

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-500812

(P2004-500812A)

(43) 公表日 平成16年1月15日(2004.1.15)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A
A 6 1 K 38/48	A 6 1 K 45/00	2 G O 4 5
A 6 1 K 45/00	A 6 1 P 1/00	4 B O 2 4
A 6 1 P 1/00	A 6 1 P 1/04	4 B O 5 0
A 6 1 P 1/04	A 6 1 P 1/16	4 B O 6 3
	審査請求 未請求 予備審査請求 有	4 B O 6 4
		(全 229 頁) 最終頁に続く

(21) 出願番号	特願2001-546939 (P2001-546939)	(71) 出願人	301005050
(86) (22) 出願日	平成12年12月19日 (2000.12.19)		インサイト・ゲノミックス・インコーポレイテッド
(85) 翻訳文提出日	平成14年6月24日 (2002.6.24)		アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160
(86) 国際出願番号	PCT/US2000/034811	(74) 代理人	100089266
(87) 国際公開番号	W02001/046443		弁理士 大島 陽一
(87) 国際公開日	平成13年6月28日 (2001.6.28)	(72) 発明者	ヤング、ジュンミン
(31) 優先権主張番号	60/172,055		アメリカ合衆国カリフォルニア州95129・サンノゼ・パークレーン 7125
(32) 優先日	平成11年12月23日 (1999.12.23)	(72) 発明者	ボーグン、マライア・アール
(33) 優先権主張国	米国 (US)		アメリカ合衆国カリフォルニア州94577・サンレアンドロ・サンティアゴロード 14244
(31) 優先権主張番号	60/177,334		
(32) 優先日	平成12年1月21日 (2000.1.21)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/178,884		
(32) 優先日	平成12年1月28日 (2000.1.28)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 プロテアーゼ

(57) 【要約】

本発明は、ヒトプロテアーゼ (P R T S)、及び P R T S を識別しコードするポリヌクレオチドに関する。本発明はまた、発現ベクター及び宿主細胞、抗体、アゴニスト、アンタゴニストを提供する。更に、本発明は、P R T S の異常発現に関連する疾患の診断・治療・予防方法を提供する。

【特許請求の範囲】

【請求項 1】

単離されたポリペプチドであって、

(a) SEQ ID NO: 1 乃至 SEQ ID NO: 14 (SEQ ID NO: 1 - 14) からなる群から選択されたアミノ酸配列と、

(b) SEQ ID NO: 1 - 14 からなる群から選択されたアミノ酸配列と少なくとも 90% の配列同一性を有する天然のアミノ酸配列と、

(c) SEQ ID NO: 1 - 14 からなる群から選択されたアミノ酸配列の生物学的に活性な断片と、

(d) SEQ ID NO: 1 - 14 からなる群から選択されたアミノ酸配列の免疫原性断片とで構成される群から選択されたアミノ酸配列を含むことを特徴とする単離されたポリペプチド。 10

【請求項 2】

SEQ ID NO: 1 - 14 からなる群から選択された請求項 1 の単離されたポリペプチド。

【請求項 3】

請求項 1 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 4】

請求項 2 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 5】

SEQ ID NO: 40 - 78 からなる群から選択された請求項 4 の単離されたポリヌクレオチド。 20

【請求項 6】

請求項 3 のポリヌクレオチドに機能的に結合されたプロモーター配列を含む組換えポリヌクレオチド。

【請求項 7】

請求項 6 の組換えポリヌクレオチドで形質転換された細胞。

【請求項 8】

請求項 6 の組換えポリヌクレオチドを含む遺伝子組換え生物。

【請求項 9】

請求項 1 のポリペプチドを生産する方法であって、 30

(a) 前記ポリペプチドの発現に好適な条件下で、請求項 1 のポリペプチドをコードするポリヌクレオチドに機能的に結合されたプロモーター配列を含む組換えポリヌクレオチドで形質転換された細胞を培養するステップと、

(b) そのように発現したポリペプチドを回収するステップとを含むことを特徴とする請求項 1 のポリペプチドの生産方法。

【請求項 10】

請求項 1 のポリペプチドに特異的に結合する単離された抗体。

【請求項 11】

単離されたポリヌクレオチドであって、 40

(a) SEQ ID NO: 15 - 28 からなる群から選択されたポリヌクレオチド配列と、

(b) SEQ ID NO: 15 - 28 からなる群から選択されたポリヌクレオチド配列と少なくとも 90% の配列同一性を有する天然のポリヌクレオチド配列と、

(c) 前記 (a) に相補的なポリヌクレオチド配列と、

(d) 前記 (b) に相補的なポリヌクレオチド配列と、

(e) 前記 (a) 乃至 (d) の RNA 等価物とで構成される群から選択されたポリヌクレオチド配列を含む単離されたポリヌクレオチド。

【請求項 12】

請求項 11 のポリヌクレオチドの少なくとも 60 個の連続するヌクレオチドを含む単離さ 50

れたポリヌクレオチド。

【請求項 13】

サンプルにおいて、請求項 11 に記載のポリヌクレオチド配列を有する標的ポリヌクレオチドを検出する方法であって、

(a) 前記サンプルをプローブでハイブリダイズするステップであって、前記プローブが、前記サンプル内の前記標的ポリヌクレオチドと相補的な配列を含む少なくとも 20 個の連続するヌクレオチドを含み、前記プローブと前記標的ポリヌクレオチドまたはその断片との間でハイブリダイゼーション複合体が形成される条件下で、前記プローブが前記標的ポリヌクレオチドと特異的にハイブリダイズする、該ステップと、

(b) 前記ハイブリダイゼーション複合体が存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含むことを特徴とする標的ポリヌクレオチドの検出方法。

10

【請求項 14】

前記プローブが少なくとも 60 個の連続するヌクレオチドを含むことを特徴とする請求項 13 に記載の方法。

【請求項 15】

サンプルにおいて、請求項 11 のポリヌクレオチド配列を有する標的ポリヌクレオチドを検出する方法であって、

(a) ポリメラーゼ連鎖反応増幅を用いて、前記標的ポリヌクレオチドまたはその断片を増幅するステップと、

(b) 増幅された前記標的ポリヌクレオチドまたはその断片が存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含むことを特徴とする標的ポリヌクレオチドの検出方法。

20

【請求項 16】

有効量の請求項 1 のポリペプチド及び医薬的に容認できる賦形剤を含む組成物。

【請求項 17】

前記ポリペプチドが、SEQ ID NO: 1 - 14 からなる群から選択されたアミノ酸配列を含むことを特徴とする請求項 16 の組成物。

【請求項 18】

機能的な PRTS の発現の低下に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項 16 の組成物を投与することを含むことを特徴とする治療方法。

30

【請求項 19】

請求項 1 のポリペプチドのアゴニストとして効果的な化合物をスクリーニングする方法であって、

(a) 請求項 1 のポリペプチドを含むサンプルを化合物に曝露するステップと、

(b) 前記サンプルにおいてアゴニスト活性を検出するステップとを含むことを特徴とするスクリーニング方法。

【請求項 20】

請求項 19 のスクリーニング方法によって同定されたアゴニスト化合物及び医薬的に容認できる賦形剤を含む組成物。

40

【請求項 21】

機能的な PRTS の発現の低下に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項 20 の組成物を投与することを含むことを特徴とする治療方法。

【請求項 22】

請求項 1 のポリペプチドのアンタゴニストとして効果的な化合物をスクリーニングする方法であって、

(a) 請求項 1 のポリペプチドを含むサンプルを化合物に曝露するステップと、

(b) 前記サンプルにおいてアンタゴニスト活性を検出するステップとを含むことを特徴とするスクリーニング方法。

【請求項 23】

50

請求項 2 2 のスクリーニング方法によって同定されたアンタゴニスト化合物及び医薬的に容認できる賦形剤を含む組成物。

【請求項 2 4】

機能的な P R T S の過剰な発現に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項 2 3 の組成物を投与することを含むことを特徴とする治療方法。

【請求項 2 5】

請求項 1 のポリペプチドに特異的に結合する化合物をスクリーニングする方法であって、
(a) 請求項 1 のポリペプチドを好適な条件下で少なくとも 1 つの試験化合物と結合させるステップと、

