TCCATAGAGA CTTCGAAA GACTGTGTG CTTCAATGCT TTGAATAAT
 42901 TGTATGATCT TTGCTCATCT ATATTATACA TATCCCTACA TATACCTCT
 42951 CATTTGCTG TAGAATAAT ACCAGACACA TTTTACAGTT TGAACATGA
 43001 TATCCCTATT TTACATTACA CCGTCACAGC ACTCTGTGTA ATTAAGTTGC
 43051 TGTCTTTGTA TACAGGCTCA GATTTGTAG CAGCTTGCCC ATAGTCACT
 43101 AGTACGAG TTCACTCTC CTCTCTCTA CAGCCATTC ACATAGAGA
 43151 TACTTAAAT ATGATGTTG ACTTTCAGT ACAGTGAGGA AATATTACT
 43201 GTTTTTGTT TGTGCTCTC CTATAGGAT GATGTGGCT GCTGGGCGAG
 43251 TAGCAGCAT GTTAGCATC ACCTTCTCT CTGTAGTGC ACTGTGTCA
 43301 CAAACCTCT AGCTGATCA ACAGGCTGAG TTGATAGGA CTAGGATTA
 43351 TATTTTAAA GTACAGAAAG TTCTAATCT GTTCTCTCT TCTTGTGAG
 43401 TGAACATATA GTATATCTC AGGGTAGAGA CTTTAAAAAT TGTCTTGTAT
 43451 ATAAACAGA AAGCGAGAT TAGGGTATT TATCCTTAGG TAGATACATA
 43501 TTCTTTTCT TCTCATTAG AATAAGTGG TTATCTGTTC TGTCTTAGA
 43551 AATTTTCTG ATGAGGCTGG CAGTGTGCT CACACCTCT AATCCACACA
 43601 CTTTGAAGG CCGAGGCGAG CAGATTCTT GAGTTCAGA GTTGAGACC
 43651 AGCTGGGCA ACATGGGGA ACCCATCAC TAAATGCAA AATTAGCTG
 43701 GCGAGGCTGG CCGATGCTTG TAGTCCAGT TACTTGGAG GCTGAGGTTG
 43751 GAGGCTCAT TGAATCTAG AGCGCGAGT TGAATGAGC CAATATCAGG
 43801 CCACTGACT CCACTGAGG CCACTGAGG ACAGGCTCTG AGAGGCTCT
 43851 CCAAAAAA TTACTGATG AATGTTGGC ACAGATAT TACTAGCAT
 43901 TAAATTTTG CTGTGTGTG CTAGACACT TGTGGATAT ATTCAATAGC
 43951 CTTTATGAA TGTGTGCTA GCACTTGGG AAGCTGAGG AGGAGGATTA
 44001 CTGAGCTCA AGATTTGAA GCTAGCTAG CACACACAG AGAGTCCA
 44051 TCTCTAATA AAAAAAAA AGAATACATA TAGTTTCAGT CATCTAGAGA
 44101 TATGTAAAC AGCAATCTCT TTAATGTAT ATGCTCTCT TTTTTTTTT
 44151 GTTGGGACA TTATTTCTT GGGATGCTA CAGATGACAA ATACTTCTTT
 44201 GAAGAGATA GTTTTCTT AGCTATAGG ATAAAAGGG CACTGAAAA
 44251 CAGTGGAGT TGTCTCTTG TGTATATG AGAAGGAAA GGGCTATGG
 44301 CAAGCAGAT TGCATTTTA ATTAAGATC CAAGAGAGAG AACAATAAT
 44351 CAAATGCTGT TAAACAGTG TTAATGGCG GCGCAGTGG CTATGCTGT
 44401 TATCTCAGT GCTTGGGAG GCGAGGCTG GTGATCACG TANGGTGAG
 44451 AATTCAGAT CAGCTGGCG AATATGCTG AACCTGTCT CTATATATA
 44501 AATCAACAG CCGTCTGTGT GCACACTGT AATCCAGCT ACTCGGAGG
 44551 CTGAGGCGAG AGATCACTT GAACCTGGA AGCGAGGTT GCACTGAGCC
 44601 GAGATCTGTC CACTGCACT CAGCTGGGG AACAGAGTG AATCTGCTC
 44651 TCAAAAAA AAGTGTGAT TTGCTGGG CAGTGGCTC AATGCTGTA
 44701 ATCAAGCAC TTGCTGAGC TGAAGGAGA GATCACTTG AGTTGATTC
 44751 AACCCACCC TAGGCAACAA AGTGAAGGCC CATCTCTACA AAAAGTACA
 44801 AATTTAGCA GGTATTTGAG CCCAGAGAT TGAGGCTGCA GTGAGCCAG
 44851 TTGCGCCAC TGCACCTCCG CCAAGTGAC CCGCACTTT ATTTTGTACT
 44901 AAGGGGGTC TTAATATCTT GCTTGGGAG AGATTAATTT TAAATTTGT

FIGURE 3L

WO 02/48366

PCT/US01/47559

44951 AGTTGTGAAA TTGTGACTGA TTTTAAACCA GCCATGTTTT GAAGGCTGTA
 45001 ATCTGGGARA ATAAATGATG TTTCGCCACT TGTTCTTAGG TGGAGCCTAT
 45051 ATATGGGTAC ATCAGAGCTC ATGTTTGCAAT AGACACACCTT ACTACACCTT
 45101 GTAGAGCTGA TAGCTTCTAA ATATTAAATG TTTTAAATGA CATTGCTATA
 45151 AATTGCTAGT GTCTGATAAG AGCCTTGATT TTAGATTAGG TGATTTCATA
 45201 TTTTACTGAG TTCTAAGGTA GGGAAAAAAR ACCAGSTAGT TTACTAAGTG
 45251 ATATATTTT TAAACATGAT AGGTGTGCTT CAGGAGATGA TACAGGGAA
 45301 TCGAGGATTA TGCATGCTC TGGGACCGGG CCTCTAAGGA TTCATTTCCT
 45351 ACATATTTCA TGTGGAACTT AAAGAACTGC CAATAACAGG AACAGACTTG
 45401 GGACAAACCA CAGGCCCTCA GCACCACTTT GACACAGTAA TTCTCATTTA
 45451 ACATGATCRA ATGTATAGST TCATTGSGCT AGATTAAAGG CTACAGGTTT
 45501 TGTGCTATA AAGCTTTTAT GTCGATATCT ATTTGAGATT GATTAAGAG
 45551 CTAAAAACCA AGAGTTTAAG GSACTGTGPT CTCTACATA CCACCTTGTG
 45601 AAAATTGGCT GTGCATATTT TTTTCCACT TGAGAAATGA AAMTTTAAAG
 45651 TACCTAGTGT GRCATGSCA TATGPAACBA ACCATTTCCT TTACTACTAT
 45701 GTCCTTTTTC AAGCTTTCTT AATAACACAG TGTGCTTAGT TACTAATCTT
 45751 TTCAATACAA AGGCAGGCTT ATGCTAATAA ACCTTTGTCT ACTGTTGCTT
 45801 ATATTTGCTT CTTTCTACT GAAATTAAT GCCAACAAT GCTGATCAT
 45851 TGGGCCATA TTGCGAGCTC ATCAGTTTCC TGGCTTATTT AGCAGAGTTC
 45901 TGACTCTATC TTCCAACCA CTAAGACTAC TTTTAAATGA AAGATGTTT
 45951 GTGGGCTATA CACAACCTG CCAGTGCAGG CTTCCTCTTA GAATGATGTT
 46001 AATTGTGGCG TAAACAAAG GTCCTTTTTC CTGSGTAACC TGTTTTCCT
 46051 TCACTTCCAG TCACCTGAAA TTATAMTTAC CTTCCTGAAA CAAAGAAATG
 46101 GTATATTTAT TTGCTACTGT AGTAATATCT GTATGCTAAC TGATTITAT
 46151 TTGTTTCTTG TTGTTCTGGA GACAGCTCTA CTCTGCTACA CAGGCTGGA
 46201 TCGAGTGGCA CAATCTTGGC TCTCTGCAAG CTCTATCTCC CAGCTTCAAG
 46251 CGATTCCTTT AACATCAGCC TCCCGGTAG CTGGGATAC AGGCATATGC
 46301 CACCATGAGC GGCATATTTT TTTTGTATTT TTAGTAGAGA CAGAGTTTCA
 46351 CATGTTGGC CAGGCTGCTC TCGAATCTCT GACCTCAGAT GATCTGCTG
 46401 CTCACGCTCT CCAAGTGTCT GGGATATCAG CTCTGAGCCA ACAGCTGCA
 46451 CCGATTTTAA TTTTTSAGAT CTTCTATATA ACGTTTTCCT CTGGACTA
 46501 CAATATATCT ATGAATATGC TGTGTTCA CA TTTGTGCTG AAGTAATGA
 46551 TGTATATATT TCGCATAGGC TGTTTTCCTA GTTGTATTC CAGAAATTT
 46601 ATAGGAGATG TGAATATATC TCTTTCTCC TAAAAAGGA ATGCTTTTA
 46651 TAAATAGGCT TTAAATAGG AGAAGAGCAH TACTCTGTA ATCTATAGT
 46701 TAACTACTAA AGGTATGTAG TTCTCTCTTA TGAGTAAAG TATTGTGTA
 46751 AATATCTTAA TTACTTTATT CTCTCTCAGA ATCCATCAT CCTGSGCCT
 46801 CCTTCTTAT TTGAGGCTTG TCGATAGCT CTGSGCTCTG TTGTTGCTT
 46851 GTTCTCTG CAGCATACCA ATTTAGCTT AGGTCCTGCG AGTTGAGAA
 46901 AGCATCTGCG CAGTACAGCG CATCTCTATA ATACACAGG CCAGGAGAG
 46951 GCCAAGAAC CTTTACTCCA GACACCAAT GTGTGACAG TGAATCAGG
 47001 AAGATTTTTC TATCAGCACG CAGGTCITAG TTTTCACTC TAGTCTGGA
 47051 TGATCATTC ATTCCATCC ACGGTGTACT TTAAGATTGT CTTAGAAAT
 47101 CTATCTGCT GACCTCGGCG GCACTAAGG GAACTGGGAA CTAGAACCT
 47151 GACAGCTTTC CAAAGATGTT ACAATTTCT TTGAATAACG TTTGTTTAT
 47201 TAGCACCAAT TTCTTGCCAC TAAGCTATTT GTTTTATAT ACATCTTTA
 47251 ATTAARAATCT ATATATGTA CTCTTAGAT ATTAGCAAT GTCTCTGTA
 47301 CATTCTCTT AAGGCTTGA CTTTACTCT TATCTACTCT CAGTGAGCA
 47351 CAGTAGGAGG TATGCTGCG GACCTATTG TTTTACATTT GTAAATTTT
 47401 GAGTCAGATT TTAATATGTT AAAATCTTGG GTCAATATAT TCAAGCCTT
 47451 ATGCGAGATG CACTAAACCA AAGAAATGTT AATGAAATG TTTCATTTA
 47501 AAAAAAARA CACTTAAGAA AACTGTACTA AACTGATTC ATGTTTGGG
 47551 CTCTTTCGA GACCTTTTAA ACATATCTA CTACTCTTT TGAAGGAA
 47601 AAGTATCAGC CATTTAGCAT TTAAGTGGG GTATTTAGAG CCGTAATCT
 47651 AATGCTGGC TCAATTTAT TCCCGAGCTA CTCTTATAC CACTATTCCT
 47701 TTAATGTTTG CATATCAATA AGCAGCTCAG CACTTGATA CATATCTTA
 47751 AATATATTA GTAAAGCTG TACCTCTGA AATSTCAGT TGAATCTAT
 47801 TATGTAGATA AATATATTA GTGCGCTTC AGGACTGTCA CAGTATCACT
 47851 TATTTACAG AGCTAATGTT TGTCTAAT TTTCAAGACC CTAGAGGAGA
 47901 GCTTTATACA ATTACCGAGG TGAATTTCTC TAAAGTGTAT ATTTTGTGT
 47951 CAGGTATAT TATTAAAAA AGTGTTACTT TGAATAAT GTATATAAG
 48001 AAGTATAGG TTTCACTCTT TTCACTCTG TGTGCGGTA TTCTTTATG
 48051 CTTTAAAC TTGACACCT CACTATGTT AATTAAGTC TTAAGAGAA
 48101 TGTAAATATT CAGTAAATA AGTTAATAT TTAATGATT TTTTTTTAA
 48151 AAGATCTTGT ATCTGTGAGT TCTACAGGG TCGTTTGG CTGTTGTACC
 48201 ATATTTTAC ATTAAGAAAT CAGAGCTTA TCAATCTTT CTCTCATCC
 48251 ATTTCTTTC CACTACATA AATGCTCTT CAGGAAGAC AAGGGAAGG
 48301 AACCAAAAT GCGATAGCG TTTTGAATA CTGATTTTC AAGTGTAGA
 48351 TCGTGCATA TAGTTTATTA ACAGAGCTCA AAGTTTACAG AATACATT
 48401 AATGCAACA CTTTGTGTT ATTAACATCT GTACATTTT ACATGCAAA
 48451 AATTTGCA CTAAGATTT AAGGCTTAA GATTAAGAC ATTTGAGAT
 48501 CTGTATACAT TTAATACAAA TTTGCAAGAT ATCAAGATGA TCTCTCTGA
 48551 AATTTCACT TCTTGTGT ATATATATGC AAGAGATAC CTATATTTA
 48601 AAAAAAGAGG CTACCTGAT GTCAATGATC CCATCCAAA CATGTACAA
 48651 TATTTTATAT ATNTACATA GTATTACAA CAAAGCCCTT TCCATATAT

FIGURE 3M

WO 02/48366

PCT/US01/47559

```

48701 TTACATATAA GTCCTTCCA AATCACCRA GAGGCITCTT TCCAGATAG
48751 CAGTCTCTTT CAGTTACTG TACCAAGTAT TCTCAATAT ACAGTAGTGT
48801 ATCTAATGG GGGAGGTGGG GAGTATCAA CTGCCATAA TTAGTTACTG
48851 AATTTTCAGA TTAGATCCCT AGTTAAAGA CACTTCAAT TGCCATAGT
48901 TTCAGATAC TGGCTTCGTA AAAAAAANA AAAAAATAGC AAGATCTCT
48951 TCTTACCATG GAGTCAGTA ATCTATTAC ATCAATATCT TGAAATCTTT
49001 CTTCATCTCT GTATTAAGT AGTTGAANA ATTTTAAAT TTCTCAGGA
49051 GCATTAGTAC TTATATAANA TTGATCTAGA AGACTCTTA GAGNATGCTA
49101 CCAGGTATTG TTGTCAATA AGAAACAAA ATGAGGCCCT ATGATCTAG
49151 GTGATTTTAA AAAAAATTTT TCCCAAGGA ACAAATCTTA AACCTTCAG
49201 TTACAGGGA GAGAGAAA TTGACCTTC TACAGATAT TTGTTTGGC
49251 TACTTATTA ACATGCAAT CAGCCCACT TCCTGTACTA TAAATATAA
49301 TGAATATCA TGAATAGTA GATTAGTAG TAGTACCAG TTITAAAGT
49351 CCACGTATAC AGATTTAAT GCATATATA ACTACAAAT GAGAAAGCT
49401 AGATGAGTAT AATCATTTAG GAGCAGCTGA AGTTTAAAA ACTTGTGAGC
49451 TCAGATGCA AAGCACTTA ATCTGCTCA ATAAAGACT TCCAAATAT
49501 TTACAGTATA TCAATTTAT TTGATTCAT CACTGCAAT TTAAATGCTT
49551 CAGGGAAT ACCTATGAA GGCCTAATC ATTTCAGTA TTTAAGGTA
49601 ATTTAGACT TTTCCTTTA AATGTCATCA ATATATATC TCAGAGGTA
49651 ATCATTTTC CTCATGATT TCAGGTGTTG AATTAAAG TTTTACAAAT
48701 GAAAGTGCA AATGATTA GTATCTTTC GTTCAACT ATGAGGTAT
49751 GAGCGGTTT TAGTTTACC AGATTTTAA TATGATAGA ATCTCCCTCA
49801 TAGAGTACCA CTCTTCCCT AGAATGAA TTTCAAGAA TAGCTAATA
49851 AATTTATAT TATTTATATA AAGCCATGT TCACAAACCA AAAATGCTA
49901 TTTACTGCA ATCTTACCA ATCCCAAGT ACAAATTTGT TCTATATTA
49951 TAAACTGCA TAAAGCAGT ACTTAAAGTA TCCA

```

FEATURES:

Transcript 1:

```

Start: 3000
Exon: 3000-3083
Intron: 3084-14681
Exon: 14682-14777
Intron: 14778-23440
Exon: 23441-23503
Intron: 23504-24652
Exon: 24653-24780
Intron: 24781-26609
Exon: 26610-26724
Intron: 26725-32754
Exon: 32755-32973
Intron: 32974-33250
Exon: 33251-33364
Intron: 33365-34391
Exon: 34392-34463
Intron: 34464-41941
Exon: 41942-42094
Intron: 42095-43227
Exon: 43228-43325
Intron: 43326-45272
Exon: 45273-45436
Intron: 45437-46779
Exon: 46780-46983
Stop: 46984

```

Transcript 2:

```

Start: 3000
Exon: 3000 3083
Intron 3084 23440
Exon 23441 23503
Intron 23504 24652
Exon 24653 24769
Intron 24774 46984
Exon 24773 24780
Intron 24781 26609
Exon 26610 26724
Intron 26725 32754
Exon 32755 32973
Intron 32974 33250
Exon 33251 33364
Intron 33365 34391
Exon 34392 34463

```

FIGURE 3N

WO 02/48366

Intron	34464	41941
Exon	41942	42054
Intron	42055	43227
Exon	43228	43325
Intron	43326	45272
Exon	45273	45436
Intron	45437	46779
Exon	46780	46983
Stop:		46984

Chromosome Map Position:
Chromosome 1

PCT/US01/47559

FIGURE 30

WO 02/48366

PCT/US01/47559

Allelic Variants (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
126	G	T C	Beyond ORF(5')			
113	G	A C	Beyond ORF(5')			
1901	C	T	Beyond ORF(5')			
2775	T	C	Beyond ORF(5')			
4917	A	T	Intron			
5406	T	G	Intron			
5626	C	T G	Intron			
7201	G	A	Intron			
10581	A	-	Intron			
10608	T	-	Intron			
12495	C	T G	Intron			
12571	C	T	Intron			
14024	G	C	Intron			
14643	T	A	Intron			
15137	A	G C	Intron			
16174	G	A	Intron			
17112	C	T A	Intron			
17204	-	T	Intron			
18636	G	C T A	Intron			
18830	-	T	Intron			
19530	G	A	Intron			
21579	G	A C	Intron			
22755	G	A	Intron			
27706	-	A	Intron			
28039	A	-	Intron			
28045	C	-	Intron			
28900	A	G	Intron			
28812	C	G	Intron			
28800	T	A	Intron			
29387	G	T	Intron			
31491	A	G	Intron			
31589	G	A C	Intron			
32608	A	G	Intron			
36093	-	G A	Intron			
37842	A	T	Intron			
41941	G	A	Intron			
41983	G	A	Exon	311	K	K
42047	G	A	Exon	333	E	K
43183	A	G	Intron			
43785	C	G	Intron			
45812	A	G T	Intron			
47215	G	A T	Beyond ORF(3')			

FIGURE 3P

PCT/US01/47559

[illegible]