(b) 請求項 1 のポリペプチドと前記試験化合物との結合を検出して、請求項 1 のポリペプチドと特異的に結合する化合物を同定するステップとを含むことを特徴とするスクリーニング方法。 10

【請求項 2 6】

請求項 1 のポリペプチドの活性を変化させる化合物をスクリーニングする方法であって、

(a) 請求項 1 のポリペプチドを、その活性が許容される条件下で少なくとも 1 つの試験化合物と結合させるステップと、

(b) 前記試験化合物の存在下での請求項 1 のポリペプチドの活性を評価するステップと、

(c) 前記試験化合物の存在下での請求項 1 のポリペプチドの活性と、前記試験化合物の不在下での請求項 1 のポリペプチドの活性とを比較するステップとを含み、 20

前記試験化合物の存在下での請求項 1 のポリペプチドの活性の変化が、請求項 1 のポリペプチドの活性を変化させる化合物の存在を示唆すること特徴とするスクリーニング方法。

【請求項 2 7】

請求項 5 の配列を含む標的ポリヌクレオチドの発現を変化させるのに効果的な化合物をスクリーニングする方法であって、

(a) 前記標的ポリヌクレオチドの発現に好適な条件下で、前記標的ポリヌクレオチドを含むサンプルを化合物に曝露するステップと、

(b) 前記標的ポリヌクレオチドの発現の変化を検出するステップと、

(c) 様々な量の前記化合物の存在下での前記標的ポリヌクレオチドの発現と、前記化合物の不在下での前記標的ポリヌクレオチドの発現とを比較するステップとを含むことを特徴とするスクリーニング方法。 30

【請求項 2 8】

試験化合物の毒性を評価する方法であって、

(a) 核酸を含む生体サンプルを前記試験化合物で処理するステップと、

(b) 処理した前記生体サンプルの核酸と、請求項 1 1 のポリヌクレオチドの少なくとも 2 0 の連続するヌクレオチドを含むプローブをハイブリダイズさせるステップであって、このハイブリダイゼーションが、前記プローブと前記生体サンプルの標的ポリヌクレオチドとの間で特異的なハイブリダイゼーション複合体が形成される条件下で行われ、前記標的ポリヌクレオチドが、請求項 1 1 のポリヌクレオチドのポリヌクレオチド配列またはその断片を含むポリヌクレオチドである、前記ステップと、 40

(c) ハイブリダイゼーション複合体の収量を定量するステップと、

(d) 前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量を、未処理の生体サンプルにおけるハイブリダイゼーション複合体の収量と比較するステップとを含み、

前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量の差が試験化合物の毒性を示唆することを特徴とする試験化合物の毒性評価方法。

【発明の詳細な説明】

【0 0 0 1】

本発明は、プロテアーゼの核酸配列及びアミノ酸配列に関する。本発明はまた、これらの配列を利用した胃腸疾患、心血管疾患、自己免疫/炎症性の疾患、細胞増殖異常、発生障 50

害、上皮の疾患、神経の疾患、および生殖障害の診断・治療・予防に関する。本発明はさらに、プロテアーゼの核酸配列及びアミノ酸配列の発現における外来性化合物の効果についての評価に関する。

【0002】

(発明の背景)

プロテアーゼは、タンパク質の骨格またはペプチド鎖を形成するペプチド結合部位においてタンパク質及びペプチドを切断する。蛋白分解は、細胞の内外で起こる最も重要かつ頻繁に起こる酵素反応の1つである。蛋白分解は、新生ポリペプチドの活性化及び成熟、誤って折り畳まれたタンパク質及び損傷を受けたタンパク質の分解、細胞内のペプチドの制御された代謝回転を担っている。プロテアーゼは、胚発生中の消化、内分泌機能、および組織形成や、創傷の治癒および正常な成長に關与する。プロテアーゼは、調節タンパク質の半減期を変えて調節作用において重要な役割を果たしている。プロテアーゼは、炎症、脈管形成、腫瘍の拡散および転移、心血管疾患、神経の疾患、細菌感染、ウイルス感染、および寄生虫感染などの病因または疾患状態の進行に關係する。

10

【0003】

プロテアーゼは、その基質をどの部位で切断するかによって分類することができる。アミノペプチダーゼ、ジペプチジルペプチダーゼ、トリペプチダーゼ、カルボキシペプチダーゼ、ペプチジルジペプチダーゼ、ジペプチダーゼ、およびペプチダーゼを含むエキソペプチダーゼは基質の末端における残基を切断する。セリンプロテアーゼ、システインプロテアーゼ、およびメタロプロテアーゼを含むエンドペプチダーゼは、ペプチド内の残基を切断する。哺乳動物プロテアーゼ、活性部位の構造、動作機構、および全体的な3次元構造によって主に4つの種類に分けられた(See Beynon, R. J. および J. S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)。

20

【0004】

セリンプロテアーゼ

セリンプロテアーゼ(SP)は、タンパク質分解酵素の広範で大きなファミリーであって、消化酵素であるトリプシンおよびキモトリプシン、補体および血液凝固カスケードの成分、細胞内および細胞外マトリックスの巨大分子の分解および代謝回転を調節する酵素を含む。20以上のサブファミリーのほとんどが、それぞれが共通の祖先を持つ6つのクランに分類され得る。これら6つのクランは、少なくとも4つの進化的に異なった祖先に由来すると推定される。SPは、ほとんどのファミリーの活性触媒部位にセリン残基が存在することから命名された。活性部位は、触媒作用に重要な一群の保存されたアスパラギン酸、ヒスチジン、およびセリン残基の3つの触媒三連構造によって形成される。これらの残基が基質の結合を促進する電荷リレーネットワークを形成する。活性部位の外側のその他の残基が、触媒作用中に形成される四面体形遷移中間体を安定化させるオキシアニオンホールを形成する。SPは広範な基質を有し、その基質特異性によってサブファミリーに分けることができる。主なサブファミリーは切断した後の残基から命名され、trypases (アルギニンまたはリシンの後)、aspases (アスパラギン酸の後)、chymases (フェニルアラニンまたはロイシンの後)、metases (メチオニンの後)、およびserases (セリンの後)がある(Rawlings, N. D. および A. J. Barrett (1994) Meth. Enzymol. 244: 19-61)。

30

40

【0005】

ほとんどの哺乳動物セリンプロテアーゼは、蛋白分解によって活性化される不活性な前駆体である酵素前駆体として合成される。例えば、トリプシノーゲンは、エンテロペプチダーゼによってその活性型であるトリプシンに変換される。エンテロペプチダーゼは、トリプシノーゲンからN末端断片を除去する腸プロテアーゼである。残った活性断片がトリプシンであって、その他の膵酵素の前駆体を活性化する。同様に、トロンビンの前駆体であ

50

るプロトロンビンの蛋白分解によって、3つの異なるポリペプチド断片が生成される。活性型トロンビンを含む2つの断片がジスルフィド結合によって結合している時にN末端断片が切断される。

【0006】

2つの最も大きなSPサブファミリーは、キモトリプシン(S1)ファミリーおよびサブチリシン(S8)ファミリーである。キモトリプシンファミリーのメンバーの中には、このファミリーに固有の2つの構造的ドメインを含むものもある。クリングルドメインは、変動するコピー数に見られる3重ループジスルフィド架橋ドメインである。クリングルドメインは、膜、その他のタンパク質やホスホリピッドなどの媒介物の結合、並びに蛋白分解の調節においてある役割を果たすと考えられる(PROSITE PDOC00020)。Appleドメインは90のアミノ酸からなる反復ドメインであって、それぞれが6個の保存されたシステインを含む。3つのジスルフィド結合によって、第1のシステインと第6のシステイン、第2のシステインと第5のシステイン、および第3のシステインと第4のシステインとが結合されている(PROSITE PDOC00376)。Appleドメインは、タンパク質-タンパク質相互作用に参与する。S1ファミリーメンバーには、トリプシン、キモトリプシン、凝固因子IX-XII、補体因子B、C、およびD、グランザイム、カリクレイン、および組織プラスミノゲンアクチベーターおよびウロキナーゼプラスミノゲンアクチベーターが含まれる。サブチリシンファミリーは、真性細菌、古細菌、真核生物、およびウイルスに見られるメンバーを含む。サブチリシンには、タンパク質-タンパク質エンドペプチダーゼであるkexinおよびfurin、脳下垂体プロホルモンコンバーターゼPC1、PC2、PC3、PC6、およびPACE4が含まれる(RawlingsおよびBarrett, 前出)。

10

20

【0007】

SPは多くの正常なプロセスにおいて作用するが、SPの中には疾患の病因や治療に関係するものもある。腸消化のインヒビターであるエンテロキナーゼは、腸刷子(intestinal brush)の縁に見られ、そこでトリプシノーゲンから酸性プロペプチドを切断して活性なトリプシンを生成する(Kitamoto, Y. 他(1994) Proc. Natl. Acad. Sci. USA 91:7588-7592)。アンギオテンシンIIおよびIII、および[des-Arg9]ブラジキニンなどのペプチドを切断するリソソームセリンペプチダーゼであるプロリルカルボキシペプチダーゼは、セリンカルボキシペプチダーゼおよびプロリルエンドペプチダーゼファミリーの両方のメンバーと配列相同性を共有する(Tan, F. 他(1993) J. Biol. Chem. 268:16631-16638)。プロテアーゼであるニューロプシンは、神経シグナル伝達に应答して海馬におけるシナプス形成およびニューロンの結合に影響を与え得る(Chen, Z.-L. 他(1995) J Neurosci 15:5088-5097)。組織プラスミノゲンアクチベーターは、発作(Zivin, J. A. (1999) Neurology 53:14-19)および心筋梗塞(Ross, A. M. (1999) Clin. Cardiol. 22:165-171)の救急処置に有用である。消化管の至る所で高発現されているある種の受容体(PAR:プロテアーゼ活性化受容体)は、細胞外ドメインの蛋白分解による切断によって活性化される。PAR、トロンビン、トリプシン、および肥満細胞トリプターゼの主なアゴニストは、アレルギーおよび炎症の状態で放出される。プロテアーゼによるPAR活性の調節は、有望な治療標的であると提案された(Vergnolle, N. (2000) Aliment. Pharmacol. Ther. 14:257-266; Rice, K. D. 他(1998) Curr. Pharm. Des. 4:381-396)。前立腺特異的抗原(PSA)は、前立腺の上皮細胞のみで合成および分泌されるカリクレイン様セリンプロテアーゼである。血清PSAは前立腺癌においてそのレベルが上昇し、癌の進行および治療による効果をモニタリングするための最も感度の良い生理学的マーカーである。またPSAによって、転移性癌の原発癌として前立腺を同定することができる(Brawer, M. K. 他(1989) Urology 33

30

40

50

: 11 - 16)。

【0008】

シグナルペプチダーゼは、全ての真核細胞型および原核細胞型に見られるSPの或るクラスであって、ある種のタンパク質からのシグナルペプチドのプロセシングに作用する。シグナルペプチドはタンパク質のアミノ末端ドメインであって、リボソーム組み立て部位からそのタンパク質を所定の細胞内或いは細胞外の部位に導く。一旦タンパク質が輸送されると、シグナルペプチダーゼによってそのシグナル配列が除去され、グリコシル化またはリン酸化などの翻訳後プロセシングによってそのタンパク質が活性化される。シグナルペプチダーゼは、酵母および哺乳動物の両方において多サブユニット複合体として存在する。イヌシグナルペプチダーゼ複合体は、5つのサブユニットから成り、その全てのサブユニットがミクロソーム膜と結合し、膜を1回或いは複数回貫通する疎水性領域を含む (Shelness, G. S. および G. Blobel (1990) J. Biol. Chem. 265: 9512 - 9519)。これらのサブユニットのあるものは、その複合体を膜上の適切な位置に移動させ、別のものは実際の触媒活性を含む。

【0009】

活性部位にセリンを含むプロテアーゼの別のファミリーは、その活性がATPの加水分解に依存する。これらのプロテアーゼは、ATP/GTP-結合モチーフであるPループの存在によって同定することができる調節性ATPアーゼドメインおよび蛋白分解コアドメインを含む (PROSITE PDOC00803)。このファミリーのメンバーには、真核生物ミトコンドリアマトリックスプロテアーゼ、C1pプロテアーゼ、およびプロテアソームが含まれる。C1pプロテアーゼは、初めは植物の葉緑体で見出されたが、原核細胞および真核細胞の両方において広範に存在すると考えられる。早発性捻転ジストニーの遺伝子は、C1pプロテアーゼに関連するタンパク質をコードする (Ozelius, L. J. 他 (1998) Adv. Neurol. 78: 93 - 105)。

【0010】

プロテアソームは、ある種の細菌および全ての真核細胞に見られる細胞内プロテアーゼ複合体であって、細胞内の生理機能において重要な役割を果たす。プロテアソームは、全てのタイプの細胞内タンパク質の分解の主な経路であるユビキチン抱合系 (ubiquitin conjugation system: UCS) と関連する。その細胞内タンパク質の中には、転写や細胞周期の進行などの細胞プロセスを活性化或いは抑制するように作用するタンパク質も含まれる (Ciechanover, A. (1994) Cell 79: 13 - 21)。UCS経路では、分解標的となるタンパク質に、小さな耐熱性タンパク質であるユビキチンが結合する。次にユビキチンが結合したタンパク質が、プロテアソームによって認識され分解される。得られるユビキチン-ペプチド複合体は、ユビキチンカルボキシル末端ヒドラーゼによって加水分解され、UCSによって再利用されるために遊離ユビキチンが放出される。ユビキチンプロテアソーム系は、有糸分裂周期キナーゼ (mitotic cyclic kinases)、オンコプロテイン、腫瘍抑制遺伝子 (p53)、シグナル伝達に関係する細胞表面受容体、転写調節因子、および変異タンパク質または損傷タンパク質の分解に関与する (Ciechanover, 前出)。この経路は、嚢胞性繊維症、Angelman's症候群、およびリドル症候群を含む様々な疾患に関与する (Schwartz, A. L. および A. Ciechanover (1999) Annu. Rev. Med. 50: 57 - 74を参照)。マウスプロトオンコ遺伝子Unpは、過剰な発現がNIH3T3細胞の癌化を引き起こす核ユビキチンプロテアーゼ (nuclear ubiquitin protease) をコードする。この遺伝子のヒト相同体は、肺の小さな細胞腫瘍および腺癌においてそのレベルが一貫して高い (Gray, D. A. (1995) Oncogene 10: 2179 - 2183)。ユビキチンカルボキシル末端加水分解酵素は、リンパ芽球白血病細胞系の非分裂成熟状態への分化に関与する (Maki, A. 他 (1996) Differentiation 60: 59 - 66)。ヒト神経変性疾患に生じる異常な構造において、ニューロンにおけるユビキチンカルボキシル末端加水分解酵素 (PGP9.5) が高発現する。

このプロテアソームは、活性部位が中心腔に向いた4つの7員環に配置された様々なプロテアーゼを含む中心触媒コアと、腔の外側のポートを覆い基質のエントリーを調節する末端ATPアーゼサブユニットとから成る巨大多サブユニット複合体(～2000 kDa)である(Schmidt, M.他(1999) Curr. Opin. Chem. Biol. 3:584591を参照)。

【0011】

システインプロテアーゼ

システインプロテアーゼ(CP)は、前駆体タンパク質のプロセッシングから細胞内分解に至る多様な細胞プロセスに関係する。既知のCPの約半数がウイルスのみに存在する。CPは、活性部位における主な触媒残基としてシステインを有する。活性部位では、触媒作用がジオエステル中間体によって進行し、近接するヒスチジン残基およびアスパラギン酸残基によって促進される。グルタミン残基も重要であって、オキシアニオンホール形成を助ける。2つの重要なCPファミリーには、パパイン様酵素(C1)およびカルパイン(C2)が含まれる。パパイン様ファミリーメンバーは、通常はリソソーム酵素または分泌酵素であるためシグナルペプチドおよびプロペプチドから合成される。ほとんどのメンバーは、構造的な重要性を有し得るプロペプチドに保存されたモチーフを含む(Karrer, K. M.他(1993) Proc. Natl. Acad. Sci. USA 90:3063-3067)。パパインファミリーメンバーの3次元構造は、2つのローブ(lobe)の間に配置された触媒部位を備えた2ローブ分子(bilobed molecule)である。パパインには、カテプシンB、C、H、L、およびS、ある種の植物アレルゲンおよびジペプチジルペプチダーゼが含まれる(Rawlings, N. D.およびA. J. Barrett(1994) Meth. Enzymol. 244:461-486を参照)。