FIGURE 3Q

PCT/US01/47559

FIGURE 3R

PCT/US01/47559

FIGURE 3S

WO 02/48366

PCT/US01/47559

CGTGATCTCGGCTCACTGCACCTCCACGCGCAGGGTCAAGAGATTCTCTGCTCAGC
 18830 AAAAGCTGTGTTTGTGGNAGATCTGCGCCAGGGGAGATGGTGAGCCTTACCAACAGG
 GGATTAGCTAAAGAGAGATTGTTGCTGACAAATATTTTAAAGTTGAGGCTCAAC
 ACTTTTGAGGGGTATGAATATATGTTGGTCAATAGTAGATAGTTCACTCAGATCAGTG
 ATTTTGTCTGATTATGAACATATTAGCTAAGTGATGAGAAATACAGTAGGATAAGGA
 TCTGTAAATGCCAAGAGGTGGAATTAACGGGTTTTTTTCTCTCTCTTTTCTTTT
 [-, T]
 CATTGAGACGGAGTCTTAATCTGGCATCCAGGTTGGAGTGCAATGGGCTGATCTCGGCTC
 ACTGCAACCTCCACGCCAAGGTTCAAGAGATTCTCTGCTCAGGCTCCCGAGTAGCTG
 GATTAACAGGTCAATGTCACCGCCAGCTAAATTTTTTTTATTTATTTTGGGA
 CAGGTTCACTCTCTGCTGAGCTGGAATTCAGTGCACATATCTGGCTCAGTGCAC
 CTTCGCTCCCGAGGTTCAAGCAGTTCTCTGCTCAGCTCCAGATATGAGGATTACAG
 19530 CCCACCTAAGCCTCCAAAGTGTGGGATTACAGGCGTAGGCCACTGTCTCTGGCGCGCT
 TTAACCTTTTGACAGACCTATGGCTCTGGAAATATAGGCCAGTCTTTGATGGTCAAGC
 TCCTAGATACACAGTCCATGTTAGGAAACACTCAAAATCCACTAGCCTCTCTTACCTA
 GATGGTTCTGTTCTCTTGGCTACAGAAACAGCCCAAGCGTTAACTTTTAAAGGATTA
 TTTACTTTCAACATTTTAAAGTTAAAGAAAGTTAAGATCCATTAATTTTGGAAAA
 [G, A]
 TGTTCACATTTTCTCTGTTCACTCTAAGACCCAGTGTCTAAGGATCTGACATCAAAAT
 CTTCACACATCTCAATTTCTGTTATATTTGCTGTAAATTTGTTAGAAATTTGAT
 GGCCCAAGAGGAAATTCCTTTGGAGAAAAGTTAGCTAGCAGGAAACAGTTGGAA
 GGGTTGGGGGTTGGCCAGATAAAGAGGGAAGAAACATTCAAAATTTGAAGAGTGCCT
 GTATAAATATGATATTGGAAAGCATGAATATTTAGAAACAGTGAAGGAAAGATTT
 21578 AGGAGGCTGAGGACAGAGAAATTTCTTGAACCTGGGAGCGAGGTTCACTGAGGCCAAGA
 TTGACCACTGCACTCCAGCTCGGCGCAGAAAGGCCAATCTGCTCAAAACAAACAAA
 AAGGAAATAGTCTGAGTAAATGTAGAGTACAGTATCTCTCAATATCTGATTTCC
 TTTTTTGGAGGGATATATCTAGTGGAGGATCTGAGTCAATGAGTACTCATTT
 TTAGGTTTGTGAGAGGCTCCACACTGTTTCTCTGATGTACTAATTTACATCTC
 [G, A, C]
 ACCAACAGTGTATGAGTGTCTCTTTCTCCACATCTCTGCTATCTTTTGGATAAAGC
 TGTTTTAACTGGGAGTGAATTTCTGCTGATGTTTGAATTCATTTCCCGATGA
 TCAATGATGTTGAGCTTTTTCATATCTCTATTTGGTCCCTTTGAGAAATCTCTCA
 GATCTTTGCCCCGTTTTTAAAGATCAGATTATGAGATCTTTCTTACAGATTTGTTG
 AGCCCCATATACATTTTGTATTATCTCTGTCAGATGATAGTTTGCAGATTTTCT
 22755 GCACATTTTGTAAATTTCTGCTATTTGGCAGGAGAAATATCTGAGAAAGGGAGC
 AATCTGATCTCTTAAAGGTTCTTCTGCAACATGTCAGAAATATATTAGCATATAG
 TTTCTCTTAAAGGAGAGCTTCCCTACCTTCTTATTAACCAATTCATCTCTG
 TGTATTATGAGCGATAGCATTTGGAATATAGAGCATAGTTCTTATGCTCAACAGGA
 CTCACTTGCCTCATATGTAAGTGTATATTTATCTAATTCAGATGTTGGGATTAAGA
 [G, A]
 GAGTACATATAGGCTGTAAAAATGCTAGCTGCTGTTATTTTCCAGTTGCTGGAATTG
 CCTTTTCATTTGATGATTTCCAGCGGTTCTCTGCTGCCACTGCAAAAAATGATACCA
 CATGTTTGAGAACAGGCTTGGAAAGGATAGATATCTGTTATATCATTTTCACTAGGTT
 GGGATTTTCTCTTATAGAAATTTCTAGATCTACTCTGCGAAATTAAGATTTACT
 ATTAATTTTCCCAATCTCTTATCTCTAGATATATATCCATCTGAAGAGCAATTAAAGTC
 27706 AGACAAATACCTTCTGTAATTAAGGACTGATATTTGTTGGGCTGAATTTAGTTTAA
 AGGACCTGATTTCTGATTCAGAGGAGTTATAGTAGAAATCATAGATTTTGGGAGG
 AATCCCTTGAAGGAGGTTAGAAAGAGCTAATATTTCTGGCTGCTGATGAGGCT
 CACACCTGTAATCTCAGCAGCTTTGGGAGTTGAGGCGAGCGGATCACTTGAGATCAGG
 TTCAGGACCGGCTGGCCACATGTTGAACCTTGTCTCTATTTAAAGAAAGAAAAA
 [-, A]
 AATAGTCAGAAAGAAAGAACTAATGATTTCAAGTTGTAACTTGGACATTAATGATAC
 AAGGCTGATGATAGCGAGGATTTAAAGATGATCAATTAAGCTATAGTTTACATACC
 ATAAAGTTATCTCTTTTATGAGTATAGTTCAAGTAAATTTAGTAAATTTATACGTTAT
 GCRAACACCCATACCCATTTGGGTTGGTGGTGGTTGGTTGGTTGGTTGGTTGGTTG
 GTTTTTTGGAGTAAATTTATTTCCCAAGGCAAGGTTTGAATATACATTTTCAATC
 28039 GTTGTAACCTTGGACATTAATGATACAGGCTGATGATAGCCAGGATTTAAAAAT
 AGCTAATTAAGCTATAGTTTACATACCTAATTTATCTCTTTTATGATATAGTTCA
 GTGATTTTATGTAATTTATATCTGTTATGCAACACCAACATATACCCATTTGGGTTGG
 TCGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG
 AAGGTTTGAATTAACATTTTCAATCTGGGTTCTCTGTTATTAAGGAGTTCTGGC
 [A, -]
 AATACAAAGCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA
 ATATATGATGATGAGAAAGGTTGGGAGGTTGGGAGGAGGAGGAGGAGGAGGAGGAGG
 TACCTCCAGTTTGGGAGGCTCAGGCGGGGAGGTTTGGGCGAGGAGTTTCAAGGCTC
 AGCTATGATCAGGCGCTACACTCCAGTCTAGGCTACAGAGTGAAGCCTGTCTCTAA
 AAAAAAGTTGAGGCGGCGAGGCTGCTGCTATATATCCAGCACTTGGGAGGCTGAG
 28045 AACTTGAACATTAATGATACAGGCTGATGATGCGAGGATATTAAAAATAGCTA

FIGURE 3T

PCT/US01/47559

FIGURE 3U

WO 02/48366	PCT/US01/47559
	GTCCACAACTCTGGCTCAGCAGTCTCCACCTTGGCTCCCAAGTCTGAGATTA
32608	CCAGGCTTGGTCTCTATACCTGTACTCCAGGCTTCGGGAGGCTGAGGTGGAGGATC GCTTGAGCTCAGAGGTTGAGACTAGCTTAGGCAACATAGCTGACTTCCACCTCTATAA AATAAACAATAATAGCTGGGCTGGTGGTGTGTGCTGTAGCCCATCAGAGGATCTT CAGGCAGAGAGATCTACTTAGGCTGAGAGGTCAAGACTACAGTGAAGCTGTATGGCACC ACGCGCTCCAGCTGGGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG [A, G] AAAAAGATAAATGGCAGGTAGGAGAGAGAGAGAGAGAGAGAGGTGTGCACAGAGGCTGA CATAAATCAGATTATGAAGGAGTATGTGGTGTGTAGGAGAGAGAGAGAGAGAGTACTAAT CTCTGTCTACTTAAATTTATGACAGGTTGAGCAGCAATTTGCTGAGGTTTATGTAACCA GTCTGTGATTTGAGGTTATCTTGGAGGTATATGGGAGCAGCTTGGTGGTGTCTATG CTACAGCAATAGCTTTGCTAGATTTTGTGTTTACTTGTGTGCTGCGAGAGTGTGTC
36093	TACTATTCTAGCTCAGGTTTGTCTAAAAATCATCAGCTGGAAAAACAATGCATTTA AATATTCTCTAGGCTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG CTTTATTAATAAATGGCTTTAATAAATTTTGGGAGGAGGTGGGCTGGTCAATCTT GTATTCCAGCAGCTTTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG CAGCTCAGCACTTGGCAAAACCCATCTCTACCAAAAAAAGAGAGAGAGAGAGAGAGAG [A, G, A] GAG GGAGGAGATTGCTTCACTCTGGAGGAGTGGAGGAGTGCAGTGAAGCAGAGTTGCTCCATGGA CTCCAGATTGGGCAACAGAGATGAGAGCCTGTCTCAAAAAAAGAGTGAAGAGAGATACAC AGTTAGTATGTGAGAGATTGATGAATATCAAGATTAAACCACTTGAATAGATGT GATCAACCTGGGCAAGATATCTTTTATCATTTGCAATGCAACATTAGATTGTCAACATTT
37642	TCAGCCCACTTCTCTGGGAGGAGTACGATATAAGAGTTCTTACACTTGCACATAGT AAGTGTTCATGAGAGATCANTATAGTTCTACACATTTCTTTTCTCAGTGTTTCTC TATCCAGGCTCTCTGGGAGGAGTACCTCCAGACTTACTCTTCCAGTCCAGTCTCTGAGC TCTCTATGAGCAGATTGGCTGGTATCTCAGCTTACAGTGAAGAGAGAGAGAGAGAGAG TTCCCTCTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG [A, T] GGTTTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG CAGGCCATCAAAATGATCTCTATCAGCTCTCTCCCAAAATATCTTGAATCAGAGC GTTCTCATATTTCTCATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG GTCTCTCTGCTTCCATCTTGGCTCAGTATAGTTCATCTGCAGTTCAGAGAGATTTT GTAAATGGAATTTATCTCATCTTACTATATCACTTCTCTTCTGCAACAGAGATAA
41941	ATTAGTTAGACTTTTGGGCTTCTCATTTTCAACTACTGATTAAAGTGGTCTCAGTGA AATGATCAAGATAGAGAGTTTATCTTCTCAGTTTCTTAACCTCAGCTCTGATCTATGTC AGCAGAGGAGCTTTTCTCTGCTTATGTGAGCTTGGAGTGTCTCAGAGATATTCAGG GCAATTTATGCTGAAGATCAGCAGAGATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG AGATTGAAGAGAGTCACTTTTAAAGCTTTATCTGATTAATCTTTTTTTTTTTTTTCA [G, A] ACCATAGCTCTTGAATTTACTTATGAGGTCAATGGAGATAGAGAGAGAGAGAGAGAG CTAGGATTCAAAATTTACAGTTGGGAGGATGAGTCTTGGTTCAGAGCTTGGTAAATA TAATATTTTAACTTAATATTTTAAATTTTGGGTAGCCCTTGTGTTTATTTCTCTCT CTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG TCTAATATAGCTCTGAGCTTGTATATTTATGCTCTCTGATAGGCTTGGAGATTA
41983	TAGTGGTCTCTCAGTGAAGTGTACTGAATAGAGAGTTTATCTTTCAGTTTCTAATC TCACTGAGATCTATGTCAGAGAGAGGAGCTTTTCACTGCTTATGTGAGCTTGGAGTATG ATCTCAGAGATTTGAGGAGATTTGCTGAAGATCAGAGAGATTTGAGAGAGAGAGAGAG ATAGTCTTTTATATGCTGAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG CTTTTTTTTTTTTTCAGAGCAGTCTTCTGAGTTTACTTATGAGGTCAATGGAGAGATA [G, A] ACAGCAGTTTACTGGGCTTAGAGTTTCAAAATTTACAGTTGGAGGAGATAGGCTTGGT TCAGAGCTTGGTGAATTAATATTTTAAATGTAATATTTTAAATTTGGGTAGAGCTT TTGTTTATTTATTTGCTTCTCACTGAGGAGTACACTGATATCTGCTCACTACTATGCT TTTATCTTTCAAAATGCTCTATATAGTCTGAGCTTGTATATTTATATGTTCTCTTA GAATGGCTTGAAGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG [G, A] AGCTTGGTGAATTAATATTTTAAATGTTAAATTTTAAATTTTGGGTAGAGGAGGAGGAG GTTTATTTTCTCTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG TCTTCAAAATGCTCTCAATAGGCTTGGAGGATTAATTTATATGTTCTCTCAAAAT GGCTTGAAGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG GATTTAGAGGCTTTGATCTTTTAAAGGAGGATTTATCTAATATCACTCTCTTACTAGGA
42047	TCGAGATCTATGTCAGCAGAGGAGCTTTTCTCTGCTTATGTGAGCTTGGAGTATGATC CAGACATATTCAGGAGGAGATTTATGCTGAAGATCAGCAGAGATTTGAAGAGGAGGAG TCTTTTATATGATGATGATTAAGAGAGTCACTTTTAAAGCTTTATCTGATTAATCTTT TTTTTTTTTTTTCAGAGCAGATGCTTGAATTTACTTATAGGAGTCAATTTGAAGAGAG ACCATTTTACTGGGCTTAGAGTTTCAAAATTTACAGTTGGAGGATGATGCTTTGGTCA [G, A] AGCTTGGTGAATTAATATTTTAAATGTTAAATTTTAAATTTTGGGTAGAGGAGGAGGAG GTTTATTTTCTCTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG TCTTCAAAATGCTCTCAATAGGCTTGGAGGATTAATTTATATGTTCTCTCAAAAT GGCTTGAAGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG GATTTAGAGGCTTTGATCTTTTAAAGGAGGATTTATCTAATATCACTCTCTTACTAGGA
43183	TCATGTGTTTGAAGATTTGATGAGTGTTCATCTATATATCACTATCTCTCACTA

FIGURE 3V

WO 02/48366

PCT/US01/47559

TAACTCTCAATTCGCTGTAGAAATAATACCCAGCACATTTTACAGPTTGGACATGATA
 TCCCTATTTTACATACACCTCCACAGCCTCTTGTGAATTAAGTTCCTCTTGTATA
 CAGGGTCAGATTGTAAAGCAGCTTGGCCATAGTCACCTAGTAGACAGTTAGTTCTCTCT
 TTTCTCTACAGCCATTTACAGTAAGAAATTAATTAATGATGATTTGACCTTCAGGTAC
 [A, G]
 GTGGAGAGAAATTAATTCGTCTTTCTTTCTGCTCTCTCTATAGGATGATGGGGTGGT
 GGGGAGTAGACAGCCATGCTAGCATCACCTTTCTGCTGTCAAGTGCATTTGTTACAGA
 ACTGCTGATGCTGATCAACAGGGTGAAGTTGATAGGAAGTACGATTAATTTTAAAGTA
 CAGATGTTCTAATCCCTGCTGCTCTCTCTATGTAAGTGAACATTAAGTATATCTTCAGG
 GTAGAGACTTTTAAATTCCTTTTATATAAACAGGAAGAGAGATTTAGGGTATTTAT
 43785 CTTAGGTAGATACATATTCCTTTCTCTCACTTAGAATATGAGCTTATCTGTCGTGTT
 CATAGAAATTTACTGATGAGGCTGGGATGGTGGCTCAGCCTGTAATCCCAACATTT
 GAGAGGCGAGGCGAGGAGATTGCTTGGGTTGAGGAGTTGGAGACAGGCTGGGCGACAT
 GGGGAACCCCAATCACTAATGAGCAAAATTAAGTGGGAGGCTGGGCGAGCTTGTAGT
 CCGAGTACTTTGGGAGGCTGAGCTGGGAGATCCTTGGTCTCAGAGGCGAGGTTGCA
 [C, G]
 TGAGCCAAATCAGGCCAGTGAATCCAGCTGGGCGACACAGTGAAGACCTTGAGGA
 CTTGCTCAAAANAAATTTACTGATGAATGTTGGCCACAGAAATTAAGTAAAGCATTAAGT
 TTTCTCTGCTTGTGGTAGAGCTTGTGGGATTAATTAAGAGCTTTTAAAGTTG
 TCCAGGCCACTTTGGGAGCTGGGCGAGGAGTACTTGAAGTCAAGGTTTGAAGTAG
 CTTAGGCAACACAGCAAGCTCCCATCTCTAATAAAAAANAAAGAAATCATATAGTT
 45812 AGCTTTTATGCTGAGTTCTTATTTGAGATTGGATAAAAATGCTTAAACACAGGTTTAAAG
 GACTGTGTTCTTCTACATACCCACTTGTGAATTTGGCTGTCATATTTTCTTCACTT
 GAGAAAGAAATTTTAAAGTACCTAGTTTGTCAATGGCTATGTAACAAACCTTTCTCT
 TTACTACTAGTCTCTTTTGAAGCTTTTCTAATATACAGTTTGTATGTTAATTAAGTTT
 TCAATCAAAAGGAGGCTTATGTTAATAACCTTTGTTCACTGTTGCTATATTTCTGCTC
 [A, G, T]
 TTTCACTGAAATTAATGCGCAACAGTGTGATCAATTTGGCCAAATTAAGCAGCTCAT
 CAGTTCTCTGGCTTATTTAGCAGAGTTCTGTACTTATCTTCCACCCACTAAGACTACTT
 TTAATTAAGAAATGTTTGTGGGCTATACACAAACTGGCAGTGAAGCTTCTGCTAAGA
 ATGATGTAATTTTGGGCTAAGCAAGGCTCTTTTCTTGGTAACTGTGTTTCTCTC
 ACTTCCAGTCACTTGAATATATTAATTAATCTTCTTGAACAGAGATGTAATTTATTT
 47215 CACAGCCATCCTCATATACACAGGCGCAGAGAGGCGCAAGAACCTTTACTCCAGGAC
 ACAATGCTGAGCAGTGAATCAGGAAGATTTTCTATCAGCAGCTAGGCTTAGTTTT
 CACTCTAGTTCTGGATGACATCCATTTCACTCAGAGTGTACTTAAAGATGCTTTA
 AGAAATGATCTGCACTGACTCCCTGGGAACTAAGGAAGTGGGACTTAGACCCAGACA
 GTTTTCCAAAGATGTTCAATTTCTTTGAANACCTTTTGTATTAGCCACCAATTTCT
 [G, A, T]
 GCGCTAAGCTATTTGTTTATTAACACCTTTAATTAAGAACTATATATTAATCTCT
 TAGATATAGCAATCTCTCTGCTACATTTCTTAAAGTGTGAGCTTTAATCTATAGC
 TGACTCAGTGAGACAGTAGGTAGTATGTTTGGGACCTATTGTTTAACTTGAAG
 ATTTTGAATCAGATTTTAAATTTGTAATATCTGGTCAATATTTCAAGGCTTATGTC
 AGTGCACCTAAGCAAGAAATGTAATGAATGTTTGGCTTAAAGAAAGAACTCTT

FIGURE 3W

WO 02/48366

PCT/US01/47559

SEQUENCE LISTING

<110> PE CORPORATION (NY)
 <120> ISOLATED HUMAN TRANSPORTER PROTEINS,
 NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
 AND USES THEREOF