【0012】

或る種のCPは至る所で発現されるが、別の種類のCPは免疫系の細胞によってのみ発現される。注目すべきは、CPが、単球、マクロファージ、および炎症部位に遊走し組織の修復に参与する分子を分泌するその他の細胞によって生成される。これらの修復分子の過剰は、或る種の疾患に役割を果たしている。リウマチ様関節炎などの自己免疫疾患において、システインペプチダーゼカテプシンCの分泌によって、コラーゲン、ラミニン、エラスチン、および骨の細胞外マトリックスに見られるその他の構造タンパク質を分解する。このような分解によって脆弱化した骨はまた、腫瘍の侵入および転移を受け易い。カテプシンLの発現によって、リウマチ様滑液の破壊を進行させる単核細胞の流入が助けられ得る(Keyszer, G. M.(1995) Arthritis Rheum. 38:976-984)。

【0013】

カルパインはカルシウム依存性細胞質エンドペプチダーゼで、N末端触媒ドメインおよびC末端カルシウム結合ドメインを含む。カルパインは、それぞれのアイソホームに固有の触媒サブユニットと様々なアイソホームに共通の調節サブユニットとから成るプロ酵素ヘテロ二量体として発現される。それぞれのサブユニットは、カルシウム結合EFハンドドメインを有する。調節サブユニットはまた、酵素が細胞膜と結合できるようにする疎水性高グリシンドメインを含む。カルパインは細胞内のカルシウム濃度の上昇によって活性化され、構造における変化および制限された自己分解を引き起こす。得られる活性な分子は、その活性のために低いカルシウム濃度が必要である(Chan, S. L.およびM. P. Mattson(1999) J. Neurosci. Res. 58:167-190)。カルパインは主にニューロンで発現するが、その他の組織でも発現する。ALS、パーキンソン病、およびアルツハイマー病を含むある種の慢性の神経変性疾患は、カルパイン発現の上昇に関連する(ChanおよびMattson, 前出)。細胞骨格のカルパイン仲介性分解が、頭部外傷による脳障害に関係すると報告された(McCracken, E.他(1999) J. Neurotrauma 16:749-761)。カルパイン3は、主に骨格筋で発現され、肢帯筋ジストロフィー2A型に関係する(50

Minami, N. 他 (1999) J. Neurol. Sci. 171: 31-37)。

【0014】

カスパーゼはチオールプロテアーゼの別のファミリーであって、アポトーシスの開始および終了に関係する。前アポトーシス信号は、イニシエーターであるカスパーゼを活性化して、タンパク質分解カスパーゼカスケードを開始させ、標的タンパク質の加水分解および細胞の古典的なアポトーシス死を導く。システインおよびヒスチジンの2つの活性部位の残基が、触媒のメカニズムに関係する。カスパーゼは最も特異的なエンドペプチダーゼの1つであってアスパラギン酸残基の後を切断する。カスパーゼは、小さなスパー領域によって分離された1つの大きなサブユニット(p20)および1つの小さなサブユニット(p10)、および可変N末端プロドメインとから成る不活性な酵素前駆体として合成される。このプロドメインは、アポトーシスをポジティブ或いはネガティブな影響を与える補助因子と相互作用する。活性化信号による自己蛋白分解によって、特定のアスパラギン酸残基(番号付け規則に従うカスパーゼ1のD297)が切断され、スパーおよびプロドメインが除去され、p10/p20ヘテロ二量体が残る。これら2つのヘテロ二量体がそれらの小さなサブユニットを介して相互作用し、触媒的に活性な四量体を形成する。あるカスパーゼファミリーメンバーの長いプロドメインが、プロカスパーゼの二量体化および自己プロセッシングを促進することが分かった。ある種のカスパーゼは、そのプロドメインに「死効果ドメイン(death effector domain)」を含み、それによってカスパーゼが、死受容体またはTNF受容体複合体に関連するFADDタンパク質およびその他のカスパーゼと共に自己活性化複合体の中に動員され得る。更に、異なったカスパーゼファミリーメンバーからの2つの二量体が結合して、得られる四量体の基質特異性が変化し得る。内在性カスパーゼインヒビター(アポトーシスタンパク質のインヒビター、またはIAP)も存在する。これらの相互作用の全てがアポトーシスの制御に影響を与えていることが明らかである(ChanおよびMattson, 前出; Salveson, G. S. および V. M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96: 10964-10967を参照)。

10

20

【0015】

カスパーゼは或る種の疾患に関係する。ある種のカスパーゼが欠損したマウスは、感覚上皮のアポトーシスの失敗による重度の神経系の障害を持ち、早期に死亡する。その他のものは炎症反応における重度の障害を示す。例えば、カスパーゼは、IL-1bのプロセッシングに関係し、その他の炎症性サイトカインにも関係する可能性がある(ChanおよびMattson, 前出)。牛痘ウイルスおよびバキュロウイルスは、それらの宿主細胞の死を回避するためおよび感染を成功させるためにカスパーゼを標的とする。加えて、AIDS、神経変性疾患、および虚血性外傷で不適切なアポトーシスが增大し、細胞死の減少が癌に関連することが報告された(SalvesonおよびDixit, 前出: Thompson, C. B. (1995) Science 267: 1456-1462)。

30

【0016】

アスパルチルプロテアーゼ

アスパルチルプロテアーゼ(AP: aspartyl proteases)には、リソソームプロテアーゼカテプシンDおよびE、キモシン、レニン、および胃ペプシンが含まれる。ほとんどのレトロウイルスは、通常はp01ポリタンパク質の一部としてAPをコードする。酸性プロテアーゼとも呼ばれるAPは2つのドメインから成る単量体酵素で、各ドメインは自身の触媒アスパラギン酸残基を備えた1/2の活性部位を含む。APはpH2-3の範囲で最も活性であり、そのpHでは1つのアスパラギン酸残基がイオン化しており、その他は中性である。APのペプシンファミリーは多くの分泌酵素を含み、その全てがシグナルペプチドおよびプロペプチドで合成されると思われる。ほとんどのファミリーメンバーは3つのジスルフィドループを含み、第1のアスパラギン酸の後に最大5個の残基からなる第1のループがあり、第2のアスパラギン酸の前に5個から6個の残基が

40

50

ら成る第2のループがあり、第3の最も大きいループはC末端に向かって形成されている。一方、レトロペプシン (retropepsins) はペプシンの1つのドメインに類似しており、1/2の活性部位を提供するそれぞれのレトロペプシン単量体を有するホモ二量体として活性化する。レトロペプシンはウイルスポリタンパク質のプロセッシングに必要である。

【0017】

APは様々な組織に作用し、あるものは疾患に関係する。レニンは、電解質バランスおよび血圧の調節を担うホルモンであるアンジオテンシンのプロセッシングの第1のステップを仲介する (Crews, D. E. および S. R. Williams (1999) Hum. Biol. 71: 475 - 503 を参照)。様々な炎症性の疾患状態においてカテプシンの異常な調節および発現が起こることは明らかである。カテプシンDの発現は、リウマチ様関節炎および変形性関節症の患者の滑液膜組織において増大する。カテプシンの発現の上昇および異なった調節は様々な癌の転移能力に関係する (Chambers, A. F. 他 (1993) Crit. Rev. Oncol. 4: 95 - 114)。

10

【0018】

メタロプロテアーゼ

ほとんどの亜鉛依存性メタロプロテアーゼは亜鉛結合ドメインにおける共通の配列を共有する。この活性部位は、亜鉛リガンドとして作用する2つのヒスチジンと、第1のヒスチジンの触媒グルタミン酸C末端とから成る。このシグネチャ配列を含むタンパク質はmetzincinsとして知られ、アミノペプチダーゼN、アンジオテンシン変換酵素、神経溶解素、マトリックスメタロプロテアーゼ、およびadamalysins (ADAMS) を含む。亜鉛カルボキシペプチダーゼに別の配列が見つかり、その配列では、2つのヒスチジン残基および1つのグルタミン酸残基の3つの保存された残基が亜鉛結合に関与する。

20

【0019】

アンジオテンシン変換酵素およびアミノペプチダーゼを含むいくつかの中性メタロエンドペプチダーゼは、ペプチドホルモンの代謝に関与する。例えば、アミノペプチダーゼBの高い活性が、高血圧のラットの甲状腺および神経下垂体に見られる (Prieto, I. 他 (1998) Horm. Metab. Res. 30: 246 - 248)。オリゴペプチダーゼM/神経溶解素は、ブラジキニンおよびニューロテンシンを加水分解できる (Serizawa, A. 他 (1995) J. Biol. Chem 270: 2092 - 2098)。ニューロテンシンは血管作動性ペプチドであって、脳において神経伝達物質として働くことができ、食事摂取の制限に関係すると思われる (Tritos, N. A. 他 (1999) Neuropeptides 33: 339 - 349)。

30

【0020】

マトリックスメタロプロテアーゼ (MMP) は、細胞外マトリックス (ECM) の成分を分解できる少なくとも23の酵素のファミリーである。MMPはN末端触媒ドメインを有するZn⁺² エンドペプチダーゼである。このファミリーのほとんど全てのメンバーが、組織によって生成されるインヒビターまたはECMの基質分子に結合可能なC末端ドメインおよびヒンジペプチド (hinge peptide) を有する (TIMP: メタロプロテアーゼの組織インヒビター; Campbell, I. L. 他 (1999) Trends Neurosci. 22: 285)。フィブルロネクチン様リピート、膜貫通ドメイン、またはC末端hemopexinase様ドメインの存在を利用して、MMPをコラーゲナーゼ、ゼラチナーゼ、ストロメライシンおよび膜型MMPサブファミリーに分けることができる。活性型では、活性部位におけるZn⁺² イオンがプロ配列のシステインと相互作用する。活性化因子が、活性部位を露出させるZn⁺² システイン相互作用または「システインスイッチ」を阻害する。これによって酵素が部分的に活性化され、次にそのプロペプチドを切断して完全に活性化する。MMPは、セリンプロテアーゼであるプラスミンおよびfurinによって活性化される場合が多い。MMPは、インヒビターとの化学量論的非共有結合相互作用、すなわちプロテアーゼとインヒビターとのバランス

40

50

によって調節される場合が多く、組織恒常性において極めて重要である。(Yong, V. W. 他 (1998) Trends Neurosci. 21:75を参照)。

【0021】

MMPは様々な疾患に関与し、その中には骨関節炎(Mitchell, P. ら (1996) J. Clin. Invest. 97:761)、アテローム斑破裂(Sukhova, G. K. ら (1999) Circulation 99:2503)、大動脈瘤(Schneiderman, J. ら (1998) Am. J. Pathol. 152:703)、非創傷治癒(Saarialho-Kere, U. K. ら (1994) J. Clin. Invest. 94:79)、骨再吸収(Blavier, L. および J. M. Delaisse (1995) J. Cell Sci. 108:3649)、加齢性黄斑変性症(Steen, B. ら (1998) Invest. Ophthalmol. Vis. Sci. 39:2194)、肺気腫(inlay, G. A. ら (1997) Thorax 52:502)、心筋梗塞(Rohde, L. E. ら (1999) Circulation 99:3063)、および拡張型心筋症(Thomas, C. V. ら (1998) Circulation 97:1708)が含まれる。MMPインヒビターは、ラットにおける乳癌および実験的腫瘍の転移を阻害し、マウスにおけるルイス肺癌、血管腫、およびヒト卵巣癌異種移植片の転移を阻害する(Eccles, S. A. 他 (1996) Cancer Res. 56:2815; Anderson 他 (1996) Cancer Res. 56:715-718; Volpert, O. V. 他 (1996) J. Clin. Invest. 98:671; Taraboletti, G. 他 (1995) J. NCI 87:293; Davies, B. 他 (1993) Cancer Res. 53:2087)。MMPはアルツハイマー病で活性化しているであろう。或る種のMMPが多発性硬化症に関与し、MMPインヒビターの投与でその症状の一部を緩和できる(Yong, 前出、を参照)。

【0022】

メタロプロテアーゼの別のファミリーはADAM(disintegrinドメインおよびメタロプロテアーゼドメイン)であり、ヘビ毒メタロプロテアーゼ(SVMP)であるadamalysinsの類似体を共有する。ADAMは細胞表面接着分子およびプロテアーゼの両方の構造を組み合わせたものであって、プロドメイン、プロテアーゼドメイン、disintegrinドメイン、高システインドメイン、上皮成長因子リピート、膜貫通ドメイン、および細胞質尾部を含む。これらの最初の3つのドメインはSVMPにも見られる。ADAMは、蛋白分解、接着、シグナル伝達、および融合の4つの潜在的な機能を有する。ADAMは、metzincin亜鉛結合配列を共有し、TIMP-1などのある種のMMPアンタゴニストによって阻害される。

【0023】

ADAMは、精子-卵結合および融合、筋芽細胞の融合や、サイトカイン、サイトカイン受容体、接着タンパク質およびその他の細胞外タンパク質ドメインのタンパク質外部ドメインのプロセッシングまたは分断などのプロセスに関与する(Schlondorff, J. および C. P. Blobel (1999) J. Cell. Sci. 112:3603-3617)。Kuzbaniaタンパク質は、NOTCH経路の基質を切断し(NOTCH自体も可能)、ショウジョウバエの神経の発達における側方の阻害のためのプログラムを活性化する。TACE(ADAM17)およびADAM10の2つのADAMは、脳におけるアミロイド前駆体タンパク質のプロセッシングにおいて類似の機能を果たすと提案された(SchlondorffおよびBlobel, 前出)。TACEもまた、TNF活性化酵素として同定された(Black, R. A. 他 (1997) Nature 385:729)。TNFは多面発現サイトカインであって、感染或いは外傷に応じる宿主防御の動員に重要であるが、過剰に発現すると重度の障害を与え、自己免疫疾患で過剰に発現される場合が多い。TACEは、膜結合プロTNFを切断して可溶型を放出する。他のADAMはその他の膜結合分子の類似のプロセッシングに関与すると思われ

る。

【0024】

ADAMTSサブファミリーはADAMファミリーメタロプロテアーゼの全ての特徴を有し、更にトロンボスポンジドメイン(TS)を含む。プロトタイプADAMTSがマウスで同定され、心臓および腎臓で発現され、炎症誘発性刺激によってアップレギュレートされる(Kuno, K.他(1997) J. Biol. Chem. 272:556)。これまでにHuman Genome Organizationによって11のメンバーが確認された(HUGO; <http://www.gene.ucl.ac.uk/users/hester/adamts.html#Approved>)。このファミリーのメンバーは、関節炎の進行により喪失する収縮能力を含む重要な力学的機能を軟骨に与える高分子量のプロテオグリカンであるaggrecanを分解する能力を有する。従って、aggrecanを分解する酵素は、間接の軟骨の分解を防止或いは緩和するための魅力的な標的と考えられる(例えば、Tortorella, M. D. (1999) Science 284:1664; Abbaszade, I. (1999) J. Biol. Chem. 274:23443を参照)。その他のメンバーは、抗血管形成の可能性(Kuno他、前出)および/またはプロコラーゲンのプロセッシングに関係すると報告された(Colige, A.他(1997) Proc. Natl. Acad. Sci. USA 94:2374)。

10

【0025】

プロテアーゼインヒビター

プロテアーゼインヒビターおよびその他のプロテアーゼ活性の調節因子は、プロテアーゼの活性および効果を調節する。プロテアーゼインヒビターは、蛋白分解に関連する疾患の動物モデルにおける病因を調節すると報告された(Murphy, G. (1991) Agents Actions Suppl. 35:69-76)。システインプロテアーゼの低分子量インヒビターであるシスタチンの低下が悪性腫瘍の進行に相関する(Callkins, C.他(1995) Biol. Biochem. Hoppe Seyler 376:71-80)。serpinsは哺乳動物血漿セリンプロテアーゼのインヒビターである。多くのserpinsが、哺乳動物における血液凝固カスケードおよび/または補体カスケードの調節に作用する。SP32は哺乳動物アクロソームプロテアーゼであるアクロシンの正の調節因子であって、プロ酵素であるプロアクロシンと結合してその酵素をアクロソームマトリックスに折り畳む助けをする(Baba, T.他(1994) J. Biol. Chem. 269:10133-10140)。セリンプロテアーゼインヒビターのKunitzファミリーは1或いは複数の「Kunitzドメイン」によって特徴づけられる。Kunitzドメインは、約50個のアミノ酸残基の間隔で規則的に離れ、3つの鎖内ジスルフィド結合を形成する一連のシステイン残基を含む。このファミリーのメンバーには、アプロチニン、組織因子経路インヒビター(TFPI-1およびTFPI-2)、インター-トリプシンインヒビター、およびbikuninが含まれる(Marlor, C. W.他(1997) J. Biol. Chem. 272:12202-12208)。このファミリーのメンバーが、カリクレインおよびプラスミンなどのセリンプロテアーゼに対する強力なインヒビター(ナノモル範囲における)である。アプロチニンは、手術中の血液損失の抑制に有用である。