<130> CL001014PCT

<140> TO BE ASSIGNED
 <141> 2001-12-11

<140> 60/254,553
 <141> 2000-12-12

<150> 09/739,457
 <151> 2000-12-19

<160> 7

<170> FastSEQ for Windows Version 4.0

<210> 1
 <211> 1473
 <212> DNA
 <213> Homo sapiens

<400> 1
 atgaccacag ggaagaagaa gaaacgggcc gcgaaccgca gtatcatgct ggccaagaag 60
 atcatcatta aggaacggagg cagccctcaa ggaatagggt ctcttagtgt ctatcatgca 120
 gttatagtca tctttttgga gttttttgct tggggactat tgacagcacc caacttggtg 180
 gattatcatg aaacctttcc taacacatac ttctgatga acggcttaat tcaaggagta 240
 aagggtttgt tgcatttctc tagtgccccc ctatttggtg ctctttctga tgtttggggc 300
 cgaaaatcct tcttgctgct aacggtgttt ttccatgtg ccccaattcc tttaatgaag 360
 atcagcccat ggtggtactt tgctgttctc tctgtttctg gggtttttgc agtgcctttt 420
 tctgtggtat ttgcatacgt agcagatata acccaagaag atgaagaag tatggcttat 480
 ggactggitt cagcaacatt tgcgtcaagt ttagtcaaca gtcctgcaat tgaagcttat 540
 ctggaacgag tatatgggga cagcttggtg gtgtcttag ctacagcaat agctttgcta 600
 gatatttgtt ttatccttgt tgctgtgcca gagtcgttgc ctgagaaaat gcggccagca 660
 tctcggggag caccatttcc ctgggaacaa gctgacctt ttgcgtcctt aaaaaaagtc 720
 ggccaagatt ccatagtgtc gctgatctgc attacagtgt ttctctccta cctaccggag 780
 gcaggccaat attccagctt tttttttatc ctacagacaga taatgaatt ttaccagaa 840
 agtgtgacag cgtttatagc agtccttgcc attctttcca ttattgcaca gaccatagtc 900
 ttgagtttac ttatgaggtc aattggaaat aagaacacaa ttttactggg tctaggattt 960
 caaatattac agtggcagt gtatggcttt agtcagaaac ctggagatgt gttagctgct 1020
 ggggcagtag cagcctatgc tagcatcacc ttctctgctg tcaatgcaat tgtttcaga 1080
 actgtgatgc ctgatcaaca ggtgtctgtt caaggaaatga taacaggaat tcgaggaata 1140
 tgcattgtgc tgggaccgyc cctctatgga ttcattttct acatattcca tgtggaactt 1200
 aaagaactgc caataacagc aacagacttg ggaacaaaca caagccctca gcaccacttt 1260
 gaacagaatt ccatcatccc tggccctccc ttcttatitt gagcctgttc agtactgctg 1320
 gctctgcttg ttgccttggt tattccggaa cataccaatt taagcttaag gtccagcagt 1380
 tgggaaagc actgtggcag tcacagccat cctcataata cacaagcgcc aggaagaggcc 1440
 aaagaacctt tactccagga cacaatgtg tga 1473

<210> 2

<400> 2
 000

<210> 3
 <211> 490
 <212> PRT
 <213> Homo sapiens

<400> 3
 Met Thr Gln Gly Lys Lys Lys Lys Arg Ala Ala Asn Arg Ser Ile Met
 1 5 10 15

WO 02/48366

PCT/US01/47559

```

Leu Ala Lys Lys Ile Ile Ile Lys Asp Gly Gly Thr Pro Gln Gly Ile
20 25 30
Gly Ser Pro Ser Val Tyr His Ala Val Ile Val Ile Phe Leu Glu Phe
35 40 45
Phe Ala Trp Gly Leu Leu Thr Ala Pro Thr Leu Val Val Leu His Glu
50 55 60
Thr Phe Pro Lys His Thr Phe Leu Met Asn Gly Leu Ile Gln Gly Val
65 70 75 80
Lys Gly Leu Leu Ser Phe Leu Ser Ala Pro Leu Ile Gly Ala Leu Ser
85 90 95
Asp Val Trp Gly Arg Lys Ser Phe Leu Leu Leu Thr Val Phe Phe Thr
100 105 110
Cys Ala Pro Ile Pro Leu Met Lys Ile Ser Pro Trp Trp Tyr Phe Ala
115 120 125
Val Ile Ser Val Ser Gly Val Phe Ala Val Thr Phe Ser Val Val Phe
130 135 140
Ala Tyr Val Ala Asp Ile Thr Gln Glu His Glu Arg Ser Met Ala Tyr
145 150 155 160
Gly Leu Val Ser Ala Thr Phe Ala Ala Ser Leu Val Thr Ser Pro Ala
165 170 175
Ile Gly Ala Tyr Leu Gly Arg Val Tyr Gly Asp Ser Leu Val Val Val
180 185 190
Leu Ala Thr Ala Ile Ala Leu Leu Asp Ile Cys Phe Ile Leu Val Ala
195 200 205
Val Pro Glu Ser Leu Pro Glu Lys Met Arg Pro Ala Ser Trp Gly Ala
210 215 220
Pro Ile Ser Trp Glu Gln Ala Asp Pro Phe Ala Ser Leu Lys Lys Val
225 230 235 240
Gly Gln Asp Ser Ile Val Leu Leu Ile Cys Ile Thr Val Phe Leu Ser
245 250 255
Tyr Leu Pro Glu Ala Gly Gln Tyr Ser Ser Phe Phe Leu Tyr Leu Arg
260 265 270
Gln Ile Met Lys Phe Ser Pro Glu Ser Val Ala Phe Ile Ala Val
275 280 285
Leu Gly Ile Leu Ser Ile Ile Ala Gln Thr Ile Val Leu Ser Leu Leu
290 295 300
Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu Leu Gly Leu Gly Phe
305 310 315 320
Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe Gly Ser Glu Pro Trp Met
325 330 335
Met Trp Ala Ala Gly Ala Val Ala Ala Met Ser Ser Ile Thr Phe Pro
340 345 350
Ala Val Ser Ala Leu Val Ser Arg Thr Ala Asp Ala Asp Gln Gln Gly
355 360 365
Val Val Gln Gly Met Ile Thr Gly Ile Arg Gly Leu Cys Asn Gly Leu
370 375 380
Gly Pro Ala Leu Tyr Gly Phe Ile Phe Tyr Ile Phe His Val Glu Leu
385 390 395 400
Lys Glu Leu Pro Ile Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro
405 410 415
Gln His His Phe Glu Gln Asn Ser Ile Ile Pro Gly Pro Phe Leu
420 425 430
Phe Gly Ala Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile
435 440 445
Pro Glu His Thr Asn Leu Ser Leu Arg Ser Ser Trp Arg Lys His
450 455 460
Cys Gly Ser His Ser His Pro His Asn Thr Gln Ala Pro Gly Glu Ala
465 470 475 480
Lys Glu Pro Leu Leu Gln Asp Thr Asn Val
485 490

```

```

<210> 4
<211> 458
<212> PRT
<213> Homo sapiens
<400> 4

```

WO 02/48366

PCT/US01/47559

```

Met Thr Gln Gly Lys Lys Lys Arg Ala Ala Asn Arg Ser Ile Met
1      5      10      15
Leu Ala Lys Lys Ile Ile Ile Lys Asp Gly Gly Thr Val Leu His Glu
20     25     30
Thr Phe Pro Lys His Thr Phe Leu Met Asn Gly Leu Ile Gln Gly Val
35     40     45
Lys Gly Leu Leu Ser Phe Leu Ser Ala Pro Leu Ile Gly Ala Leu Ser
50     55     60
Asp Val Trp Gly Arg Lys Ser Phe Leu Leu Thr Val Phe Phe Thr
65     70     75
Cys Ala Pro Ile Pro Leu Met Lys Ile Ser Pro Trp Trp Tyr Phe Ala
85     90     95
Val Ile Ser Val Ser Gly Val Phe Ala Val Thr Phe Ser Val Val Phe
100    105    110
Ala Tyr Val Ala Asp Ile Thr Gln Glu His Glu Arg Ser Met Ala Tyr
115    120    125
Gly Leu Val Ser Ala Thr Phe Ala Ala Ser Leu Val Thr Ser Pro Ala
130    135    140
Ile Gly Ala Tyr Leu Gly Arg Val Tyr Gly Asp Ser Leu Val Val Val
145    150    155
Leu Ala Thr Ala Ile Ala Leu Leu Asp Ile Cys Phe Ile Leu Val Ala
165    170    175
Val Pro Glu Ser Leu Pro Glu Lys Met Arg Pro Ala Ser Trp Gly Ala
180    185    190
Pro Ile Ser Trp Glu Gln Ala Asp Pro Phe Ala Ser Leu Lys Lys Val
195    200    205
Gly Gln Asp Ser Ile Val Leu Leu Ile Cys Ile Thr Val Phe Leu Ser
210    215    220
Tyr Leu Pro Glu Ala Gly Gln Tyr Ser Ser Phe Phe Leu Tyr Leu Arg
225    230    235
Gln Ile Met Lys Phe Ser Pro Glu Ser Val Ala Ala Phe Ile Ala Val
245    250    255
Leu Gly Ile Leu Ser Ile Ile Ala Gln Thr Ile Val Leu Ser Leu Leu
260    265    270
Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu Leu Gly Leu Gly Phe
275    280    285
Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe Gly Ser Glu Pro Trp Met
290    295    300
Met Trp Ala Ala Gly Ala Val Ala Ala Met Ser Ser Ile Thr Phe Pro
305    310    315
Ala Val Ser Ala Leu Val Ser Arg Thr Ala Asp Ala Asp Gln Gln Gly
320    325    330
Val Val Gln Gly Met Ile Thr Gly Ile Arg Gly Leu Cys Asn Gly Leu
335    340    345
Gly Pro Ala Leu Tyr Gly Phe Ile Phe Tyr Ile Phe His Val Glu Leu
350    355    360
Lys Glu Leu Pro Ile Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro
365    370    375
Gln His His Phe Glu Gln Asn Ser Ile Ile Pro Gly Pro Pro Phe Leu
380    385    390
Phe Gly Ala Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile
395    400    405
Pro Glu His Thr Asn Leu Ser Leu Arg Ser Ser Ser Trp Arg Lys His
410    415    420
Cys Gly Ser His Ser His Pro His Asn Thr Gln Ala Pro Gly Glu Ala
425    430    435
Lys Glu Pro Leu Leu Gln Asp Thr Asn Val
440    445    450
455

```

```

<210> 5
<211> 49984
<212> DNA
<213> Homo sapiens

```

```

<400> 5
agtcctactg tattttttac ttgtattttt gttgttttgt gggattttaa aaatattttt 60
attctgagga tagttgaatc cacaggatac tgagggctag ctgtattcac sacccaaatc 120

```

WO 02/48366

PCT/US01/47559

```

acatacaaaag cgacaagtto atacacaata ggcctattag aacaggactg ttctctcttg 180
ttatccattg cagcctttct agcacaagc ctgggacatt ctggacattt agtatgtgtt 240
aaattttctct tactacaatta ttccaacag tatttactgc aatctgcaat taacctctctt 300
ttgttttgta actgtgtccc caactagaat gaaagctctg tgcagatagt gtctcaatta 360
ttgatgtatc cctggcaatct aataaaacac tgacaacaca agcacccagt aaatattttt 420
tgaatgactg aacaataacc agttcaatag gctgataaaa ttgttatagc tagatgaagt 480
atgattttga gggactatga aaatcaaagt aacocacaa taaattatcc gccctctact 540
tccatttoaaa scaagctcot gggaattgaa ttatgaatc tatcatatta ctttctctaa 600
agaacttcaa gttgggtgtc aactaaaaag ttgcaaggca ggcgogggtg ctcaaacctg 660
taatcccagc actttggtag aactgagtat ccttgaggc caggtttgaa accagcctgg 720
tcaacataac cagactctgt ctttacaata gaaaaattaa aattagccag gcatgtgtgt 780
gtgcatttgt agtccagat acttggaagc ctgaggcaga aggatcgttt cggagagggc 840
tgcaggaggc catgatggca caactgacct caagcctggg tgcagagtg agacccctgc 900
tcagaaaata ataataggcc acgcatgggt gctcaacct gtaatccag caacttggga 960
ggctggggcg ggaacatcac ctacggttag gagttcaagc ctggccaaca cggtgaaatt 1020
ccatctctac taataatata aaaaaaatta gccaggcatg gtagtgggga cctgtaatcc 1080
cagctactcg ggagcctgag gcaggagaa cacttgaacc tgggagctgg aagtgtcagt 1140
gagccaagtt ggcactattg cactgcagcc tgggcaacaa gagcaaaact ctgtctcaa 1200
aaataaataa taataaaagt ttcaaaatga gaatatatgt ttcaaaacaa gtataatgaa 1260
tatacttatt gatgggaaaa tataattaga agtatctatc aggtatataa ttgcttttct 1320
ttctccctcc atggaaatta gtttttttct caatttttag tccgltatga aatacaagg 1380
aaaggaaatt caatcaaat tactttttaa cattttattt ggaataaatt tcaaaattac 1440
agaaagtttg caaacaagct acaagaact catcatttca ttactgttt ttctttttac 1500
cctatatatt agttatttat agctgtgtaa caataaccc caagcttagt tggcttacc 1560
caagtacttt tcatcttaca ctgtttctga gtcaggatlc caggagtgcc taagctaggt 1620
ggctctacct ctgggtctct catgaagttg tagtcagcca aaggcttagc caaggttgga 1680
ggatctactt caaaagtgac tcaactccgt gcaatttgta ggaggctaca aacagttcct 1740
ggacaactgg atctctccat aggtctgtgt agtgctctga aaacacggaa gcaggtctcc 1800
ccaggtctca agccccaaaa taataataga agagagccgc aaaggaaagt gcatgtcct 1860
ttatgacctt gccctcgaag tcaatactgt caactctgtt ttgcttctac aagagtcact 1920
aagctagtc tacaactcag gggaggggaa tttagtccca cctcttccag ggaggaaat 1980
catgaaatct gtgaacatat cttagaacta ccatacttag tttagtact tttaacatt 2040
cgccattttg cttgtccctt cctttttccc caactacata tacatacaca tacatgttac 2100
tccataacca tctgagagta gggagcatgc ggtgtatccc tatccctcgt gttttctct 2160
taagasaag gatattctat tatacacac ggtgttctc aatatctaatt tttaacattg 2220
tgatacttta aagttcaactt caacttggtt aaattgcctt ttctagcaat gttcccatca 2280
aatttctttt taacaactca agtaaaaaa gaaagggcca caaagggtga catgtgtcag 2340
gtaaagtatt ttcttggca gggaaataaa aagttctctg gcttagggag gtaaaacagc 2400
gtgagccagc tgaattctag cgggggtccc tgaacaccaa aaggacaaaga ctgttctga 2460
aacatacat tatctcttaa gttaccatt acttacggaa aatgattttt taactgtccc 2520
ttcgtttcct gtcttggtta gaacacagct ggagattgtg ttaatagctt agaagctctg 2580
ttcogtgag caggttaacaa ctttttgaaa caaattccct catctgctga agaaggggga 2640
caaaaacggc cctatcgcc cagaacogtt gcagagattt agctagctgg tgacgccgga 2700
gcagaaagtt gtacaggtag caagcagcac caacgcagc ccgcggttac cctggccggg 2760
cgctacttgt agagtgggtt ggcggcgagc gggcggggca gtatcacgcy yaaagggcgg 2820
ggcccgctgg tccgtgata gcaagattgt gaagcgccgc cggagcgag cggcgccctc 2880
aagatggcgg cggcgcccca gagcggctcg gcccgcgagt agtgtggga cggcactaga 2940
tgctggggcc tcccggccc ggaatggctg cagcagcgcc aggaatcgag gatgtaaaa 3000
tgacccaggg gaagaagaag aaacgggccc cgaacccgag tatcatgctg gccaaagaa 3060
tcatcaltaa ggaaggagcc acgttgagct gagttccgag ccggcgagcg tccctggggg 3120
ccccatccg gtctctctct cagaccccca cactgccgtc tctaggcgtc ccggtgcctc 3180
cctcccttcc cccacccctgt cogaactgcc ggtgctctgg ggtcggggac cgcagtcgg 3240
ccgtccgggg aatgtgcctc cgtgctcgg gcttggggcc tccggggccc gtctctttc 3300
tttccgcaac ctgctgacct ctgtctggc cgtctctgca ggcctcgag cctcgaaacc 3360
caogtgcgcc tccgcgcggg ggaagaaagt gcggggctcc cccggcgccc cgcgcgcgc 3420
gccctcctc gccgcagcct cgcctcgcct tccgcgcgag gcccccgga gccgtgcgc 3480
cgctgttcaa ggggtctgga accatccctg ctctcccata gttgtctaac ggggtggcgt 3540
ctggcggggg gatcccgctg cggcccgcta gtacgttgc tttctgtttc cagctctctc 3600
tcgctcgggt ctccgctct gggctgctta cagtaaaacc tgacgggaga tgggttccc 3660
tcaattcccc gactcggaag catgaacgca gacacctggg gctacattc gaacctgcta 3720
gttttaaaag aaaaagtcac actgtgtgtc ttaagabcaa aagtattaga atcagtcagt 3780
gcctaagat cggggagga cacttggag ggaagaaag ttgtgtttt agaaacagt 3840
gcatcacag taaactttat gaagtgtgta gtaaccagc tggggagcta ggaattttaa 3900
tgaaaaaaca aaacaaaaca aaactgttt atgctaaccat ttctcgttg ctacactgtg 3960
tggtctttgt tgcatccgct gatcccgct tctgaatat aatggaaagg tgatatatat 4020
gtttcaacta cctgaagtg gcagaaattg taccattaat tccattttct ttatatctt 4080
attggaagcg cgtacaactg ctacacagat agtaaatgtg taagtaggac acaattgagg 4140
atttgcgtaa ttcagttgaa aaactgtaca aaattttat acatttcaga acacggccca 4200