20

30

40

【0026】

新規のプロテアーゼおよびそれらをコードするポリヌクレオチドの発見により、新規の組成物を提供することで当分野の要望に応えることができる。この新規の組成物は、胃腸疾患、心血管疾患、自己免疫/炎症性の疾患、細胞増殖異常、発生障害、上皮の疾患、神経の疾患、および生殖障害の診断・治療・予防において有用であり、また、プロテアーゼの核酸配列及びアミノ酸配列の発現における外来性化合物の効果についての評価にも有用である。

【0027】

(発明の要約)

50

本発明は、総称して「PRTS」、個別にはそれぞれ「PRTS - 1」、「PRTS - 2」、「PRTS - 3」、「PRTS - 4」、「PRTS - 5」、「PRTS - 6」、「PRTS - 7」、「PRTS - 8」、「PRTS - 9」、「PRTS - 10」、「PRTS - 11」、「PRTS - 12」、「PRTS - 13」、及び「PRTS - 14」と呼ぶプロテアーゼである精製されたポリペプチドを提供する。本発明の一実施態様では、(a) SEQ ID NO: 1乃至14 (SEQ ID NO: 1 - 14) からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 14 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含む単離されたポリペプチドを提供する。別法では、SEQ ID NO: 1 - 14 のアミノ酸配列を含む単離されたポリペプチドを提供する。

【0028】

更に本発明は、(a) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドをコードする単離されたポリヌクレオチドを提供する。別法では、このポリヌクレオチドは、SEQ ID NO: 1 - 14 からなる一群から選択されたポリペプチドをコードする。別法では、このポリヌクレオチドは、SEQ ID NO: 15 - 28 からなる一群から選択される。

【0029】

更に、本発明は、(a) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドをコードするポリヌクレオチドと機能的に結合されたプロモーター配列を含む組換えポリヌクレオチドを提供する。別法では、本発明は、この組換えポリヌクレオチドで形質転換された細胞を提供する。更なる別法では、本発明は、この組換えポリヌクレオチドを含む遺伝子組換え生物を提供する。

【0030】

また、本発明は、(a) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 14 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドの生産方法を提供する。この方法は、(a) このポリペプチドの発現に好適な条件下で、このポリペプチドをコードするポリヌクレオチドと機能的に結合されたプロモーター配列を含む組換えポリヌクレオチドで形質転換された細胞を培養するステップと、(b) このように発現したポリペプチドを回収するステップとを含む。

【0031】

更に、本発明は、(a) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) S E

Q I D N O : 1 - 1 4 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドに特異的に結合する単離された抗体を提供する。

【 0 0 3 2 】

更に、本発明は、(a) S E Q I D N O : 1 5 - 2 8 からなる一群から選択されたポリヌクレオチド配列と、(b) S E Q I D N O : 1 5 - 2 8 からなる一群から選択されたポリヌクレオチド配列と 9 0 % 以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記 (a) に相補的なポリヌクレオチド配列と、(d) 前記 (b) に相補的なポリヌクレオチド配列と、(e) 前記 (a) 乃至 (d) の R N A 等化物とで構成される一群から選択されたポリヌクレオチド配列を含む単離されたポリヌクレオチドを提供する。別法では、このポリヌクレオチドは、少なくとも 6 0 個の連続するヌクレオチドを含む。

10

【 0 0 3 3 】

更に本発明は、(a) S E Q I D N O : 1 5 - 2 8 からなる一群から選択されたポリヌクレオチド配列と、(b) S E Q I D N O : 1 5 - 2 8 からなる一群から選択されたポリヌクレオチド配列と 9 0 % 以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記 (a) に相補的なポリヌクレオチド配列と、(d) 前記 (b) に相補的なポリヌクレオチド配列と、(e) 前記 (a) 乃至 (d) の R N A 等化物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチド配列を有する標的ポリヌクレオチドをサンプルにおいて検出する方法を提供する。この方法は、(a) 前記サンプル内の標的ポリヌクレオチドと相補的な配列を構成する少なくとも 2 0 個の連続するヌクレオチドを含むプローブと前記サンプルをハイブリダイズさせるステップであって、前記プローブと前記標的ポリヌクレオチドまたはその断片とでハイブリダイゼーション複合体が形成される条件下で、前記プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズする、該ステップと、(b) 前記ハイブリダイゼーション複合体の存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含む。別法では、前記プローブは、少なくとも 6 0 個の連続するヌクレオチドを含む。

20

【 0 0 3 4 】

更に本発明は、(a) S E Q I D N O : 1 5 - 2 8 からなる一群から選択されたポリヌクレオチド配列と、(b) S E Q I D N O : 1 5 - 2 8 からなる一群から選択されたポリヌクレオチド配列と 9 0 % 以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記 (a) に相補的なポリヌクレオチド配列と、(d) 前記 (b) に相補的なポリヌクレオチド配列と、(e) 前記 (a) 乃至 (d) の R N A 等化物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチド配列を有する標的ポリヌクレオチドをサンプルにおいて検出する方法を提供する。この方法は、(a) ポリメラーゼ連鎖反応増幅を用いて、前記標的ポリヌクレオチドまたはその断片を増幅するステップと、(b) 増幅された前記標的ポリヌクレオチドまたはその断片が存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含む。

30

【 0 0 3 5 】

更に本発明は、(a) S E Q I D N O : 1 - 1 4 からなる一群から選択されたアミノ酸配列と、(b) S E Q I D N O : 1 - 1 4 からなる一群から選択されたアミノ酸配列と 9 0 % 以上の配列同一性を有する天然のアミノ酸配列と、(c) S E Q I D N O : 1 - 1 4 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) S E Q I D N O : 1 - 1 4 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含む効果的な量のポリペプチド及び好適な医薬用賦形剤を含む組成物を提供する。一実施例では、S E Q I D N O : 1 - 1 4 からなる一群から選択されたアミノ酸配列を含む組成物を提供する。更に、本発明は、患者にこの組成物を投与することを含む、機能的 P R T S の発現の低下に関連した疾患やその症状の治療方法を提供する。

40

【 0 0 3 6 】

更に本発明は、(a) S E Q I D N O : 1 - 1 4 からなる一群から選択されたアミノ酸

50

配列と、(b) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1-14 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドのアゴニストとして効果的な化合物をスクリーニングする方法を提供する。この方法は、(a) このポリペプチドを含むサンプルを化合物に曝露するステップと、(b) このサンプルのアゴニスト活性を検出するステップとを含む。別法では、本発明は、この方法によって同定されたアゴニスト化合物と好適な医薬用賦形剤とを含む組成物を提供する。更なる別法では、本発明は、この組成物の患者への投与を含む、機能的 P R T S の発現の低下に関連した疾患やその症状の治療方法を提供する。

10

【0037】

更に、本発明は、(a) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1-14 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドのアンタゴニストとして効果的な化合物をスクリーニングする方法を提供する。この方法は、(a) このポリペプチドを含むサンプルを化合物に曝露するステップと、(b) このサンプルのアンタゴニスト活性を検出するステップとを含む。別法では、本発明は、この方法によって同定されたアンタゴニスト化合物と好適な医薬用賦形剤とを含む組成物を提供する。更なる別法では、本発明は、この組成物の患者への投与を含む、機能的 P R T S の過剰な発現に関連した疾患やその症状の治療方法を提供する。