```

WO 02/48366

PCT/US01/47559

```

gtcaatatgc caaagttagt aaaacttgag acatatgtaa tgactttgga atatattttt 4260
agtttaacgt ttattatatg ttatagcttt gacatttatt gaaaaaaga aacaaattcc 4320
tcaagttctt ttatttgaa ctagtaattt aaacatttac ttgtattaga tcygttatga 4380
agagctatag ctcttttgac caagtagtta agacatttgt gggagagaaa atactgtgc 4440
cttgtctac cttagaag agcaatat ttacattctt cataaaatct acaaaatagt 4500
ggcaatgaaa gattgtattt tgtaagacca agtgatattt aagatcagta tttttacaa 4560
aatgtaagaa tgaactgat taagaacac agcttccttt tocttgaaa gttcagtttt 4620
attacotctt tttgggtttt tgtttgattt gctttacagc agatgttttc tttccaaatc 4680
ctgtgagttt tggaaaagat cgtttttaaa ctttctgtc ctattattaa ggttgtaatt 4740
aattcttagc ctgctttggg acacaaaata aatgttttgc accagcaata ggtttcacat 4800
agaacaaatg aagacttttc ttgagggtct tgaacatggg ggcattatc atttctcatc 4860
tttatcacct taatatttca ttctctatct taagagcact gggcactcct ttgaaaagg 4920
ggcttggttt tglagtgttg gatccacag ggcctagat gfgaattta aagatataa 4980
aacacttcca tttgtacta gacacttctc agatgattt ttattgtaatt tttgtttatt 5040
cttatacgta atcagaggat atatttcaat aatatcagg ggaatttttt gcattatttt 5100
tattttaatc catccagct tttaaattta aagatatac taitgcagtc atagaaatga 5160
ttgtaaaatg gtatgtgctt atctacctct ctacttcaaa tagttcagac tactattatg 5220
aacctttttt gttgtttgt ttgagatgga gtctcactct gttgccagg ctggaggagt 5280
gcagggcgag gatctggct cactgtaacc accgctctct ggttcaagt gattctcctg 5340
cctcagcttc cagagtagct gggactacag gcagtgcca ccatgctgg ctaatttttt 5400
atattttcag tagagacaaa gtttccacct alttgccagg ctgtgttga acttctaac 5460
tcatgattca ccaactgtg cctcccaag tgcgggatt acaggttga gccacgtgc 5520
ccagactgaa cattttttaa gaaaggggaa aaaattgcca ttgtatctc tgttgttgt 5580
tgttttttaa ttcatctgat catagaatat ttcagtgcta ttgtgttga cctcagagt 5640
tcagagtttt tataaagttc cggcaatggg tagattcatt cagtgagatg tctgaggctc 5700
tttgtctgtt acatgacagt cgtgaacagt atttccata cctgtcaat ggtactgatt 5760
tgatccctct ctgatttct tcttttcaac aatgttaata aaattcttct cgttgtctct 5820
gctaagtaca tatatgtaag cctatttggc cagtttaaat atttataac aaacttctga 5880
agagtttca atgatttttc tgaatttag agcagatag agcagatttg tagtttcaa 5940
cgtctgaaga aataaactct tctaaatgag cagatttaat cgttaagtac tgattttttt 6000
attgaaattg tatttcaatt aattgtattt cattcagctg aatgaaaaac aggcaggagt 6060
aaagctaaac agtaggtctac ctatgtgagt agacacaaat aagataaatt acattaaggt 6120
gtgtgatttt atattaggtt tttttaacct gggtctgttc accctgaagt tgtttgcata 6180
attttctttt ggcatacat gtttcttggg agagtcccaa aaggtccata cttcccaaaa 6240
gtttaagagc aattgttctg tttgaaaaca gcataagtaa ctaaaagata agttccacat 6300
attatattca gtaaatattt aatcatatc tgtatactac ttcatctgat aaagtaacca 6360
tatatttgaa ttgtctttaa aagcatctac atatagaat agtttttag ctaggggag 6420
tggtccagc cgttaactcc agcacttggg gaggccagtg tgggcagatc acttgaagcc 6480
aggagtctga gactagcctg gccacatggt tgaaccccca tctttaccaa aattacaaaa 6540
atgagctagg tgtgtgtgta tgtgcctgta atccagcta cccggaggcg ctaggcacga 6600
gaatcacttg aacctgggag gccaaagatt cagtgagct agatcacgac actgcattcc 6660
agcctgggtg atggagtgaa actgtctcaa aaaaaaaa aaagaagttt ttagtacag 6720
gttttcatgt atgtaacatt cagtgtaggt alttaagaca gctgaataa aatatccttc 6780
tgacatttcc aatactaga attctgtttt gttttattaa agcattacaa cttgttttta 6840
agcattcttg ttagaggcta agagctaaag agttatttat agttttcaa ttgaatttc 6900
ctatctcttt aaagtctca tcttaaaaa tgacttttat tgttttggcc ataacattgt 6960
ggaaactcat ctctgacagt ggaanaatga tagttcttct agaagtttgt ggtaaaatga 7020
ctttaaagat ttgatagaaa tgaaggcata tctgaattgc atgttcgaaa gtacctgaaa 7080
aaagtaaaa tgatatatca ttgaaaagt aatgcattat ccoctgataa gcagagcacc 7140
agattttttt tttcttgga tccctgattt taattaaata ggaagtcagca accgtttcaa 7200
gagcaggacc caagctctga ccccttgcac tcttcaactg caaggatggc tgaagtatg 7260
gcaggaagac tctctgggt gtagggtctt ttagaccca gagagctgtt aaataacctt 7320
tggttgtag catgcaagca atagaagggt cctgtgttgc ttctctttt ctctctttt 7380
ttttctctt tgagacagag ttgtgtctt gtgtctcagg ctgggtgca atgggtgat 7440
cttggctcnc agcaacctct gctccctggt tcaaggaaat tctctacct tagcctctg 7500
aatagctggg attacaggca tgtgccatca tgcacagcta atttttgat ttttttagta 7560
gagacgggat tttacatgt tggcaggct ggtctcgac tcttgacttc ggtgatcca 7620
cctgctcag tcttccaaag tgggattaca ggtgtcagcc actgcgcctg gccctgtgt 7680
gcttttcaaa aagcctagaa acatcagggt gtttatattg tctttggcag gtgtgtggt 7740
ggcagcatca ttaattactt agctccttcc ctccatgggt cagtttttg tttagattgg 7800
tgtgtttggg gataaatata tatgcagttt tttttcaga tggctatag catccagttc 7860
atctcatgt agttagaaga ctgtcatcac aaataatca gacgctgcg agaatcttc 7920
ctacagttga aatgtactc ctttgcagct actgaagtt taagtttaa gtaaaaaaat 7980
gaatagcttt cttcaggtta cattctgaca agtctgtat atttaaaagt ttaattata 8040
aggaactctg attgtctttt agcattattt taattggaa gtgtgaaagt aacagttgac 8100
agtttcagcc aggtatcatc aagaagagat gaatatgggt ataatatag tctcaaaatt 8160
tccagtactt tatacaaaag aaatatccct cccactgccc tgttttttaa aaataaata 8220
atcatggtt tcttccagc cgtgggaac ttaatagaa ggttcaggag ggacaagtat 8280

```

WO 02/48366

PCT/US01/47559

```

atgcagcata cctgtcattt tccattcaag ttttaactta tttttaaaat ttattatttt 8340
ttaaaatatt tcaatagttt tggggtacag gtgggttttt ggttacatag atgtttttta 8400
gtgatgattt ctgagatttt agtgcactgt tccactggac agtgtatact gtaccataa 8460
tatagtcattt tatcctcttc aagcttcccc cccattccca aagtcocattc tattagtctt 8520
acgcctttgc gtccctcatag ctgaactctc acttgttaag gagaaacatac gaacatttgg 8580
tttccattcc tgaattactc acttagaata atggcctcca attccatcca agtttctgca 8640
aaagacatta ttctattcct ttttatggct aagtattcaa tggatatata acaccacatt 8700
ttctttatcc acttgtttgt cattgggcac ttgggtttgt tccatatctt tgcagtttgt 8760
aatgtgtctg ctataaacat gcattgtcat gtgtcttttt catataatga cttottttac 8820
tttgggtggg taccagtag tgggattgct ggatcaaaaca gtagttctat ttttagttct 8880
ttaaggaatc gcatactgt tttccatagt ggtgtacta gttacattc ccaacagcag 8940
tgtcaaaagtg ttcatgtgc accacatcca cccactctat tatttttga tttttaaaat 9000
atggccattc ttgcaggagt aaatgatata tcaattgtgt ttttaattgc atttccctga 9060
taattgttga tgttgagcat cttttcatat gtttgtgac ttattgtatg ccttttgaaa 9120
aatgtctatt catgtctttt gcctactttt gatgggattg tttgtttttt ttctgtctga 9180
tttgagttcc ttgtagattc tgggtactag tcttttgtca gatgcacagt tcataaata 9240
ttctcccaac tgtatgggtt gtctgtttac tctgtgtatt tttttttttt tttttttttg 9300
agatggaaat ttgtcttgtt ttcccaggct ggagtgcact ggcatgatct tggctccctg 9360
caacctctgc ctctcaggtt caagccattc tctgtcctca gccctccaaag tagctgggat 9420
taacaggcaaa cccacacatg cctggataac ttttttgtat tttaagtaga gacpagtttt 9480
ctctatgttg gccaggctgg actcaaaata ctgaccttag gtgatccacc cgccttggcc 9540
tcccagatg ctgggattac aggcattgct aggcgctat aagttatttg ctttatttct 9600
gggttatctg ttgtgttcca ttggtcttca tgccatttt tataccaagta ccatgcggtt 9660
ttgtaactg tagcttttt tataatttaa agtcgggtaa tgtgatgct ccagatttgt 9720
tttttgctta gtctgtcttt ggctatgtgg gctctttttt ggttccatat gaattttagg 9780
attgtttttt ctgttctctg gaagtatgat gctggtattt tgaatggaaat tgcattgaa 9840
ctatagattg ttgtgtcag tatagtcatt ttcaaatgt tgaattcttc cttccatgaa 9900
catgggatgt gtttcccttt gtgtcattta tgattcttt taacagtggt ttgtactttt 9960
cctgtaaaag atcttctact tcttggta agtgtattcc taggtgtttt gttttttttg
10020
cagctattgt aaaaaggatt gagttcttga ttgtattctc agctttgtcg ttgtggcat
10080
atagcagtgc tattgatttg tgtcattgat ttgtatctc gagactttac tgaatcgttt
10140
atcagatctc ggagcttttt ggatgcgtca ttagggtttt ctaggtatac agtcatatca
10200
ttggcaaaaca gtggcagttt gatttctctt ttccaattt gcattgctgt tattcttttc
10260
tcttgtctga ttactctggt taggaactct aattttttta attactatgg gtacaaagta
10320
gatacagata ttatcaggt acatctgata ttttgataca agcatatggt gatacaggta
10380
taccagtgtat aataaatcag ggatactggg gtatcoatta cctcaaaact ttatcatttc
10440
tttgtgttag gaacatgcca attccacttt tatttttatt tattttttat ttttgagac
10500
agagtctcgc tctgtgcccc aggcagcata catagtacag tagtgtactc cagcctgggt
10560
gacggggaga ctctgtctca aaataaataa ataataaaat aaataaatct gttcagacta
10620
atgtcctaga gtgtattccc aatgttttct tctagtcgtt tgtggtttca ggttttagat
10680
ttaagtcttt aatccatttt gatttgattg ttgtacatgg caagaggtag ggtataat:
10740
ttattcttct gtatatggat atccactttt cctagcacca tttaggagac tatccttttc
10800
ccaatgtata cttcgggtgc gtgtcamaa atgagttgac tgtaaatgca tggatttatt
10860
tctgggttct ctattgtgct ctattgtcta tgtatctggt ttataccag tattatgctg
10920
ttttggttac tatcaatttg tagtataatt tgaagtgaag taatgtgatt cctccaagcc
10980
tcggtttttt tttttttttt tttttttttt tttttatgag acagagtota gctctgtcgc
11040
ctaggctgga gtgcagtgcc gcaatcttgg ctactgcaa cctctgcttc cctggttcaa
11100
gtgattctcc tgcctcagtt tcccaggaa ctgggattac aggtccacc accacgctg
11160

```

WO 02/48366

PCT/US01/47559

gctaattttt gtatttttag tagagacggg gtttcaacttt gatggccagg ctggctctga
11220
actcctgacc tcaggtgatc cgcctcctt ggccctccaa agtgctggga ttacaggcgt
11280
gagtcactgt gccacgcctc cagccttgtt ctttttgctc aggattgctt tggctgttct
11340
ggctcttgtg gttccatata agtttttaga tttaaaagaa aaaattctgt gaggaaatgc
11400
atttgtatgt tgatagtaac tgcattgaat ctgtagattg cttttgtag tattaaaatt
11460
ttaacagtat tgattcttc aatttatgaa catgaaatat cttccatttt gtgtgtgtgt
11520
ctctctcaat tcgtgtcatc aatgttttgt agtcctgtaga catctttcac ttctttaagt
11580
ttattcttag gtattacatc tgtagctatt gtaagtggga ttattttctt ggtttctttt
11640
tcagatattt gctgttggca tatagaaatg gtactgattt ttgtatctg caacttcagt
11700
gaatttgctt ccattctgat agtttttttg tggagtattt agggttctct ctatataagg
11760
tcctgtcatc tgtaagagg gacagttttg acttctgtt ttctaatttg catgcctttt
11820
atttcttact catgctaat tgccttagt ggtaactttc agtaactttg tgaataagag
11880
tgccgaaagt ggccatctt gtctgttcc agatctttga ggaaggctt tcagggtttc
11940
ccgttccagc atgatatgctc tgtgtctgtc atatatggct tttatcatat tgaggatagt
12000
tccttctata ccatttttga gagtttttat gaagcagtg tgaatttttg taaatgcttt
12060
ttcatcttta attgaaatga tcattttatt ttctctcatt cttttgaaat gatgtatcac
12120
cttgatagat ttatgtatgt tggactatcc ttctatacct ggaatgaatc cacttgaaca
12180
tgatgaatga tttttttgtt ttttaatttt ttgagacgga gttttgctct tgytgcacag
12240
gctggaatgc aatggcgcaa tottggtcca ccgcaacttc cgcctccgc gttcaagcga
12300
ttctcctgcc tcagcttctc gagtagctgg gattacaggc atgcgccacc acgcctggct
12360
aattttgtat ttttagtgga gacgggggtt cttcatgttg gtcaggctgg tottgaactc
12420
ctgaacctcag gtgatccacc cgccttggcc tcccaaatg ctggaattac aggtgagagc
12480
cactgcgccc ggccgatgaa tgatcttttt taagacctcc ttctgaagg aggtttgcta
12540
gtattttgtt gaggattttt gcataaatgt tcatcagaga tattgtccca tagtttattt
12600
tgttttcttc catgctagtt tttagtaatt ttctctttaa ataaacaaag cttttcttc
12660
ctaaagtcca agcatgctta ttgaaaaga tatggaaaat tcagaatagc atagttaaca
12720
atgtgatata acttaaaatc attacctaat ataaatttta ttacattga ggtcagttat
12780
tattgttttt cagagtgaa attacctac ctatacatgt tatatctac tttagatttt
12840
aaaaaaatta gcatgcttta agccctgaga agttgtacca agctttctgc taggggctgg
12900
gtatatgtgg tggatgaacat ggtggacaaa acagggttaa agcttatcaa atttgtggcc
12960
aatttttttt tttttttcag agtcctgtgc gccagggctg ggggtcagtg gtgggatctc
13020
aactcactgc aacctctgcc tcccagggtc aagcgattct cotgcctcag cttttctgag
13080
tagctgggat tcacggcaca cgcacccatg cccagctaat tttgtattt tttagtagaga
13140
cagggttttg ccatgttggc aaggtctgtc tcgaactctt gaacctcaagt gatctgcccc
13200

WO 02/48366

PCT/US01/47559

ccttgccctc ccaaaagtgt gagattacag gcatgagcca ccatgcctgg cctcctgtgg
 13260
 tctttttttg accttatatt actatgcttg ttccatttga tactaggcat cactccctcc
 13320
 ttctgaaac tgctccattg tcatatgtga cactgtattc tcttactctc cctgatattt
 13380
 tcttactgtt ccttttatct tctcttttcc ttatcaaaag aagaaaaacc ttcaactatt
 13440
 tctgttgcoct cctacttaaa atgttgccgc ttcttggggt tctgtctcag cccactgctg
 13500
 ttttccactc tgacactcct gataactcca tctactctgg tggtttcaga tatcaacttt
 13560
 gctgctaatt gttttaatca gtgtcaaaag acaatacaaa tgtttcaagt aaagagggca
 13620
 gttttgtaga taggacctga agtaaatctg agcctcgtgg ggggaagtgc tgggaagcca
 13680
 ccagctttaa ctgctagaca accaagctaa acacttggaa gttgttcttg atttccctt
 13740
 ccgtatatta tcaagotcct cccaattcca aatcctgaat ccttaattcg ttccctccct
 13800
 tccaabatto atactgtgcc actgttttat gccctcattt cttgtttgag ctgattcaga
 13860
 tagcttcctt tttagatgcc ttgtctctc cattttatcc tttaggaat caccagagtg
 13920
 ataactctgc agtgagctct aagacatctc tggcagcggg ataaacttaa tttgtattt
 13980
 tctttctcat gtatatcaaa ttccaaatct cttacatact ttgcctgggg attgtctgc
 14040
 ttttgagcca tgttgatata gtgtttatat tttgccact tgcttcattt atggttttt
 14100
 tttttttttt tggttacatc ttggccagaa taactttaaa actttcatct gattgtgtca
 14160
 gtcttaatat ctttttagtg ctcccatagg ccttcagaat taaatataga ctccctagca
 14220
 tggaaagctgg tcttttagta cctgtagctt gtctttcaat acaccaacg tgcagcccat
 14280
 gcactgggtg tactgaactc gatatatgag acccataatg ccgcaagcct tggaaagctt
 14340
 gtacaggctg agccatcttt tccacctat acctcgcct gtctaactct gttgtgtcct
 14400
 ttcaagcttc ctctggaag tctgatatct cccacctccc aagctccctt ggaactctga
 14460
 tgttccaaact gcatactgtg cttatgctaa tgaatttcgt tgttgccttg tcgtccctc
 14520
 tgactttgaa gacagaggca gtgagtcag atgtttgaca cagtgcctag tscatatag
 14580
 atcttaatat ttgttgacta ttaacatcgt tgttattgtt aataattata gaatgtactg
 14640
 ttaacttttt ttaacttttt aaaaaactct gttttttata gcctcaagga ataggttctc
 14700
 ctagtgtcta tcatgcagtt atogtcatct ttttggagtt ttttgccttg ggaactatga
 14760
 cagcaaccac ctgtgtgta agtaactctt taaattattt aacactgact ccaaaatctc
 14820
 ttcttcttca gttttggagg aaatgtggg ccttttccct ttgcacggtt aattctccca
 14880
 ccagtattgt tcagtattca ccagtatttt actggttgtc ttttccact gtttaactctc
 14940
 ccttaacctt ttttggagg ggggtggcgt ggaggtgttt gaatttggac ttgtcactgg
 15000
 gcaagtctca gcagaggctc tgttaactact ctgagtaaaa tggaaagagat tcttaaacg
 15060
 acaggttttag aaaagatgat gtctgtgacc tgcataactc ggcataatta ctttgaggtt
 15120
 catttatgca gctgtacttt ccasaaacag gtttctgttc atttgggcta agtacctaga
 15180
 agggctattc ttttaatagat ctaagctgat tttaccocaa ttctccacag tttgaaactt
 15240

WO 02/48366

PCT/US01/47559

tagaaaagac ctcocctgcc gaccaaaaca ctcagaagat agccagtttt cttatatgg
 15300
 tgtagataag gggaatggaa ggagggaggg actatctatg gtaaatatct ataccatctt
 15360
 gaaaggagta attatgataa atgtacagtt taccaaatcc tagagggaata gagttttaaa
 15420
 gtaatatact atgttttcat gaaggttttt ataaaaagt tatttaatag aaaaattatg
 15480
 taagtagatt gaactagcct aagaacattt acagtacata tttcttgata tatttattga
 15540
 cagctgtgta attgttacta totatacata aaatattgat gtttagcagt tgcattatgcc
 15600
 tgtaatccca gcattttggg aggcctgggtg ggcagatcgc ttgagctctg gagttgagac
 15660
 cagcctgggc aacatggtaa aaccttgcct ctacaaaaaa tgcacaaatt agttgtgcat
 15720
 ggtggcatal gtttgtagtc ccagctactc gggaggctaa ggcaggagaa tcacttgagc
 15780
 ccaggaggca gaggtttagt tgaccgcata tcgtgccacc acactccagc ctggcgagc
 15840
 ggcgtgaaac cttgtctcaa aaaaaaaaaa caaaaaaa aaacagccgg cgcggtggct
 15900
 cacacctgta atcccagtac ttggggagcc caaggcgggt ggatcacgag gtcaagagat
 15960
 tgaaccatc ctgaccaaca tgggtgaancc ctgtctctac taaaaataca aaaaatagct
 16020
 gtgcgtgggt gtagcgaact gtaatccag ctacttgga ggctgaggca ggagaatctc
 16080
 ttgaacccgg gaagtggag ttgcagttag ccgagactgc accactacc tcacagctgg
 16140
 atacagggtg agactctgtc tcaaaaataa aaagtcattt tgaatatata gacatgttc
 16200
 atgagtattg ctatnaaaa atatcagagg gtttttttt tttttttagt ttactgattt
 16260
 cagatagaaa tctttnaaaa attaatctac acatttcccg gcttcaaat ccaagtacaa
 16320
 cgatttggaa ctctctcaga tgaatgaagt tgattatgac attcataact tcatigaatt
 16380
 gtaataacct gtttttgtca aggttactg aagtgctgta ataacttttt ggcctcatga
 16440
 cttaacatta gcttctctaa tgcgccagcg tgctttttat aatctgtcag ttaacatac
 16500
 aaatctgtct ggtagacct cactctacc attbaaagta cttgagcttt gtaatagtaa
 16560
 acagccacagc tgtatttata ttatgatgga tctaggcaca gtcctttaat gtattcaag
 16620
 tggactatgt taagcacagt cotaatgtat tccactttga atcattatc attttttcat
 16680
 cttacaacc accctatcag tgagatatta gcctcstaat acagatgagg aaaccggggc
 16740
 ttagaagaagt taagcaattt gattgctact ctgacagtaa gctgcagtgt tggtatttgc
 16800
 acccaggctt ccttgactcc tccagtgtc agtcttttgg ggaatgcagg tagtaacttg
 16860
 tttgtacca tgttttagat agttgaggtt gtcaggcagc caaccacta gtaagttagg
 16920
 gtgatcaaaa tgtggatgag ctgttagcaa gctatgaaaa aaagcatttt gtgatgttcc
 16980
 cataattgt tatcagtatt tcaagtgtgt atagctattt ttaaaatttg cttottgttt
 17040
 aaattttttt aggtatgtta tcttctgtgt takttttgta catttttttc ctagtggac
 17100
 aaaggagggc tatctttttt aagsacaagg aaggagtccc ctaattaga aaggcttgtt
 17160
 tattcatttt tcatagacta atgtgtctaa tatattcctt tttttttttt tttttttttt
 17220
 tggagaggag tctcgtctgt tctgtcccca ggctggagtg caggagcagc atcttggctc
 17280

WO 02/48366

PCT/US01/47559

actgcatccc ccacctccca ggttcaagtg attttcctgc ctcagcctcc caagtagctg
 17340
 gggactacagg cacatgccac catgcccagc taatttttgt acttttagta gagatggggt
 17400
 ttaccatgt tgaccagaat ggtctcgatc tcttaacctc gtgatccgcc cgccttgccc
 17460
 tccaaaagtg ctgggattac aggtgtgagc cactgtgcct ggcgaatata ttottattat
 17520
 otttaatttt tgttttcttt ttcttttttt ttttattttg ttgtttgtg tattttgaga
 17580
 tggagtctca ctctgttgcc caggctggag tgcagtgggt caaccttgcc tcaactgaac
 17640
 ctcacactcc caggttcaag caattctcct gctcagcct cccgagtagc tgggactata
 17700
 ggcacgtgcc aacataccag gctaattttt gtatttttaa tagagacggg gtttcaccac
 17760
 attggccagg ctggtcttga actcctgacc tcagggtgat cgcctgcctc ggcctcccag
 17820
 agtgttgga ttacaggcat aagccaccat gccagcctg gcatatctac tttttagaag
 17880
 tggacctgtt atatattcag tatatgtcac taattaagaa caatatatta attcaatag
 17940
 ggctttttaa aaaggtttta ctcatctcaa ggctttttgc ttacaatttt tgtttttttg
 18000
 ttgttggcct tgttggcagt ctttgttttg ggcctcagta ctactccac tctctccag
 18060
 cattgtgtgt gagagggtgt taaagggtg ggtttctggg taaagaagg ctctcctct
 18120
 aaaaagctgt taattcttaa acatttcatt tctgttttat gtgttttgaa actgattata
 18180
 aatggtgcat gccacaagag tcaaaagttt taacattcat ttaaaagga aatgaggat
 18240
 gaagacataa ttaatttat attttaagtc agtatcttct attcctgtt ccctccctca
 18300
 acagtatat catagtttgt ttcagcattt cagatttcaa agatatctct tgaagtatt
 18360
 ttttaactag ataaccagtt tttagacatat taattttgaa tglctgggtt gggatttatg
 18420
 atagccttaa tttcttaatt tttaaaaacta atgtgcatt ttaagaccaa aaaaactgtg
 18480
 tgttgcaatt atctttcact ttttagccct catagaacag tcaaaaaaca aaagctgtgt
 18540
 tttgtggaag atctgcccag ggaagatgg tgaacctcta caacaaggg gatttagcta
 18600
 aaaaagga tttgtactg acaaatattt ttaagattg aggtctaaca cttttgagag
 18660
 gttatgaata tatggttgt catagtagat agttcagtca gaatcagtga ttattgcttg
 18720
 attatgtaac atattagcta agtgatgaga ataacagtag gtataaggat ctgtaatgcc
 18780
 aaggagtgga atttacgggt tttttttttt ctttcctttt tttttttttt cattgagag
 18840
 gagtcttaac ctggcatcca ggttgagtg cagtggcgtt atctcggtc actgcaacct
 18900
 ccacggccaa ggttcaagag attctcctgc ctcagcctcc ccagtagctg gaattacagg
 18960
 tgcattgcac cagcccccag taattttttt ttttattatt ttttttgaga cagagtttca
 19020
 ctctgtctgc taggctggaa ttcagtggca ctatctcgcc tcaactgaac ctctgcctcc
 19080
 cagggtcaag cagttctctg cctcagcctc ccaagtatgt gggattacag gcacctgcca
 19140
 ccatgcctgg ctaatttttt tatttctagt agaggcgggg ttttaccatc ttggtcaggc
 19200
 tgggtttgaa ctctttacct tgtgatccac ccacctaac ctcccaagt gctgggatta
 19260
 cagggtgag ccaactgttc tggccggcct tacccttttg acagacctat ggctctggaa
 19320

WO 02/48366

PCT/US01/47559

ataataggcc agtgtttgat ggttcaagct cctagataca cagtccatgt tacggaacac
 19380
 tcasaatcca ctatcatctc ttctacctag atggtttcgt gtccctggct acsgaacaag
 19440
 ccccaagcg tttaacattt taaggattat ttactttcaa catittttaa gttaaaaaaa
 19500
 agttaagatc cataaaattt ttggaaaag tgttacattt tctctgttca cctctaesga
 19560
 ccagtgttaa aggatcctga catcaaaaat ctttacaaca ttogaattac ttgttatatt
 19620
 tgtctgttaa aattttgtta gaaattgtat ggcccaaaag gagaatttgc ttggagaaa
 19680
 aagtttaggt agcagaggaa cagtttggaa ggttggggg ttggccagat aaagaagggg
 19740
 aagaacattt caaaattgaa aggatgcctg gtataaata tgaatttgg aaagcataga
 19800
 atatttcaga aacagtgaag cgaacagatt gattggaatg gaatacagct tggcaaatg
 19860
 aatcattagt gataagatct agcatagtat aaaacttctt atagacattc ataattgttt
 19920
 tcatttcttc taacacaaa cctgttcttc atacctaga agatttggct tgcagtaggc
 19980
 cctatgtgat tattgaaaga saagcataat acattttagt ccgttaaaa gttttgagat
 20040
 actagtttaa ggagttttaa tcttatcctt tagcacaag actgggaana tatggctaga
 20100
 ggactagatc ttttttgcaa tttttttttt tttttttag ttgcttctgg caactttctc
 20160
 ttgtgtgtgt tgttttatatt ccttttcaca agtatgttga attgaacttt tctctaata
 20220
 tcacttagct acttagttaa tgcattgcagt agaactctaa aaagaacttc taggagtttc
 20280
 tcaaagacct cccagtaatt cttttcaatt agagagggca tgcatttttt ccttttttat
 20340
 ttttaataaa tatitttattt tttatttttg tgggtacata ggtgtatata tttatggggg
 20400
 tacatgagat attttgatac aggcatacag tatgtaata ttacgtcagg gtaaatgggg
 20460
 tatccatcat ttcaagcatt tatcctttct ttgtgttata aacaatocaa ttttaggttt
 20520
 tgttttgttt tgttttttgt ttttttgaga cggagtcctg ctcgtcacc aggcaggagt
 20580
 gcagtggcgc gatctcagct tactgcaatc tctgcttctt ggaattcaagc aattctcctg
 20640
 cctcagcttc ccaagtagct gggactacag gcaactgcca ccataccag ctaatttttg
 20700
 tatttttagt agagatgggg tttaaccgta ttggccagga tggctcatt ctcctgacct
 20760
 tgtgctctgc cgcctcagc ctcccaagggt gctgggatta caggcgtgag ccaccacgcc
 20820
 tggccaaatt ttagttattt tcaaatgtag aataaatgtt ggcgttagtc aacctgtgt
 20880
 gcatctcag tactagatct tatttatctt atttttttgt gccactaac catcctctct
 20940
 ccactacce ttccggcct ctggtaacca tcatctgct ctgtctcat gagttcagtt
 21000
 gttgtaattt ttagctctca caaataattg agaactgtg aagtttctcc ttttgtcct
 21060
 ggcttatctt acttaacata atgacctcca gttccatcca tgttgttgca aatgacagga
 21120
 tctcattctt tttctgtgtg tataaatata ttttttttat cctttcattc atctgttgat
 21180
 ggaacttag gttgttcca aatcttagct attaagaata gtgtgcata caaaattag
 21240
 ccaggcatgg tgggtcacac cgtaatccca gctactcagg aggcgtgagg agggagaattg
 21300
 cttgaacctg ggaggcggag gttgcagtga gccaaagatt caccatcgca ctccagcctg
 21360

WO 02/48366

PCT/US01/47559

ggccacaaga gcccaactcc gtctcaaaaca aacaaaaaaa ggaatagtc tgcagtaaat
 21420
 gtaggagtag agctatctct tcaatatact gatttccttt ttttggagg gatatatact
 21480
 agtagtgaga ttgctggatc ataggttagc tccattttta ggttttttga ggagccttcc
 21540
 aactgttttc cttagtattt gtactaattt acattccac caacagtgtg tagtggttcc
 21600
 cttttctcca catccttgcc tatcttttgg ataaaagctg tttttaactg ggtgagatg
 21660
 atatttcaact gtagttttga ttgtcatttc cccgatgac agtgatggtt gagcattttt
 21720
 tcaataacct attgttcaact ttgagaaatg tctattcaga tcttttggcc gttttttaaa
 21780
 aatcagatta tgagattctt ttcttaccaga attgttttag ccccttctac atttttgtta
 21840
 ttaatccctt gtcagatgga tagtttgcag atattttctc ctattctgtg ggtgtctctt
 21900
 tcaatttgtt gtttgcttgg ctgtgcagct ttaaaactga tgtgatctca ttgtccatt
 21960
 ctcaacttgg ctttggctgc ctgtgcttgt ggagtattat caagaaatct ttgccagtc
 22020
 cagtgtcctg gagatcacat actattttta taaataaaat ttattggaa cacagtcaaa
 22080
 cccattcaat tacatacagt ctgcgactgt tttttttttt ccttttcttt tctttttttt
 22140
 tttttttaag acaggctatt actctgttgc tgaggatgga gtgtggtggc acgatctcag
 22200
 ctcaactgcaa cctctgccct ccgggtttaa gcgattcttc tgcctcagcc tctgagtag
 22260
 ctgtgattgc aggcgcctgc caccacgctt ggctaatttt tgtattttta gtagagatga
 22320
 ggtttgcaca tgttggccag gctgtcttg aactccacc tcagggtgac catccgcctc
 22380
 ttccctccaa agtgttggga ttacaggtgt gagccaccac aactggcctc taaattgatt
 22440
 tttaacttaca atgagcacat ttttgtttaa ttctctgcta ttggcaggag aagaataact
 22500
 gaagaaaggg gagcaattct gatccttcta aaggttcttc tgcacatg tcagaaagta
 22560
 tatttagcat aatgtttctt cttaaaggga agaccttccc tacttctctt attaccaca
 22620
 ttccattctt ctgttgttat tactgagcga tagcattgga taatagaagc attagtctt
 22680
 aagtcacaaca ggaactcagt tgcctcatat gtaaatgcat aatattatct aattcacagt
 22740
 gttgggatta aacaggagta catataggct gtaaaaatgg tagctgctgt ttatttttcc
 22800
 agttgcctgg aattgccttt tcaittgatg cattccagcg gttctcttgc tgcceactgc
 22860
 aaaaaattga taccacatga ttgagaaca agccttggaa aggatagaat aactgttat
 22920
 acattttcat aggttgggat ttttttctt tatagaatct ttctagatct acttcgtggc
 22980
 aattaaaaat tactatttaa ttttccaat ctctatctct agataatata tccatctgaa
 23040
 agagaattat aagtcagtta ttttgggga gcagcatagc atagttagta aaaacatagg
 23100
 ctttcaagtc tgatctctta ggttcagctt cagctttgcc atttacttac tgactgtaat
 23160
 cttagacaag atgtttaacc tctgcatac agttttctga tagggctggt atgagggatt
 23220
 aaatgagata atatatgtaa agtgcccaat gtagtgccct gtgacattt aegtaaccata
 23280
 taaatatttg gttattaatg gtcatatgca tgtcatacaa atctgaatat gtaaaatana
 23340
 tcagattgta gtatgaatg atgtcaaaa aggtaaatgt agaaatttta ttaagactga
 23400

WO 02/48366

PCT/US01/47559

```

aatatagcat gtgattttta ttttggtttt tattcttttag gtattacatg aaacctttcc
23460
taaacatata tttctgatga acggcttaat tcaaggagta aaggtaggat cagtcataca
23520
tataatgtg tatatatata tacatacaca tacacgcaca cggatgaaca tacataaata
23580
catatatata actatgcgaa tatatgtttg tttaactgatt aataacttaa tttttataat
23640
tagatggagt atattttgag agataactta atactttcta gacttgagtt ttaaatatga
23700
tctaactaca atatagccaa tcatgcataa taataaacca catcataaca gattccctgt
23760
ttttatgttg gcattttaat tccagagatc tcaatgattt tgtaagacta cagattgaag
23820
aggaataaac tatactaaaa aaccccatc aatataaac tgatcagta gggtaagag
23880
ttggttatta tatgaattct ttgctctttc tttttcagtt tttttttttt tttttttttt
23940
ttttttgaga cagagctttg cactgtcacc caggctggag tgcaagtggc atgatctcag
24000
ctcaactgct agaacctctg cctcccaggt tcaagcgatt ctctgcctc agcctcctaa
24060
gtagctggga ccacagacat gtgccaccac acccggttaa tttttgtat ttttagtaga
24120
gacagggttt tgccatgttg gccaggctgg tctcgaactc ctgacctaa tgatctgccc
24180
tcgaactccc aaagtgttga gattatttgt gcaagtactg tgcccagctc aaattcttta
24240
ctcttgatcc agtttgacaa acaagttttc gaataaggta attagcctgt tttatatata
24300
tataatata tgttttatgt tttatatata tatatacaca cacacataca tatacacaca
24360
catacacaaa caggttaata gcatatgga ttgctattgt ggatttattg tgattgagaa
24420
tttattagag cagttcataa ttaataccta cctggagccc cacagatgat tctaaactat
24480
cttgggaat tttaaatta tatatagata agcaattggt tatttaaaag ttgtgatata
24540
atttacttta gaaacaaagt tgttgaaaat tttctatag gagtaaatg atttattttt
24600
ggttcatgct catagatgtg tggatttcatt ttttttcatt taattttttt agggtttgtt
24660
gtcattcctt agtgcccgc ttatttggtgc tctttctgat gtttggggcc gaaaatcctt
24720
cttgcctgcta acggtgtttt tcacatgtgc cccaattcct ttaatgaaga tcagcccatg
24780
gtatgtgcac atttagatta tagcaactaa atatcacitt cagctattgt tttctaatg
24840
ttcctttatt tctttcactt gtgccatcct ggttatgagg ttttaatttt atttttctaa
24900
tacatttatt ctttcacaca gataagaagg cctctaataa tacagtaggt aatagttcat
24960
aaagatttta gaaatagttg acactgttgg aggcactcag acttctggct aacttaattt
25020
tggaattcag aacccaatt tcaaaccaat atttgggga ctggatagga ttgctaactt
25080
ttttcttgt atagaatgtt aaaaacagac aaaaatttg tcattatcta tattagttag
25140
gaatagggtt tgctatatat caaaaccaa aataagagtt gcttaaaat caaatttctc
25200
ttttcatgtg aaaaaggttc tgcaagttga gcagtctgga gctggtatgg cagttccagt
25260
ggagcagact ttttctatct tctgttctt ccatcctcat ctctcacctc actgaacaaa
25320
gtggcttctt gagctctagc catcaagtca tactactgat agcaaggtgg aaggagttgc
25380
taagaagagc acatcacctt cttttaagga aactttacag aattcccatc ccaactctgc
25440

```

WO 02/48366

PCT/US01/47559

```

ttgcatttct ttggtcagat ttcagggata tgctgtcctt agttacaaga gtctggactc
25500
aaggccaccc ttgattaaat cagtattttt tttttttcca ctccaggaggc aagggaaggga
25560
aagctaacca aaagggaata attgtcctaa agtcagagcc caatttggtt gtagatttat
25620
agattttatc tgactaatgt tctttttact ccaactgagta ttaatgttag ttctacatca
25680
cgatggggtt ttattttccat ggtgttatct acasaggccc aaagggttat atggggcagg
25740
atcttcttct ttagatgag atacaattct aatagacca gcctttgatg aaggccactc
25800
atgcacatga aatcttacca aaaattgaac atggaatgaa caaagaattt gctgtgttaa
25860
gcagtggtgtt agaaaaattt tagtagtgat agtcttggtg gaatgatatt ttgaaagctc
25920
attatgatta tatgtgattt tcagaaacta aacatcagta ttacaataga aacttcttat
25980
tccagctctt ttgggaatta ttatagcaa aatatagttt actcttttaa tattttgttc
26040
actgtttata gtaactgtct attctcagtt ttatgaacac tgaactctgc catataggtt
26100
ttttgtgtgt aaaactgagt tattgtttt gccagcattt gaaaagctaa agataactta
26160
tgaaaaacat ttagtacta tatagaggct caasataaat atgaagaaac ttgttctca
26220
gtcttggtta ctggtatttg ttttttatct agtgtgtggt tgaanaacag taatggatgg
26280
attataatag acttttagtt gttgctgttt ttggcagaaa aagcaaaact acttaacagt
26340
tcaacagttg tcaaaaagat ttgctgnaat acatgatcta attttaagta ttctgtata
26400
gtagctgagc tgctattgaa gggctccttc tagccctatg ttttcataat atctgtggct
26460
tatatttatg gaatagttaa tccatgaact atcctagtaa gctgttgact gaaatgagct
26520
gctcttaacg ttaattaaat tataaaaatg aaagaagatt aaacacatgg taattgtctt
26580
aaccatttct tgttatcttt cattcctagg tggtaacttg ctgttatctc tgtttctggg
26640
gtttttgcag tgacttttct tgtgttattt gcatacgtag cagatataac ccaagagcat
26700
gaaaagaagta tggcttatgg actgggtatgt atgtttatc tatacctttt gtatctgctg
26760
agaaalgctt tgtttttaag ataaatatta ttataaggag tgcaaacctt tgcatataca
26820
gatttttgcg taaaatatat ttgtataaa aatcattcat tagactacat ttaaaatttt
26880
tttgcggtat gaggcctatgt aagttttgat tcttttcatt tagtagatat tcaatagtca
26940
catgtcagaa ttgaaattat agtatatttt acctttaga gttcttttta acagaatcct
27000
aaaaataaga attatttagt atgtcaagag ttaaaaaaaa tcaactactca ttaattgtct
27060
aatctaaaat acaacaggct aacatctagc tcagggatca gcaaaccttt tttgtggctt
27120
tgtgggccac gtacagtcct tglctcattc tttgttttt gcattgtgat ttatgtttat
27180
aaactcttta aaatgttag aaacagccag atttgagcca tagtggtagg tgcacaactc
27240
ctggacactg ttttggtaaa ctaaaattat gcagtataat gtgtcatcta tcaaatctag
27300
gaattaaagg aaaaagcctt agtaatagaa tgactactat aggcacaata atagatcact
27360
actgaatagc cagaantagg acagtgtatg atttcggtaa atgtgagaca aataccttgt
27420
gataaataag gactgaatat tgtgttgggc tgaattagtt ttaaaaggga ctgattctg
27480