20

【0038】

更に本発明は、(a) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1-14 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドに特異的に結合する化合物をスクリーニングする方法を提供する。この方法は、(a) このポリペプチドを好適な条件下で少なくとも1つの化合物と結合させるステップと、(b) このポリペプチドとこの試験化合物との結合を検出して、このポリペプチドと特異的に結合する化合物を同定するステップとを含む。

30

【0039】

更に本発明は、(a) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1-14 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1-14 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドの活性を調節する化合物をスクリーニングする方法を提供する。このスクリーニング方法は、(a) このポリペプチドを、その活性が許容される条件下で少なくとも1つの化合物と結合させるステップと、(b) この試験化合物の存在下でのこのポリペプチドの活性を評価するステップと、(c) この試験化合物の存在下でのこのポリペプチドの活性と、この試験化合物の不在下でのこのポリペプチドの活性とを比較するステップとを含み、この試験化合物の存在下でのこのポリペプチドの活性の変化が、このポリペプチドの活性を調節する化合物の存在を示唆するという特徴を有する。

40

【0040】

50

更に本発明は、SEQ ID NO: 15 - 28 からなる一群から選択された配列を含む標的ポリヌクレオチドの発現を変化させるのに効果的な化合物をスクリーニングする方法であって、(a) この標的ポリヌクレオチドを含むサンプルを化合物に曝露するステップと、(b) この標的ポリヌクレオチドの発現の変化を検出するステップとを含む、該スクリーニング方法を提供する。

【0041】

本発明はさらに、試験化合物の毒性を評価する方法を提供する。この方法は、(a) 核酸を含む生体サンプルを前記試験化合物で処理するステップと、(b) 処理した前記生体サンプルの核酸をプローブとハイブリダイズするステップと、(c) ハイブリダイゼーション複合体の収量を測定するステップと、(d) 前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量を、未処理の生体サンプルにおけるハイブリダイゼーション複合体の収量とを比較するステップとを含み、前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量の差異が試験化合物の毒性を示唆する。この方法における前記プローブは、(1) SEQ ID NO: 15 - 28 からなる一群から選択されたポリヌクレオチド配列と、(2) SEQ ID NO: 15 - 28 からなる一群から選択されたポリヌクレオチド配列と少なくとも90%の配列同一性を有する天然のポリヌクレオチド配列と、(3) 前記(1)に相補的なポリヌクレオチド配列と、(4) 前記(2)に相補的なポリヌクレオチド配列と、(5) 前記(1)乃至(4)のRNA等価物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチドの連続する少なくとも20個のヌクレオチドを含む。また、前記ハイブリダイゼーションは、前記プローブと前記生体サンプルの標的ポリヌクレオチドとの間で特異的なハイブリダイゼーション複合体が形成される条件下で行われる。また、前記標的ポリヌクレオチドが、(1) SEQ ID NO: 15 - 28 からなる一群から選択されたポリヌクレオチド配列と、(2) SEQ ID NO: 15 - 28 からなる一群から選択されたポリヌクレオチド配列と少なくとも90%の配列同一性を有する天然のポリヌクレオチド配列と、(3) 前記(1)に相補的なポリヌクレオチド配列と、(4) 前記(2)に相補的なポリヌクレオチド配列と、(5) 前記(1)乃至(5)のRNA等価物とを含む。代替的に前記標的ポリヌクレオチドは前記ポリヌクレオチド配列の断片である。

【0042】

(本発明の記載について)

本発明のタンパク質及び核酸配列、方法について説明する前に、本発明は、ここに開示した特定の装置及び材料、方法に限定されず、その実施形態を変更できることを理解されたい。また、ここで用いられる用語は、特定の実施例のみを説明する目的で用いられたものであり、後述の請求の範囲によってのみ限定され、本発明の範囲を限定することを意図したものではないということも理解されたい。

【0043】

本明細書及び請求の範囲において単数形を表す「或る」、「その(この等)」は、文脈で明確に示していない場合は複数形を含むことに注意されたい。従って、例えば「或る宿主細胞」は複数の宿主細胞を含み、その「抗体」は複数の抗体は含まれ、当業者には周知の等価物なども含まれる。

【0044】

本明細書で用いた全ての科学技術用語は、別の方法で定義されていない限り、本発明の属する技術分野の一般的な技術者が普通に解釈する意味と同じである。本明細書で記述したものと類似、或いは同等の全ての装置及び材料、方法は本発明の実施及びテストに使用できるが、好適な装置及び材料、方法をここに記す。本明細書に記載の全ての文献は、本発明に関連して使用する可能性のある文献に記載された細胞系、プロトコル、試薬、ベクターを記述し開示するために引用した。従来発明を引用したからと言って、本発明の新規性が損なわれると解釈されるものではない。

【0045】

(定義)

用語「P R T S」は、天然、合成、半合成或いは組換え体など全ての種（特にウシ、ヒツジ、ブタ、マウス、ウマ及びヒトを含む哺乳動物）から得られる実質的に精製されたP R T Sのアミノ酸配列を指す。

【0046】

用語「アゴニスト」は、P R T Sの生物学的活性を強めたり、模倣する分子を指す。このアゴニストは、P R T Sに直接相互作用するか、或いはP R T Sが関与する生物学的経路の成分と作用して、P R T Sの活性を調節するタンパク質、核酸、糖質、小分子、任意の他の化合物や組成物を含み得る。

【0047】

用語「アレル変異配列」は、P R T Sをコードする遺伝子の別の形を指す。アレル変異配列は、核酸配列における少なくとも1つの変異によって生じ、変異mRNA若しくは変異ポリペプチドになり、これらの構造や機能は変わる場合もあれば変わらない場合もある。ある遺伝子は、天然型のアレル変異配列が存在しないもの、1つ或いは多数存在するものがある。一般にアレル変異配列を生じる変異は、ヌクレオチドの自然な欠失、付加、或いは置換による。これらの各変異は、単独或いは他の変異と同時に起こり、所定の配列内で一回或いはそれ以上生じる。

10

【0048】

P R T Sをコードする「変異」核酸配列は、様々なヌクレオチドの欠失、挿入、或いは置換が起こっても、P R T Sと同じポリペプチド或いはP R T Sの機能特性の少なくとも1つを備えるポリペプチドを指す。この定義には、P R T Sをコードするポリヌクレオチド配列の正常な染色体の遺伝子座ではない位置でのアレル変異配列との不適當或いは予期しないハイブリダイゼーション、並びにP R T Sをコードするポリヌクレオチドの特定のオリゴヌクレオチドプローブを用いて容易に検出可能な或いは検出困難な多形性を含む。コードされたタンパク質も変異され得り、サイレント変化を生じP R T Sと機能的に等価となるアミノ酸残基の欠失、挿入、或いは置換を含み得る。意図的なアミノ酸置換は、生物学的或いは免疫学的にP R T Sの活性が保持される範囲で、残基の極性、電荷、溶解度、疎水性、親水性、及び/または両親媒性についての類似性に基づいて成され得る。例えば、負に荷電したアミノ酸にはアスパラギン酸及びグルタミン酸が含まれ、正に荷電したアミノ酸にはリシン及びアルギニンが含まれ得る。類似の親水性の値をもち極性非荷電側鎖を有するアミノ酸には、アスパラギン、グルタミン、セリン、トレオニンが含まれ得る。類似の親水性の値をもち非荷電側鎖を有するアミノ酸には、ロイシン、イソロイシン、バリン、グリシン、アラニン、フェニルアラニン及びチロシンが含まれ得る。

20

30

【0049】

用語「アミノ酸」及び「アミノ酸配列」は、オリゴペプチド、ペプチド、ポリペプチド、タンパク質配列、或いはそれらの任意の断片を指し、天然の分子及び合成分子を含む。「アミノ酸配列」が天然のタンパク質分子の配列を指す場合、「アミノ酸配列」及び類似の用語は、アミノ酸配列を記載したタンパク質分子に関連する完全で元のままのアミノ酸配列に限定するものではない。

【0050】

用語「増幅」は、核酸配列の複製物を作製することに関連する。一般に増幅は、この技術分野で周知のポリメラーゼ連鎖反応（PCR）技術によって行われる。

40

【0051】

用語「アンタゴニスト」は、P R T Sの生物学的活性を阻害或いは減弱する分子である。アンタゴニストは、P R T Sに直接相互作用するか、或いはP R T Sが関与する生物学的経路の成分と作用して、P R T Sの活性を調節する抗体、核酸、糖質、小分子、任意の他の化合物や組成物などのタンパク質を含み得る。