```

WO 02/48366

PCT/US01/47559

attcaaaagga cgttatagt aagaatcata agatttttgg ggaggaaaca cctatagaga
 27540
 gaaagttaga aaaaagaacta ataattcttg gctgtttcag tggtctacac ctgtaatctc
 27600
 agcaacttgg gaggttgagg caggcggatc acttgagatc aggagttcac gaccagcctg
 27660
 gccaacatgg tgaasccttg tctctattta aaaaaaaaaa aaaaaaaaga gtgaagaaga
 27720
 aaagaactaa tgatttcagt tgtaacttg gaacattaaa tgatacagg ctgatgatag
 27780
 ccaggatatt taaaaaatag tctaattaag ctatagtta cataccataa aatttatcct
 27840
 ttttatgagt atagttcagt gaattttagt aaatttatac tgttatgcaa acaccaccat
 27900
 aacccaattt ggggttggtc ggttggttgg ttggttggtt ggtttggttt ttttgacgta
 27960
 atttattttc ccatagccaa agttttgaaa ttaacaattt tcaatctgga ggttctgtgt
 28020
 atttaagccat gttctggcaa aaacaaaac aaacaaaac aaacaaaac aaacaaaac
 28080
 acactgaat cttctagaaa taatatggat gcagaaaaa ggtggggaag tggcaggca
 28140
 cagtggcatg tgctgtaat accaccagt tgggaggcca aggcaggggg attgcttgag
 28200
 gccaggagtt tgaggctgca gctatgatca tgccactaca ctccagtcta ggttacagag
 28260
 tgagcccttg tctctaaaa aaaaagttg gagggggcag gtgcagtggc ttataatccc
 28320
 agcaacttgg gaggttgagg caggaggatt gtttagagcc aggagtttga gactagcctg
 28380
 ggcaacatag tgagaccoca tctatacagg aactttaaaa attagccagg tgtggtgtg
 28440
 tgtgctgtga gtcccagcta cctgggaggc ttaggtaga ggaatccttg ggcctgagag
 28500
 gttgaggctg cagttagccg tgatcgacc actgcgctcc aacatgagcc acagagcgag
 28560
 acctgtctcc aaaaagggg gttgggggg gcggggtgac cctgtgtatc tttttctga
 28620
 gcagaaagaa atggctacca agtggagaga actgaggaga agggaaatga catgaacaa
 28680
 ctgtactgac ttgctcactg tgtcacaaat gtgactctg taaatgccct caaatgtctt
 28740
 cagttagccct catagttaga accattttcc ctttccacac acttgtgcca gagccctgct
 28800
 gagatctggg tccctctgaa accacaccta gggctgcaat aacaaaataa ccactacatt
 28860
 tgaatatata tatttatatg tatgtgtgtg tgtgtatgta tgtgtgtgta tatatatata
 28920
 gtttgttttt tgttagagag gagtctcgtc ctgtcaacca ggctggagtg cagaggtgtg
 28980
 atcttggctt actgcaacct cggctccttg ggttcaaacg attctgtctg ctcagcctcc
 29040
 ccagtagcta tgcccaccac catgcccagc taatttttgt atttttagta gagacgggtg
 29100
 ttaccatata tggccagtct tgtcttgaa tctgacctt tggctcgctt gctcggcct
 29160
 cccaaagtgt tgggattata ggctgtagcc atggcgcttg gcccccatgt gaatatatta
 29220
 aataccattt aaaaaccac cacsaccag ttatagaaca ttccatcag cccaaatgt
 29280
 tccctcagcc ctgtttgcca tctgtcccca tgcctcacct gtgacccaa gcaaccaaca
 29340
 atttagcttc tgtcaccatg gttttgcctt ttctagaac ttcatagaaa ttaataata
 29400
 caaaacatct tttgtgcta acttcttcca ctggcataa tcttttgaga ttgatccatg
 29460
 ttgatactat agatcaatag gttctatttt tgtctctttt cctttttttt ttttttgaga
 29520

WO 02/48366

PCT/US01/47559

cagggctctg ctctatcccc caggctggag tgcagtggca tgcacatggc tcaatgaagc
 29580
 ctgggcttcc tgggctcaag cgatccttct gctcagcct ccaaagcagt tgggaccaca
 29640
 ggcatgatcc accatgccca gctaattttt ttctttttga gacagggtct cactctattg
 29700
 cccaggctgg agtgcagtg tgcctttaca gctcactgca gctctgtct cccctctacc
 29760
 tccctgcttc aagtgatcct tccacctcag cctcccgagt agctgggact actaattttt
 29820
 gtatgttttg tagtgatgga gtttcccat gttgcccagg atggtctcaa actcttaaac
 29880
 tcaagtatc tgcctgtctc agcctcccaa agtgcaggaa ttacaggcat gagacactgc
 29940
 accctggccag tagttttttt tgattgtgtg gtatgtatt ctatccatc agttgatgga
 30000
 catttgattg atagctagat gtttgaatt actagaattt tatgtacttg ttcaataaat
 30060
 tgacctttga aaattgaatt gcttgcttta agcaatagag ttgcaagtaa gcattcttgt
 30120
 gaagtttaag ttctccatcc aaaagtcaaa aatggcatag aaacagata aaattccaac
 30180
 attaatctct atgctttgaa agaataatgt ccttttccct tccctccct ccccttccct
 30240
 ttcttttgcc ctctctctcc cctcccccct cctcccccct tccacttttc actttcactt
 30300
 tcccttttcc ttctgctttt cacttttctc ctctcttttc ctctctctct ttcccttccc
 30360
 ttcccttccc ttcccttttt ccttttccct tccctccccc ttctcccttt cctctctttt
 30420
 ttctttttct ttctttttcc ttctccctcc cctttccccc ttctcttaag ccttttccct
 30480
 aagccttttc ccttttttaa ccttttccct ccccttttct ctcccccct ccccttttcc
 30540
 ttctctctcc tctcctttcc ttctctcttc tctctctccc tctctcttc ttcccccctt
 30600
 ctccatttct ttcccttttg ctctcttctc ctcttctct tccagacagg gtgttgcca
 30660
 gactggactc actcttgga tcaagtatc tccacactca gctctctgag tagctgggac
 30720
 tataggcagg tgcacactca cctgactaag agtgctattt ttatgaagt ttctctgtg
 30780
 tcacatctgc taatttgtag gctgtgtcc agtaggctag aaatgctgc ggttaacagg
 30840
 ttgtctctac tctgttctct ttcaacttta atcttctct tccaccaggt taaaaaata
 30900
 gacttctcta gattttttaga gatgttctta atttatctgt gatttcttc ttctaaacc
 30960
 tgccaactaa aaagattacc aagctcagtt ttgttccagg gcttaacata ttattcatga
 31020
 gaacagggaac ctccaagctc ttaagcttta ttccagctag ccttccagta tgcatacaga
 31080
 taaacgttca ttaatttta atattgaaa agtcacagt aaattggatt tcttagagc
 31140
 agtggaattt agactccttt cccagagagc acttaagggt ttatggagat gcccttaacc
 31200
 aagcttgtcc aaccatgggt ccacaggctg caccatggc ccaacagaaa ttcataaagt
 31260
 ttcttaaaac attatgcagt tttttttct ttaagctcat cagctattgt tagtgtattt
 31320
 tatgtgtggc ccaagacgtt tcttccagcg tggcccaggg aagccaaaag attagacacc
 31380
 cctcccttaa ggaccagcat gactggcagt caaggagggg tgtttgtaca gtgcccaggc
 31440
 tctcaaccct tctcaacta aaagagttaa aaatttaaa taggcccggc atggtggctc
 31500
 acgctgttaa tccagcact ttggggagcc gaggcgggcg gatcacagg tcaggagatc
 31560

WO 02/48366

PCT/US01/47559

gagaccatct tggctaacac gggggaaacc coactctctac taaaaatata aaaaatttagc
 31620
 cgggacgaggt ggcgggcggcc tgtagtccca gctattcggg agggctgaggc agggagaatgg
 31680
 cgttaacccc gggggcggag cctgcagtga gccgagatog cggcactgca ctccagcctg
 31740
 ggcgcacagag cgagactccg tctcaaaaaa aaaaaaaataa tttaaataga
 31800
 ggcagggtct tgctgtgttg cccaggctgg tcacaaactt ctggcttcc gcagtcctcc
 31860
 caacctggcc tcccgaagt ctgagattac aggcattgaac catcacccc agtcttctta
 31920
 aaaaaatctc ttttacctat gaatttgcca gtaggattta ttggaacaga gggctccaag
 31980
 gcttagaaga tttagaagaca gtgtcctgag aggcctatcat ttattttatt ttatttttga
 32040
 gatgcggtct cactctgtca ccttggtcgg agtgcaagtg tgctgtcatg gctcactgca
 32100
 acctccgctt ttctggctca aaggaattct gccacctcag cctctgaagt agctgagact
 32160
 acagggtgac accagcatgc ccagctaat tttctttttt cttttttgat acagacaggg
 32220
 ttctccatg ttgtccaggc tgtttttaag gcaagaatct aattctttac ttctcctgcc
 32280
 aaaggagaga gtataagaaa agtggggcca ggcttgggt ctcatacctg tactccccgc
 32340
 ccttcggggg gctgaggttg gaggatcgct tgagctcagg agttcgagac tagcctaggc
 32400
 aacttagcgt gacttccacc tctataaaaa ataaacaaaa tttagctggc gtggtggtgt
 32460
 gtgcctgtag ccccatcagg agatcttcag gcagggaagt ctacttgagc ctgagaggtc
 32520
 aagactacag tgagccgtga tggcaccact gcactccagc ctgggcgaca gagcaagacc
 32580
 cagtcccccc actctcgccc ccacaagaaa aaaagataaa tggcacaggt aggaagagaa
 32640
 aagggaaggt gtgcaacaga aggcctgaca taaatcaaga ttatgaagg agttatgttg
 32700
 tgttagagaa aaaagtagcc tgactaatct ctgtctatcc ttaatttatt gcaggtttca
 32760
 gcaacatttg ctgcaagttt agtcaccagt cctgcaattg gagcttatct tggacgagta
 32820
 tatggggaca gcttgggtgt ggtcttagct acagcaatag ctttgctaga tatttgtttt
 32880
 atccctgttg ctgtgccaga gtcgttgctt gagaaaaatgc ggcagcatc ctggggagca
 32940
 cccatttctt gggaacaaga tgaccctttt gcggttagtt tatacttttt ccttctcctt
 33000
 gataaaaaag tgcatgattc agtgagcat taatatattt ttgtggtat ttcttttagg
 33060
 aaaaacatctt gggtttttct tttaacattt tgaataactt ttcagaatag ttgggggaat
 33120
 atgaatatasa taaaaaggac aactagtgtt ccattgagtag actgccaaaa ggaatctctg
 33180
 catattctta agaatagttc agtgggtttg atagaaacc ttatatcaat cagtcttttt
 33240
 cttgttctag tcttaaaaaa aagtgggcca agattccata gtgctgctga tctgcattac
 33300
 agtgtttctc tctacctac cggaggcagg ccaatattcc agcttttttt tatacctcag
 33360
 acaggtaaaa tctcttcca ctaagggtga ctttttttct attgtctagt gctttaataa
 33420
 aaatatlttaa tcttgagaga actgtaatag aagtggcctt taaaaatga tatcattgga
 33480
 cttggtatag tggctcacgc ctgtaatctc agcacttttg taggccaagy tgggtggatc
 33540
 agctgaggtc aggagatcaa gccagcctg gcccaacatg caaaatccta tgcttactaa
 33600

WO 02/48366

PCT/US01/47559

aaatacaaca actagccagg cgtagtggcg ggcacctgta atcccagcta cttggggaggc
 33660
 tgaagcgagg gaatacattg aaccaggagg gcggaggctg cagttagcca agattgcgcc
 33720
 attcaactcca gctggggtaa caagagcgaa actcgtctc aaattaaat aaataaataa
 33780
 ataaataaaa agaataattat ttggtctact agactttacc tcctattctg tgtggctgat
 33840
 gttccttatg catttctaata gggggttatt tggatatata agatatttgg cttaggggaga
 33900
 aagatagttt tcctgttcca tatagtggtt ttgagtttgt tggctataga attgatggga
 33960
 tgaatttaaco ccttcacotg ctccagcttc ttgtgtattt agagcatatg taagtagagc
 34020
 agctagccaa aatgagagca aaacaagta tttctctctt tgacactagt ctacatagac
 34080
 ggaatccaag cctttaacca atacatgtaa aatgcacaaa atactgcaat atttatttgt
 34140
 aaaaatgattc tgagttcttg ataagtatct ccaatttagt ataccacat tgaggggacc
 34200
 accatggata agtaggcatt tttagtatgt taagatatgt tgtatttctc ctgagggaatt
 34260
 cttgttttat aaatgaatta catttatttt tttctggccc attaaatgtt aaatacaag
 34320
 tagrtgcagt ggttttctat ctggtatata ttgtatttca ttaagtgtac ctttctgga
 34380
 aaccttctca gataatgaaa ttttcaccag aaagtgttgc agcgtttata gcagtccttg
 34440
 gcattcttcc cattatttga caggtgagtt tcttttttta gttagagtga tgtcagtac
 34500
 cctggctggc catccaaact ggggcctcat ctagtgtagg tatcttggtg gaatcaaca
 34560
 agtcaggagt ctgaggttga taggttcaga aattcaattt gtttgtgtag ctagtgttga
 34620
 tcagatittt gggactgaca catttaatat aggatataata aacgtaaaag ctagtittat
 34680
 gatgtatact aggatataat tctggatgaa gttagaagtt tcatgtcttg tagtccttgt
 34740
 cctatttatt cactcattca acaaatattt attgaatacc taccgtgtgt caggcactgg
 34800
 ggatacagca gtgaacttaa ctaaattctt gctttgtag agtttacatt ctggtgggat
 34860
 aagaacagaaa ataaaaaac caatgtaaac atatcagatg tattgagaac tacagagaaa
 34920
 aagctaaggg tcatatggag agtgacttgg gagaatggtc ttaaggccac agggggcatg
 34980
 gttgagaaga ctttgctaata gcaagtgac atttcagcag agacgcaag gtatgaggaa
 35040
 caagctgtag gattatcttg aagaagagaa ttttaaacag tacaaacaaa atttccagga
 35100
 gagactgtag ctggcttctg taagaaacaa cagggggcca gtgcggctcc agttgagtga
 35160
 acagttgaga gcattcttag aactgaagtc caagaggtag tgaggagcca gatcgtgcgg
 35220
 ggcttacagg tcctcataag acttgagtaa gataggaagt cattggtggg ttttgagcag
 35280
 cagaatgaca tgaaagtgtg tatcttggca gccagttgg taacagttg ttgtgtgatg
 35340
 aataagatat taatcaacac aggaatttgg attttctgag gagatatttc atcctggcct
 35400
 ccaactgtt tctgtttaac aataagaata ctatottttt tccagtgaac accataattc
 35460
 tctcacatca cagaacattg tcctccttcc tgtgaaaaag cccaccccc tttcttgcta
 35520
 ttggcttcc tgtttctaag acactatgaa gacaacttag tctaagttaa tacttttct
 35580
 cacttagat gtaacttact agattacaac ttagttttat tctagggtat ttttaattga
 35640

WO 02/48366

PCT/US01/47559

ctcttggatg gttccaccatt tctcgagtag gattccttct gcaggtctca tgattcattc
 35700
 tgttttgggt agtttagcaa caaatttcaa atttaaatcc tatatgcctc cccaagcctc
 35760
 tgcacataca tatacttgggt gttgagtatt agtactattt ctgcttcag gttgttcta
 35820
 aaaatcatca gtctggaaaa acaatgcatt taaatattca ttcttagcca tgagaaaaagt
 35880
 gctttttaac tttggaggaa aatatactgt agcctttata taasaatggc tttaaaaaa
 35940
 gtttttgggt ccaggtgctg tggttcatgt ctgtattccc agcactttgg gtggccaagg
 36000
 tcaggggaccg cttgagccca ggagttcaag accagctcaa gcaacttggc aaaacccat
 36060
 ctctaccaaa aaaaaaaaaa aaaaaaaaaa aaagaaagaa agcccggtgt ggtggtgtgt
 36120
 gcctgtagtcc ccagctactc aggaagctga gatgggagga ttgcttgatc ctgggaggtc
 36180
 gaggggtgag tgaagccacag ttgtcccatg gtactccagt atgggcaaca gaatgagacc
 36240
 ctgtctcaaa aaaaaaggtt aaggaataa caccgttagt atgtgtagaa cttgatgaat
 36300
 tatcaaaagt taacccaacc ttgcaataga tgtgatcaac ctgggcatga gtatcttttt
 36360
 catacattgc caacattaga tttgtcaaca ttttgttaa gatttagatg aagagagcag
 36420
 actaatctg taatgacaca taaaagattg atagctataa ggtottaagt tctgtttctt
 36480
 acttaaatga ccatgggagc tgtatgcac taataaatgt agtcagtac actgcagacc
 36540
 cagtgtatgag tggagggtgc ttttggggtt attttccctc tgtttagcag atggcatctg
 36600
 gcactagggtt acaagatgaa aacagctctc tgagagtcca gaacttggca gggcagacag
 36660
 gtaaaaggaga ggaatgtagt tggatttcta ttgaaagat cagtgcagca gcagagccct
 36720
 gaaggcagta agaccgcaag gtgcaagcaa cctccaaac accattgctg acacagcatg
 36780
 acacacacag tccatgtagg aacaaggtcc ttaagtgaac atttaggttt tgggtgctat
 36840
 gtagaagtaa gaattaacac tttatattcat atatgagcta ctacattatt acatgtatag
 36900
 cctcaagggt agaagaatga tcttgtatc tcatttatgc agaaatatac aattgagaat
 36960
 gacttgtccc ctgttgacct tccataccac tgaagacca gtgtatttct cctctctctc
 37020
 cagcagcacg gagctgctaa tagctctcca ttaetgcatg tttgtttat ttcttaccac
 37080
 cgtgcttttg ttctgtctat tctttctgct tcccttctc ccaccctaag tagccctttt
 37140
 ctgggtcccc atgctcatgt ggcataatct ccatcattgt acagaataca gtgtagttaga
 37200
 gattttatct ctgtttattg ctttagtttg tgagctctag cggacctctg agtagatgat
 37260
 gaggtcagga ttatatcata ttcatttttg tcaccctagc accctgaact gccaggagggt
 37320
 gctaaactaa aggtcatttg tttttttcca taatgttaaa aaaaaaaact taataaaaa
 37380
 ttagtagtaa gcactttaaa attgggaagt gttacgtgaa attatagtta tgtagcttct
 37440
 cccccaaaat gattagatct ggtaacoggg tgtgggccaa ccttccagcg agagccaagt
 37500
 agttgctgtc ccttttgga ccttttcttt ttggtttatc atcagcccca ttacttctgt
 37560
 ggcaccggta cgtataagag ttccatacac ttgcactaag taagtgttta catgagaaca
 37620
 tcaatatagt tctacacatt tctttttttt cagtgttttc ctatccagcc ctctctgttg
 37680

WO 02/48366

PCT/US01/47559

ggggtgactcc cacacttact ctccagtc agtcctctgag ctccatatg acacattggc
 37740
 tgggtatctc agcottaaca tggccaasat taaaatctgg gtccatccc ttgcccacca
 37800
 cccctatgct cctcatctca ttcagtggct tcaccaccgc caggttttgg gggccagaag
 37860
 ccttagcgac attcatgaat ctttctccc taaccttaca ttcagcccat caaatgatac
 37920
 ttccatatac ctctctctcc aaaatatatc ttgaatcaga gcgtttctga tattctccat
 37980
 tagtaggaac ccaatctgag cgtggccat atcttccatc tggctctctc gcttccatct
 38040
 tgcttactta tagttctatc tgcacttggc agagtaattt ttgtaaaatg gaaatttaat
 38100
 cgcattctac ctataactca ctttctcttg caaccagaat aaaaatctaga ctcttatta
 38160
 tgtcattttc ctccactctc cacatgggtg agtatgttct gtttcagtcc agggcctttg
 38220
 caattgctgt cagccttctt tgggggtgtc tctttcccca gatctttgca taactaggtc
 38280
 tctcccttta gtgagctctc acttcaaaag gccctccctg gatcctgggt taagcagcat
 38340
 ctccatcaca ctgcctgttt ttattttgtt catagcagct acaacctgga atatcttggt
 38400
 aataaagcag gctcatatgc atcttctctc catgagaagg tagtccgtgt gttgatggat
 38460
 actttgactt actcaccatc gcgtcttcag gccaaagata agagttggtc ttattaggaa
 38520
 ttggccaat atttgttgac agactaacca aagctgagtg tagtgtagc ttactgttc
 38580
 atcagctatg ccatcttctt taagtctatt aaectttggg actcagtgct gtcctctca
 38640
 aaatgagagt taaattaaag gacacaaaa atttttccg ttcaagaaca ctatatatt
 38700
 aagtattttg ttccattatt attaatatta ttgagaatat tataacattt agaattatgc
 38760
 atgctttcca taaacactta ttaagaactt actgggtgct ggggatataa atgtaataa
 38820
 gagaaaggt cctgccttca agaagaaaat cagtgttccc agttacagtg ttgtaagtat
 38880
 cgytgtcttt tagggctttg ggtcacaaga taggctgtag gggaaggagt ccaagaggag
 38940
 gtctccacag cagtgggctt gctgcagtaa ttctgcacat aggaacgttac ccaggaagag
 39000
 ggggaattgg gtatctcatt gtttccaatt ttaatgactc atctcatggc cttagaazaa
 39060
 tatattcttg gccgggcacg gtggcccaca cctataacct tagcactttg ggaggccaag
 39120
 gtgggtgaat cactgaggtt caggagttcg agaccagcct ggccaacatg gtgaacctc
 39180
 atctctacta aaaatgcasa aaaaatttag ctgggcatgg tggcacgtgc ctgtaatccc
 39240
 agctactcgg gaggtgaagg caggagaatc acttgaacct gggaggtgga ggttgcatg
 39300
 agccagcatt gcgcacagc actccagcct gggcaacaag agcgaaactc catctcgggg
 39360
 ggaaaaagaa aggcctgggca cgtggtggct cagcctgag ataccagcac tttgggaggc
 39420
 cgaagcagct ggatcacaag gtcagaagtt cgaaccagc ctggccaata tggtggaacc
 39480
 ctgtctttac taaaaatca aaaattagca ggcattggtg tgggcacctg agtccagct
 39540
 actcgggagg ctgaggcaga aaaatcgctt gaaccccgga ggcagaggtg gcagtgaacc
 39600
 aagattgtga cactgactc tagccttggc aacagagcga gactccgtct caaaagaaaa
 39660
 aaaaacccaa aagaaaaaga aaatatattc tccagttaat cttatctata aaaaaggaaat
 39720

WO 02/48366

PCT/US01/47559

gaggttaata atgcattcta agcctttttt attgaattgg gatttattct ttgagaacag
 39780
 ctttccacaa aggggaagat agtcatttct gcagataagt acttactggc tagatgggtt
 39840
 gggtgaaggc ctatgagatg accgcatttt ataagtactt tctgtaata ttaatgatct
 39900
 ctgcttgaga agtgcagctt ttcttagact agcatttctg aaagaacacg tgcctocagg
 39960
 tacatggctg gcccatggca tctgtttccc actgagctgg gaggtagtgc tcagccttct
 40020
 tcaactgtct gtctcttggc tgaagtggca gggatttcat cattaagtga aatctatttt
 40080
 ttaacagcat ttcttcttgg taatgcttat catcatctca atgaatgatg agaacaatg
 40140
 ttttctgctc tgtgagttcc tagagccata taaagaatca cagcttttta gatgaagtg
 40200
 ctaccttccg ggatgttctt aaaagtagtt tcccagaagt tcttcaactt tgcattatag
 40260
 ttccctgttc ttctcagaca gttggaggcc acagagcttt ggttatactt acagtttct
 40320
 ttttcataat taattgaagc ccagtcctct atatatgcca caaataaaat catttttaat
 40380
 tatttttaac cctaattaga atcctaatac ttctaatctt tctgtattat tttttataat
 40440
 attggccacc agttccacc cacaagtcac gggtagcttg agatggtctc tgcacccag
 40500
 gctggagtg c agtgggtgga tcatggctta ctgcagctc gacctcctgg gctcagtg
 40560
 tctctgggac tcagcctctt gaggtagctg gaccacaggt gbtgcccac agcctggcta
 40620
 atttttcaat ttttttga gaagtgtct ctttatgtg tccaggctgg tctcaaatc
 40680
 ctgggttcaa gcagtccttc catctcagtt ccccaagtg ctgggattac agatatag
 40740
 cactgtgctt ggccataaat ttctttcag ttgaagtitt tctcagctaa ctgctggct
 40800
 tgggcaagtt tctttttctg gtttggttgc cgaagaatat aattctacat ggaactcag
 40860
 ctctactca cgtgatttaa gtgaagtct tagacaagaa agctgtgagt tctaattgct
 40920
 caaaaatcct tgaatgaatc atttgctttt cactggcata ttgtgatta aattcttaag
 40980
 tggccatctc aatttaaaga attcaaaact tatattttat gaggttttaa gtgtcagccc
 41040
 atctataaat agtgatttcc tgaattattt ttacttctct atggacttac tagctatctt
 41100
 atgtagtata ttaaaaaact tagttagaat agcaaaaaaa aattttctaa tgttcccaa
 41160
 tcaactgctga accttgttga ctttgaaga aaaaggaggc acaagaaac ccacccactg
 41220
 ataattattg ttattaaacg ctgattattg ccaggcacia ttctaagaac tttatgtaaa
 41280
 tatatctcat ttaattccca tgacaagggt ttgaatagg tgcctttatc occatttgca
 41340
 aagagacaaa agtgaggctc aaagtgaagt gaattgctga gagccacaca aagccaagat
 41400
 tatgatccag gctgtttttc taaagtctgc tgtatagtat actgcattta tccnaatca
 41460
 agcttattaa tttactgttt ataaaaggca tcatgggttc aacaacagat taacttaggt
 41520
 aaataatata tgggtaatac ttgttctgtt agtttttctt tagttgtgga aataagcatt
 41580
 tttagactaa tggaccta acaagcttta aggtatttat gtaatgggga tctcaaatc
 41640
 attagttaga acctttgacc cttccatttt caactactga ttttaagtggt cctcagtaga
 41700
 aatgtactga ataggaagtt ttatctttca gttttctaac tctcagctg gatctatgtc
 41760

WO 02/48366

PCT/US01/47559

agcagaggga cttttcatct gcttatgtga cctggactag tgatctcaga catattcagg
 41820
 gcaatttattg ctgaaataca gccaaattgt agaaaagtgc caatagtcct tttatagtgt
 41880
 agattgaaag aagtcacttt ttaaaacttt attctgataa atcttttttt ttttttttca
 41940
 gaccatagtc ttgagttttac ttatgaggtc aattggaaat aagaacacca ttttactggg
 42000
 tctaggattt caaatattac agttggcatg gtatggcctt ggttcagaac cttggtaagt
 42060
 ataaatattt taatgttaat atttttaatt ttggtgttag cccttggtgt tttatttgct
 42120
 tctcaactga ggggtagact gtaatctgtc tcatactatg ctttttatct ttcaaaatgt
 42180
 gtctaataata agtcctgccac ttgtatatatt atatgttctc ctagaatggc ttgaggatta
 42240
 aaaagggtgac cttttatagc taaatgacag gctgaatttt tgaatgagat tatacagctt
 42300
 ttgaatcttt aaggagcatt taatctaat cagtccgtta ctaggaaaaa gtatgtaac
 42360
 ccattgcaac aaggccctga aagtatttta cattttttgt ttttctgttc aggaaaaaga
 42420
 aagttgtata accagtgatc attactgata gcaaacaga agtgaaaaca atccagttat
 42480
 ttgggtgccac gcttcactgt gtgtctcagc tttctgacc agtcgggttt cctgcagggg
 42540
 acttgagtag tgacgcttgt tgtcaggctg ccacagtaga tagtattatt gtgctcaggc
 42600
 attatggcac tgaactccat ggtttgact aatatltcaa acaactgact gcctgtctc
 42660
 tgttttgact tggatataag caagcatcaa gagggtagt attgttctc caaactagcc
 42720
 ttgcacaaga ggtgctcaca attgaatta cctaaacat ttcttttaa catcaagcca
 42780
 ggaacatcca agttacttgt tctttacaat ttaaggatta gatcaaatca gogatattct
 42840
 caaaaatcca tccataagaa ctttgccaaa gacttggttg cttcatgtgt ttggaatat
 42900
 ttgatgatgt ttctgtcatct atattataca ttatctcaa tataacctct caattgcctg
 42960
 tagaaataat acccagcaca ttttacagtt tggacatga tatccctatt ttacattaca
 43020
 ccctcacagc actcttgta attaatgtgc tgtctttgta tacagggtca gattgttaag
 43080
 cgacttgccc atagtcaact agtaagcaag ttcagttctc cttttctcta cagccatttc
 43140
 acagtaagaa ttacttaatt atgtagtttg actttcaggt acagtgagga agaatttact
 43200
 gtttttggtt tgcctctctc cttataggat gatgtgggct gctggggcag tagcagccat
 43260
 gtctagcatc acctttctct ctgtcagtc acttgtttca cgaactgctg atgctgatca
 43320
 acagggtgag ttgataggaa ctacgataa ttatttaaaa gtacagaatg ttctaactct
 43380
 gtgttctgtc tctatgtac tgaacataa gtatatctc agggtagaga cttttaaat
 43440
 tgcttttgat ataaacagga aaagcagatt ctagggtatt tatccttagg tagatacata
 43500
 ttcccttttc tctcacttag aatatgtggtc ttatctgttc tgttcataga aaatttactg
 43560
 atgaggctgg gcattgtggtc tcacacctgt aatcccaaca ctttgagagg ccgaggcagg
 43620
 cagattgctt gagttcagga gttggagacc agcctgggca acatggggaa acccatcac
 43680
 taaaatgcaa aaattagctg ggcagggttg cgcagcttg tagtcccagt taactgggag
 43740
 gctgagggtg gaggatcact tgagttctag aggcggaggt tgcactgagc caaatcag
 43800

WO 02/48366

PCT/US01/47559

ccagtgaact ccaecctggg cgacacagtg agagaccttg agagacctgt ctcaaaaaaa
 43860
 ttactgatg aatgttggcc acagaatatt aactaagcat taagtttttg ctgtgttggtg
 43920
 ctagecacct tgtgggatat attcatagac cttttatgaa tgttgtcca gccacttggg
 43980
 aagctgaggg aggaggatta cttgagctca agagtgtgaa gctagcttag goaacacagc
 44040
 aagactccca tctctaataa aaaaaaaaaa agaaatcata tagtttcagt catctgaaga
 44100
 tatgtaaacac agcaacatct ttaatgtcat atgototctt tttttttttt ggtgggaaca
 44160
 ttaattctctt gggatgctaa cagatgacaa atacttcttt gaaaaggata tgttttgtct
 44220
 agtcataagg ataaaagggg cactgcaaaa cagtggcagt tgcacttgg ttgataaatg
 44280
 aggaaggaaa ggcctatggc caaagcaagt tgcattttta attaaagtc caaagagag
 44340
 aaacaaaaat caaatgctgt taaacagtg ttattggccg ggcgcagtg ctcatgcctg
 44400
 taattctcagt gctttgggg gccaaaggct gtggtacacc taaggtcagg agttcgagat
 44460
 cagcctggcc aacatgggta aacctgtct ctaactaaaa aatcaaccag gcgtcgtggt
 44520
 gcacacctgt aatccagct actcgggagg ctgagggcagg agaatcactt gaaccttgga
 44580
 agcggagttt gcagtggcc gagatcgtgc cactgcactc cagcctgggc aacaagagtg
 44640
 aaactccgtc tcaaaaaaa aagtgtgatt ttggctgggc acagtggctc aatgctgta
 44700
 atcacagcac ttgttgagg tgaggcagga ggaatccttg agttgagtc aacaccacc
 44760
 taggcaacaa agtgagacc catctctaca aaaagtacaa aaattagcca ggtatttgag
 44820
 cccaggaagt tgaggctgca gtgagccaa tttgcgccac tgcactccg ccagttgacc
 44880
 ccccaacttt atttttgaat aaaaggggtc ttaagtactt gcttggggac agataaatt
 44940
 taacatttgt agttgtgaaa ttgtgactga tttttaacca gccatgtttt gaaggctgta
 45000
 atctagggaa ataaagtagt tttgccact tgtttctatg tgagacctat atatgggtac
 45060
 atcagagctc atgtttgcat aggacagctt actacacttt gtagagctga tagcttctaa
 45120
 atattaatag tttttaatga cattgctata aattgctagt gtctgataag aagcttgatt
 45180
 ttagattang tgatttcata ttgtactgg ttctaaggta gggaaaaaaa accaggtagt
 45240
 ttactaagt ataatgtgtt taaacaatgt aggtgtggt caagggaatga taacaggaaat
 45300
 tccaggatta tcaaatggtc tgggaccggc cctctatgga ttcatittct acatattcca
 45360
 tctggaaactt aaagaactgc caataacagg aacagacttg ggaacaaca caagccctca
 45420
 gcaccacttt gaacaggtaa ttctcattca acatgatcaa attgtatggt tcaattggct
 45480
 agattaaagg ctactgggtt ttggtcatga aagcttttat gtgagattct attgagatt
 45540
 ggataaaaatg cttaaaaaac agagttttag ggaactgtgt cttctacata ccaccttggt
 45600
 aaatttggt gtgcataatt tttttcoact tgagaatgaa aaattttaag tacctagttt
 45660
 gtccatggca tatgtaacaa accatttctc tttaactacta gtctcttttg aaacttttct
 45720
 aatatcacag ttgtgtatgt taactaactt ttcatacaaa aggcaggctt atggtaaata
 45780
 acctttgttc actgtttgtc atatttcctt cttttcaact gaaataaat gccaaacaat
 45840

WO 02/48366

PCT/US01/47559

gctgatcatt ttggccaata ttagcagctc atcagtttcc tggcttattt agcagagttc
 45900
 tgtacttata ttccaaccca ctaagactac ttttaataaa aagaatgttt gtgggtata
 45960
 cacaaaactg gcagtgcagg ctctgtctaa gaatgaatgt aattttgggc taaaacaaag
 46020
 gtctcttttt ctggtaacc tgtgtttttc tcaactccag tcaactgaaa ttataattac
 46080
 ctctcttgaaa caaagaaatg gtaatttatt ttgtctactgt agtaataact gtagctacc
 46140
 tgattttatt ttgttttttg ttgtttctga gacagtctta ctctgtcaca caggctggaa
 46200
 tgcagtggca caatcttggc tctctgcaac ctctatctcc caagttcagc cgattctttt
 46260
 aacatcagcc tcccggttag ctgggattac aggcataatgc caccatgacc ggctaatttt
 46320
 ttttgtattt ttagtagaga cgaggtttca ccatgttggc caggctggtc tcgaactcct
 46380
 gacctcaagt gatctgcctg cctcagcctc ccaaagtgtc gggattacag gcttgagcca
 46440
 acacgtgcag cctgatttta tttttgagat cttttataaa acgttttccc cttggactaa
 46500
 caaatlaatc atagaaatgc tgtgttcaca ttttgtctg aaagtaatga tgtaattatt
 46560
 tcgcataggc tgttttgcca gtgtttatcc ccagaatttt atagagaatg tgatagtatc
 46620
 tcttttctcc taaaaaggga atgcctttta taattatggc ttttaaaag agaagagcaa
 46680
 ttagcttgta attcataagt taatactaaa aggtatgtag tttctctta tgagtaaaag
 46740
 tattttgtga aaaaacttaa ttactttatt ctctctcaga attccatcat ccoctggcct
 46800
 cccttccat: ttggagcctg ttcagtactg ctggctctgc ttgttgcctt gtttattccg
 46860
 gaacatacca atttaagcct aaggtccagc agttggagaa agcactgtgg cagtcacagc
 46920
 catcttcata atacacaagc gccaggagag gccaaagAAC ctttactcca ggacacaaat
 46980
 gtgtgacgac tgaatcagg aagatttttc tatcagcacc caggtcttag ttttcacctc
 47040
 tagttctgga tgtacattcc atttccatcc acagtgtact ttaagattgt cttagaagaa
 47100
 gtatctgcat gaactccgtg ggaactaaag gaagtgggaa cttagaacca gacagttttc
 47160
 caaagatgtt acaatttctt ttgaaaaccc ttttgtttat tagccaat ttcttgccac
 47220
 taagctattt gttttattat acatctctta attaaaaact atatatgtaa ctcttagat
 47280
 attagcaaat gtctctgcta ccatcttctt aaggtgttga gcttttaact tatgctgact
 47340
 cagtgcagaca cagttaggtag tatggttttg gacctattt ttttaacatt gtaaaatttt
 47400
 gagtgcagatt ttaatatgtt aaatcttgg gtcataaat tcaaagcctt aatgcagatg
 47460
 cactaaaacs aagaatggg aaatgaattg ttgcattha aaaaaaana ctcttaagaa
 47520
 aactgtacta aatctgaatc atgtttttgag ctgttttgca gtacttttaa acattattca
 47580
 ctactgtttt tgaagtgaag aagtatcagc catttagcat ttaagttggg gtatttagag
 47640
 cctgtaattc aaatgctggc tcaaatattt tccccagcta ctctttatc cactattctt
 47700
 ttaattgttg cataatcata agcaccctaa cacttgaata cstaacttaa aaattatata
 47760
 gtaagctgg tagccttga aatgtcagtg tgatatctat tatgtagata aatatatata
 47820
 gtggcctttc aggactgtca cagtaacact ttatttacag agctaattgt tgcctaaat
 47880

WO 02/48366

PCT/US01/47559

```

tttcaggacc cttagaggaga gctttatata attaaccgatg tgaatttttc taaagtgtat
47940
atTTTTgtgt ccagttatat tatttaaaa agtggtactt tgtaaaaat gtatataaag
48000
aactgtatag ttacactgtt ttcatcttg tgggtggta ttgctaatg ctttttaac
48060
ttggaacact cactatggtt aaataaggtc ttaaaagaaa tgtaaatatt ctgttaataa
48120
agttaaatat tttaagtatt ttttttttaa aaaagtcttg atctgtgagt tcctacaggg
48180
tctgttttgg ctttgttacc atatttttac agtaacaaat caaagcatta ttcattcttt
48240
ctttgcatcc atttcttttc cctcactaaa aaatgctctt caaggaagac aagaggaaag
48300
aaacaaaaat gcagtatgag ttttgaagta cttagtttc aaagtgtaga tgctgcataa
48360
tagtttatta acaaaagctca aagtttacag aaataacatt aaatgcacaa ctitttgatt
48420
attaacacat gtacaatttt acactgcaa acaattgcaa ctaaaagttt aggaggtaaa
48480
gaattagcac acttggaagt ctgtatacat ttaatacaaa ttgcaagat atcagaatga
48540
tctctcctga aaatccaact ttcttctgtt atatatatgc aagagatcac ctatatttta
48600
aaaaagagag ctacctgtat gtcatgcat ccatccaaaa catgtcacia tattttatat
48660
atattacata gtatttacaa caaagccctt tccatagtat ttacaataaa gtcctctcca
48720
aatcaccaa gaggcctctt tccagaatag caagtgcctt caagttactg taccaagtat
48780
tctcactaat acagtatgtt atctaattgg gggaggtggg ggagtatcaa ctgcctaata
48840
ttagtactcg aattttcaga ttagatccct agtttaaga cacttgcaat tgccaatagt
48900
ttcaagatgc tggcttcgta aaaaaaaaa aaaaatiagc aaagattctt tcttaccatg
48960
ggactcagta atgttattac atcaatatct tgaatgttt ctcatctgt gtaataaagt
49020
agttgaaaat aattttaaat tatcacatga gcattagtag ttataaaaa ttgattatga
49080
agactcttta gagaatgcta ccaggtattg ttgtcaata agaaaaacaa atgaggccct
49140
atgattctag gtgaatttta aaaaaatttt tcccaagga acaattctta aaactttcag
49200
ttacagggaa gagaagaaaa ttgtaccttc taacagttat ttgttttgc tagtttatta
49260
acattgaaat caaccaagt tctgatacta taaaaataa tgaattatca tgatatagta
49320
gaatttagtag tagtaccag ttttaaaagt ccactgatac agatttaact gcataataa
49380
actacaatt gagaaaaact agatgagtat aaacatttag gaagcactga agttttaaa
49440
acttgtcagc tcaaatgaca aaagcactta atctggtcac ataaaaactg tccaaattat
49500
tttaagtata tcaaatttat ttgattcact actagcaat ttaaatgctt caaggaaat
49560
acgctatgaa ggcttaattc atttcagtta tttaaggtaa atttaggact ttcccttta
49620
aatgtcatca ataatatatc tcaagaggta atcattttcc ctcatgattt tcaggtgttg
49680
aatttanaag ttttaccat gaaagtgtat aaatgaatta gtattctttg gttgcatact
49740
atgaggttat gagcaggttt tagtttaccg agatttttaa tatgatcaga atctccctca
49800
taaggatcaa ctctttcctt agaactgaat ttctaagaa tagcttaata aaattcatat
49860
tatctataa aacgccatgt tcacaaacca aaaaatgtca tttactcaca atctttacca
49920

```

WO 02/48366

PCT/US01/47559

atcccccaagt acaaatitgt tccatattta taaactgccca taaaagcagt acttaagta
 49980
 tcca
 49984

<210> 6
 <211> 490
 <212> FRT
 <213> Mus musculus

<400> 6
 Met Thr Gln Gly Lys Lys Lys Lys Arg Ala Ala Asn Arg Ser Ile Met
 1 5 10 15
 Leu Ala Lys Lys Ile Ile Ile Lys Asp Gly Gly Thr Pro Gln Gly Ile
 20 25 30
 Gly Ser Pro Ser Val Tyr His Ala Val Ile Val Ile Phe Leu Glu Phe
 35 40 45
 Phe Ala Trp Gly Leu Leu Thr Ala Pro Thr Leu Val Val Leu His Glu
 50 55 60
 Thr Phe Pro Lys His Thr Phe Leu Met Asn Gly Leu Ile Gln Gly Val
 65 70 75 80
 Lys Gly Leu Leu Ser Phe Leu Ser Ala Pro Leu Ile Gly Ala Leu Ser
 85 90 95
 Asp Val Trp Gly Arg Lys Ser Phe Leu Leu Leu Thr Val Phe Phe Thr
 100 105 110
 Cys Ala Pro Ile Pro Leu Met Lys Ile Ser Pro Trp Trp Tyr Phe Ala
 115 120 125
 Val Ile Ser Val Ser Gly Val Phe Ala Val Thr Phe Ser Val Val Phe
 130 135 140
 Ala Tyr Val Ala Asp Ile Thr Gln Glu His Glu Arg Ser Met Ala Tyr
 145 150 155 160
 Gly Leu Val Ser Ala Thr Phe Ala Ala Ser Leu Val Thr Ser Pro Ala
 165 170 175
 Ile Gly Ala Tyr Leu Gly Gln Met Tyr Gly Asp Ser Leu Val Val Val
 180 185 190
 Leu Ala Thr Ala Ile Ala Leu Leu Asp Ile Cys Phe Ile Leu Val Ala
 195 200 205
 Val Pro Glu Ser Leu Pro Glu Lys Met Arg Pro Ala Ser Trp Gly Ala
 210 215 220
 Pro Ile Ser Trp Glu Gln Ala Asp Pro Phe Ala Ser Leu Lys Lys Val
 225 230 235 240
 Gly Gln Asp Ser Ile Val Leu Leu Ile Cys Ile Thr Val Phe Leu Ser
 245 250 255
 Tyr Leu Pro Glu Ala Gly Gln Tyr Ser Ser Phe Phe Leu Tyr Leu Lys
 260 265 270
 Gln Ile Met Lys Phe Ser Pro Glu Ser Val Ala Ala Phe Ile Ala Val
 275 280 285
 Leu Gly Ile Leu Ser Ile Ile Ala Gln Thr Ile Val Leu Ser Leu Leu
 290 295 300
 Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu Leu Gly Leu Gly Phe
 305 310 315 320
 Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe Gly Ser Glu Pro Trp Met
 325 330 335
 Met Trp Ala Ala Gly Ala Val Ala Ala Met Ser Ser Ile Thr Phe Pro
 340 345 350
 Ala Val Ser Ala Leu Val Ser Arg Thr Ala Asp Ala Asp Gln Gln Gly
 355 360 365
 Val Val Gln Gly Met Ile Thr Gly Ile Arg Gly Leu Cys Asn Gly Leu
 370 375 380
 Gly Pro Ala Leu Tyr Gly Phe Ile Phe Tyr Ile Phe His Val Glu Leu
 385 390 395 400
 Lys Glu Leu Pro Ile Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro
 405 410 415
 Gln His His Phe Glu Gln Asn Ser Ile Ile Pro Gly Pro Pro Phe Leu
 420 425 430
 Phe Gly Ala Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile
 435 440 445
 Pro Glu His Thr Asn Leu Ser Leu Arg Ser Ser Ser Trp Arg Lys His

WO 02/48366

PCT/US01/47559

450 455 460
 Cys Gly Ser His Ser His Pro His Ser Thr Gln Ala Pro Gly Gln Ala
 465 470 475 480
 Lys Glu Pro Leu Leu Gln Asp Thr Asn Val
 485 490

<210> 7
 <211> 490
 <212> PRT
 <213> Mus musculus

<400> 7
 Met Thr Gln Gly Lys Lys Lys Lys Arg Ala Ala Asn Arg Ser Ile Met
 1 5 10 15
 Leu Ala Lys Lys Ile Ile Ile Lys Asp Gly Gly Thr Pro Gln Gly Ile
 20 25 30
 Gly Ser Pro Ser Val Tyr His Ala Val Ile Val Ile Phe Leu Glu Phe
 35 40 45
 Phe Ala Trp Gly Leu Leu Thr Ala Pro Thr Leu Val Val Leu His Glu
 50 55 60
 Thr Phe Pro Lys His Thr Phe Leu Met Asn Gly Leu Ile Gln Gly Val
 65 70 75 80
 Lys Gly Leu Leu Ser Phe Leu Ser Ala Pro Leu Ile Gly Ala Leu Ser
 85 90 95
 Asp Val Trp Gly Arg Lys Ser Phe Leu Leu Leu Thr Val Phe Phe Thr
 100 105 110
 Cys Ala Pro Ile Pro Leu Met Lys Ile Ser Pro Trp Trp Tyr Phe Ala
 115 120 125
 Val Ile Ser Val Ser Gly Val Phe Ala Val Thr Phe Ser Val Val Phe
 130 135 140
 Ala Tyr Val Ala Asp Ile Thr Gln Glu His Glu Arg Ser Met Ala Tyr
 145 150 155 160
 Gly Leu Val Ser Ala Thr Phe Ala Ala Ser Leu Val Thr Ser Pro Ala
 165 170 175
 Ile Gly Ala Tyr Leu Gly Gln Met Tyr Gly Asp Ser Leu Val Val Val
 180 185 190
 Leu Ala Thr Ala Ile Ala Leu Leu Asp Ile Cys Phe Ile Leu Val Ala
 195 200 205
 Val Pro Gln Ser Leu Pro Glu Lys Met Arg Pro Ala Ser Trp Gly Ala
 210 215 220
 Pro Ile Ser Trp Glu Gln Ala Asp Pro Phe Ala Ser Leu Lys Lys Val
 225 230 235 240
 Gly Gln Asp Ser Ile Val Leu Leu Ile Cys Ile Thr Val Phe Leu Ser
 245 250 255
 Tyr Leu Pro Glu Ala Gly Gln Tyr Ser Ser Phe Phe Leu Tyr Leu Lys
 260 265 270
 Gln Ile Met Lys Phe Ser Pro Glu Ser Val Ala Ala Phe Ile Ala Val
 275 280 285
 Leu Gly Ile Leu Ser Ile Ile Ala Gln Thr Ile Val Leu Ser Leu Leu
 290 295 300
 Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu Leu Gly Leu Gly Phe
 305 310 315 320
 Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe Gly Ser Gln Pro Trp Met
 325 330 335
 Met Trp Ala Ala Gly Ala Val Ala Ala Met Ser Ser Ile Thr Phe Pro
 340 345 350
 Ala Val Ser Ala Leu Val Ser Arg Thr Ala Asp Ala Asp Gln Gln Gly
 355 360 365
 Val Val Gln Gly Met Ile Thr Gly Ile Arg Gly Leu Cys Asn Gly Leu
 370 375 380
 Gly Pro Ala Leu Tyr Gly Phe Ile Phe Tyr Ile Phe His Val Gln Leu
 385 390 395 400
 Lys Glu Leu Pro Ile Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro
 405 410 415
 Gln His His Phe Glu Gln Asn Ser Ile Ile Pro Gly Pro Pro Phe Leu
 420 425 430
 Phe Gly Ala Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile

WO 02/48366

PCT/US01/47559

```

      435      440      445
Pro Glu His Thr Asn Leu Ser Leu Arg Ser Ser Ser Trp Arg Lys His
450      455      460
Cys Gly Ser His Ser His Pro His Ser Thr Gln Ala Pro Gly Glu Ala
465      470      475      480
Lys Glu Pro Leu Leu Gln Asp Thr Asn Val
      485      490
```

【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 02/048366 A3

- (51) International Patent Classification: C12N 15/12, 15/11, 15/63, 1/21, 5/10, C07K 14/47, 14/705, 16/18, 16/28, C12Q 1/68, G01N 33/68, A01K 67/027
- (21) International Application Number: PCT/US01/47559
- (22) International Filing Date: 11 December 2001 (11.12.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/254,553 12 December 2000 (12.12.2000) US
09/739,457 19 December 2000 (19.12.2000) US
- (71) Applicant: PE CORPORATION (NV) [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).
- (72) Inventors: MERKULOV, Gennady; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). GUEGLER, Karl; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). WEY, Ming-Hui; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). DI FRANCESCO, Valentina; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GR, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, HU, IT, LU, MC, NL, PT, SE, SI, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report
- (88) Date of publication of the international search report: 15 May 2003
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/048366 A3

(54) Title: ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the transporter peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the transporter peptides, and methods of identifying modulators of the transporter peptides.

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No PCT/US 01/47559
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/12 C12N15/11 C12N15/63 C12N1/21 C12N5/10 C07K14/47 C07K14/705 C07K16/18 C07K16/28 C12Q1/68 G01N33/68 A01K67/027		
According to International Patent Classification (IPC) or to both regional classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K C12Q G01N A01K		
Documentation searched other than minimum documentation to 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)		
EMBL, SEQUENCE SEARCH, EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE, CHEM ABS Dat		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MATSUO N ET AL: "CLONING OF A CDNA ENCODING A NOVEL SUGAR TRANSPORTER EXPRESSED IN THE NEONATAL MOUSE HIPPOCAMPUS" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US, vol. 238, no. 1, 1997, pages 126-129, XP002933745 ISSN: 0006-291X the whole document ---	1-23
X	WO 98 39448 A (HUMAN GENOME SCIENCES INC) 11 September 1998 (1998-09-11) SEQ ID NO: 152; SEQ ID NO: 454 page 88, line 17 -page 89, line 4; claims 1-23 page 153 --- -/-	1-23
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members 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 published on or after the international filing date "L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "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 "T" 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 "Z" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
22 January 2003		21 02 2003
Name and mailing address of the ISA European Patent Office, P.B. 6818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2940, Tx. 31 051 epo nl, Fax: (+31-70) 340-3016		Authorized officer Devijver, K

Form PCT/ISA210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

 Internat. Application No.
 PCT/US 01/47559

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 37674 A (GENETICS INST) 29 July 1999 (1999-07-29) SEQ ID NO: 13; SEQ ID NO: 14 page 41, line 12 -page 42, line 16; claims 1-41 page 53 -page 82	1-23
X	--- DATABASE EMBL [Online] 14 October 2000 (2000-10-14) NIH-MGC: "501458251F1 NIH MGC 66 Homo sapiens cDNA clone IMAGE:3861876 5', mRNA sequence." Database accession no. BF036183 XP002228193	1,2,4,5, 8,9
X	--- DATABASE EMBL [Online] 3 April 2000 (2000-04-03) BIRREN B ET AL: "Homo sapiens chromosome 1 clone RP11-84617 map 1, WORKING DRAFT SEQUENCE, 19 unordered pieces." Database accession no. AC027557 XP002228194	1,2,4,5, 8,9
E	WO 02 22684 A (YUE HENRY;INCYTE GENOMICS INC ; LEE ERNESTINE A (US)) 21 March 2002 (2002-03-21) SEQ ID NO: 5; SEQ ID NO: 31 claims 1-107; tables 1-3	1-23
P,X	--- DATABASE EMBL [Online] 30 October 2001 (2001-10-30) MATHUR B N ET AL: "Homo sapiens putative tetracycline transporter-like protein mRNA, complete cds." Database accession no. AF427492; Q96NY0 XP002228195	1-23
L	--- DATABASE EMBL [Online] 14 December 2001 (2001-12-14) PADIGARU M ET AL: "Sequence 28 from Patent WO0188135." Database accession no. AX317853 XP002228196 L document cited to provide information on the relevant sequence disclosed in WO0188135	1-23
P,X	--- -& WO 01 88135 A (BAUMGARTNER JASON C ;BURGESS CATHERINE E (US); MACDOUGALL JOHN R) 22 November 2001 (2001-11-22) claims 1-52 --- -/--	1-23

Form PCT/ISA/210 (continuation of second sheet) (July 1999)

INTERNATIONAL SEARCH REPORT		Internat. application No. PCT/US 01/47559
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] 14 December 2001 (2001-12-14) PADIGARU M ET AL: "Sequence 27 from Patent WO0188135." Database accession no. AX317852 XP002228197 L document cited to provide information on the relevant sequence disclosed in WO0188135	1-23
P,X	-& WO 01 88135 A (BAUMGARTNER JASON C ;BURGESS CATHERINE E (US); MACDOUGALL JOHN R () 22 November 2001 (2001-11-22) claims 1-52	1-23
P,X	WO 01 85767 A (MACDOUGALL JOHN R ;SPADERNA STEVEN K (US); SPYTEK KIMBERLY A (US);) 15 November 2001 (2001-11-15) SEQ ID NO: 27; SEQ ID NO: 28 page 59, line 1 -page 63, line 9; claims 1-41	1-23
P,X	WO 00 77021 A (HUMAN GENOME SCIENCES INC ;ROSEN CRAIG A (US); RUBEN STEVEN M (US)) 21 December 2000 (2000-12-21) SEQ ID NO: 17; SEQ ID NO: 113 page 21, line 19 -page 25, line 2; claims 1-23	1-23
P,X	DATABASE EMBL [Online] 10 August 2001 (2001-08-10) KAUL R K ET AL: "Homo sapiens chromosome 1 clone RP4-714D9, complete sequence." Database accession no. AC093019; AL356262 XP002228198	1,2,4,5,8,9
A	WO 00 00633 A (MILLENNIUM PHARM INC) 6 January 2000 (2000-01-06) the whole document	1-23

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 01/47559
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.	<input checked="" type="checkbox"/>	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: Although claim 18 (in part) 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.
2.	<input checked="" type="checkbox"/>	Claims Nos.: in part: 12,17,18 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: see FURTHER INFORMATION sheet PCT/ISA/210
3.	<input type="checkbox"/>	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 8.4(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:		
1.	<input type="checkbox"/>	As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2.	<input type="checkbox"/>	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	<input type="checkbox"/>	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.	<input type="checkbox"/>	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.:
Remark on Protest		
	<input type="checkbox"/>	The additional search fees were accompanied by the applicant's protest.
	<input type="checkbox"/>	No protest accompanied the payment of additional search fees.

International Application No. PCT/US 01/47559

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: in part: 12,17,18

Claim 17 refers to an agent identified by the method of claim 16 without giving a true technical characterization. The claim covers all compounds having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for only a very limited number of such compounds, namely: antibodies that selectively bind to a peptide of claim 2. In consequence, the scope of said claim is ambiguous and vague, and its subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). An attempt is made to define the compound by reference to a result to be achieved. This lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claim which appear to be clear, supported and disclosed, namely those parts relating to: antibodies that selectively bind to a peptide of claim 2.

The above comment also applies to claims 12 and 18.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 01/47559

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
WO 9839448	A	11-09-1998	AU	6545398 A		22-09-1998
			EP	0972029 A1		19-01-2000
			EP	0972030 A2		19-01-2000
			JP	2002519990 T		02-07-2002
			JP	2002510192 T		02-04-2002
			WO	9839446 A2		11-09-1998
			WO	9839448 A2		11-09-1998
			US	6420526 B1		16-07-2002
			US	2002164669 A1		07-11-2002
			US	6444440 B1		03-09-2002
			AU	6545298 A		18-09-1998
WO 9937674	A	29-07-1999	AU	2456499 A		09-08-1999
			CA	2318303 A1		29-07-1999
			EP	1049714 A1		08-11-2000
			JP	2002508936 T		26-03-2002
			WO	9937674 A1		29-07-1999
			US	2001039335 A1		08-11-2001
WO 0222684	A	21-03-2002	AU	9101101 A		26-03-2002
			WO	0222684 A2		21-03-2002
WO 0188135	A	22-11-2001	AU	6806001 A		26-11-2001
			WO	0188135 A2		22-11-2001
			AU	6805901 A		26-11-2001
WO 0185767	A	15-11-2001	AU	5952601 A		20-11-2001
			WO	0185767 A2		15-11-2001
			AU	5952601 A		20-11-2001
WO 0077021	A	21-12-2000	AU	4073600 A		14-11-2000
			AU	5457600 A		02-01-2001
			EP	1189917 A1		27-03-2002
			WO	0061774 A2		19-10-2000
			WO	0077021 A1		21-12-2000
WO 0000633	A	06-01-2000	AU	4729499 A		17-01-2000
			EP	1092038 A1		18-04-2001
			WO	0000633 A1		06-01-2000

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	C 0 7 K 14/47	4 H 0 4 5
C 0 7 K 16/18	C 0 7 K 16/18	
C 1 2 N 1/19	C 1 2 N 1/19	
C 1 2 N 1/21	C 1 2 N 1/21	
C 1 2 N 5/10	C 1 2 P 21/02	C
C 1 2 P 21/02	C 1 2 Q 1/02	
C 1 2 Q 1/02	C 1 2 Q 1/68	A
C 1 2 Q 1/68	G 0 1 N 33/15	Z
G 0 1 N 33/15	G 0 1 N 33/50	Z
G 0 1 N 33/50	G 0 1 N 33/53	D
G 0 1 N 33/53	G 0 1 N 33/53	M
G 0 1 N 33/566	G 0 1 N 33/566	
	C 1 2 N 15/00	F
	C 1 2 N 5/00	A

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,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,OM,PH,P L,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZM,ZW

(72)発明者 ゲグラー カール

アメリカ合衆国 メリーランド州 ロックビル シー 2 - 4 # 2 1 ウェスト グーデ ドライブ
4 5 セレラ ジェノミクス内

(72)発明者 ウェイ ミンーホイ

アメリカ合衆国 メリーランド州 ロックビル シー 2 - 4 # 2 1 ウェスト グーデ ドライブ
4 5 セレラ ジェノミクス内

(72)発明者 ディフランチェスコ バレンティーナ

アメリカ合衆国 メリーランド州 ロックビル シー 2 - 4 # 2 1 ウェスト グーデ ドライブ
4 5 セレラ ジェノミクス内

(72)発明者 ビースリー エレン エム.

アメリカ合衆国 メリーランド州 ロックビル シー 2 - 4 # 2 1 ウェスト グーデ ドライブ
4 5 セレラ ジェノミクス内

F ターム(参考) 4B024 AA01 AA11 BA80 CA01 CA04 CA09 CA11 DA01 DA02 DA05
DA11 EA01 EA02 EA03 EA04 FA01 GA11 HA01 HA03 HA11
4B063 QA01 QA18 QQ05 QQ13 QQ42 QQ52 QR08 QR33 QR42 QR55
QR59 QR62 QR74 QR80 QR82 QS05 QS25 QS34 QS36 QX02
4B064 AG01 AG27 CA01 CA19 CC24 DA01 DA13
4B065 AA01X AA57X AA87X AA93Y AB01 BA01 CA24 CA25 CA44 CA46
4C084 AA17 NA14 ZB21 ZB26
4H045 AA10 AA11 AA20 AA30 BA10 CA40 DA00 DA76 EA20 EA50
FA72 FA74