【0052】

用語「抗体」は、抗原決定基と結合可能なFab及びF(ab')₂、及びそれらの断片、Fv断片などの無傷の分子を指す。P R T Sポリペプチドと結合する抗体は、抗体を免疫する小ペプチドを含む無傷の分子またはその断片を用いて作製可能である。動物（例え

50

ば、マウス、ラット、若しくはウサギ)を免疫化するのに使用されるポリペプチド或いはオリゴペプチドは、RNAの翻訳から、或いは化学的に合成可能であり、必要に応じて担体タンパク質と結合させることも可能である。ペプチドと化学的に結合した一般に用いられる担体は、ウシ血清アルブミン、チログロブリン、及びキーホールリンペットヘモニアン(KLH)を含む。次ぎに、この結合したペプチドを用いて動物を免疫化する。

【0053】

用語「抗原決定基」は、特定の抗体と接触する分子の領域(即ちエピトープ)を指す。タンパク質或いはタンパク質の断片が、宿主動物を免疫化するのに用いられるとき、このタンパク質の種々の領域は、抗原決定基(タンパク質上の特定の領域或いは三次元構造体)に特異的に結合する抗体の産生を誘発し得る。抗原決定基は、抗体と結合するために無傷の抗原(即ち、免疫応答を引き出すために用いられる免疫原)と競合し得る。

10

【0054】

本明細書において「アンチセンス」は、特定の核酸配列のセンス(コーディング)鎖と塩基対を形成し得る任意の組成物を指す。アンチセンス成分には、DNAと、RNAと、ペプチド核酸(PNA)と、ホスホロチオネートやメチルホスホネート、ベンジルホスホネート(benzylphosphonate)などの修飾された骨格(backbone linkage)を有するオリゴヌクレオチドと、2'-メトキシエチル糖または2'-メトキシエトキシ糖などの修飾された糖を有するオリゴヌクレオチドと、5-メチルシトシンまたは2'-deoxyuracil、7-deaza-2'-deoxyguanosineなどの修飾された塩基を有するオリゴヌクレオチドを含み得る。アンチセンス分子は、化学合成や転写を含む任意の方法で作ることができる。相補的アンチセンス分子は、一度細胞に導入されると、細胞によって作られた天然の核酸配列と塩基対となって二重鎖を形成し、転写や翻訳を阻害する。「負」または「マイナス」という表現はアンチセンス鎖であり、「正」または「プラス」という表現はセンス鎖である。

20

【0055】

用語「生物学的に活性」は、天然分子の構造的、調節的、或いは生化学的な機能を有するタンパク質を指す。同様に、用語「免疫学的に活性」または「免疫原性」は、天然或いは組換え体のPRTS、合成のPRTSまたはそれらの任意のオリゴペプチドが、適当な動物或いは細胞の特定の免疫応答を誘発して特定の抗体と結合する能力を指す。

【0056】

用語「相補的」は、塩基対合によってアニールする2つの一本鎖核酸配列間の関係を指す。例えば、配列「5'AGT3'」が相補的な配列「3'TCA5'」と対をなす。

30

【0057】

「所定のポリヌクレオチド配列を含む組成物」または「所定のアミノ酸配列を含む組成物」は広い意味で、所定のヌクレオチド配列若しくはアミノ酸配列を含む任意の組成物を指す。この組成物は、乾燥した製剤或いは水溶液を含み得る。PRTS若しくはPRTSの断片をコードするポリヌクレオチド配列を含む組成物は、ハイブリダイゼーションプローブとして使用され得る。このプローブは、凍結乾燥状態で保存可能であり、糖質などの安定化剤と結合させることが可能である。ハイブリダイゼーションにおいて、プローブは、塩(例えば、NaCl)及び界面活性剤(例えば、SDS:ドデシル硫酸ナトリウム)、その他の物質(例えば、デンハート液、乾燥ミルク、サケ精子DNAなど)を含む水溶液に展開され得る。

40

【0058】

「コンセンサス配列」は、不要な塩基を分離するためにDNA配列の解析を繰り返し行い、XL-PCRキット(PE Biosystems, Foster City CA)を用いて5'及び/または3'の方向に伸長され、再度シーケンシングされた核酸配列、またはGELVIEW断片構築システム(GCG, Madison, WI)またはPhrap(University of Washington, Seattle WA)等の断片構築用のコンピュータプログラムを用いて1つ或いはそれ以上の重複するcDNAやEST、またはゲノムDNA断片から構築された核酸配列を指す。伸長及び重複の

50

両方によって構築されるコンセンサス配列もある。

【0059】

用語「保存的なアミノ酸置換」は、元のタンパク質の特性を殆ど変えない置換を指す。即ち、置換によってそのタンパク質の構造や機能が大きくは変わらず、そのタンパク質の構造、特にその機能が保存される。以下に、あるタンパク質の元のアミノ酸が別のアミノ酸に置換される保存的なアミノ酸置換を示す。

元の残基	保存的な置換
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

10

20

一般に、保存されたアミノ酸置換の場合は、a)置換された領域のポリペプチドの骨格構造、例えば、シートやヘリックス高次構造、b)置換された部位の分子の電荷または疎水性、及び/または、c)側鎖の大半が維持される。

【0060】

用語「欠失」は、1個以上のアミノ酸残基が欠如するアミノ酸配列の変化、或いは1個以上のヌクレオチドが欠如する核酸配列の変化を指す。

30

【0061】

用語「誘導体」は、化学修飾されたポリヌクレオチドまたはポリペプチドを指す。ポリヌクレオチド配列の化学修飾には、例えば、アルキル基、アシル基、ヒドロキシル基、或いはアミノ基による水素の置換がある。誘導体ポリヌクレオチドは、自然分子(未修飾の分子)の生物学的或いは免疫学的機能の少なくとも1つを維持するポリペプチドをコードする。誘導体ポリペプチドとは、もとのポリペプチドの生物学的機能、或いは免疫学的機能の少なくとも1つを維持する、グリコシル化、ポリエチレングリコール化、或いは任意の同様のプロセスによって修飾されたポリペプチドのことである。

40

【0062】

「検出可能な標識」は、測定可能な信号を生成し得る、ポリヌクレオチドやポリペプチドに共有結合或いは非共有結合するレポーター分子や酵素を指す。

【0063】

用語「断片」は、PRTSまたはPRTSをコードするポリヌクレオチドの固有の部分であって、その親配列(parent sequence)と同一であるがその配列より長さが短いものを指す。「断片」の最大の長さは、親配列から1つのヌクレオチド/アミノ酸残基を差し引いた長さである。例えば、ある断片は、5~1000個の連続するヌクレオチド或いはアミノ酸残基を含む。プローブ、プライマー、抗原、治療用分子、またはその他の目的に用いる断片は、少なくとも5、10、15、16、20、25、30、40

50

、50、60、75、100、150、250若しくは500個の連続するヌクレオチド
或いはアミノ酸残基の長さである。断片は、優先的に分子の特定の領域から選択される場
合もある。例えば、ポリペプチド断片は、所定の配列に示された最初の250若しくは5
00のアミノ酸（或いは、ポリペプチドの最初の25%または50%）から選択された連
続するアミノ酸の所定の長さを含み得る。これらの長さは一例であり、配列表及び表、図
面を含む明細書に記載の任意の長さが、本発明の実施例に含まれ得る。

【0064】

SEQ ID NO: 15 - 28の断片は、例えば、この断片を得たゲノム内の他の配列と
は異なる、SEQ ID NO: 15 - 28を明確に同定する固有のポリヌクレオチド配列
の領域を含む。SEQ ID NO: 15 - 28のある断片は、例えば、ハイブリダイゼー
ションや増幅技術、またはSEQ ID NO: 15 - 28を関連ポリヌクレオチド配列から
区別する類似の方法に有用である。ある断片と一致するSEQ ID NO: 15 - 28
の正確な断片の長さや領域は、その断片の目的に基づいて当分野で一般的な技術によって
日常的に測定できる。

10

【0065】

SEQ ID NO: 1 - 14のある断片は、SEQ ID NO: 15 - 28のある断片に
よってコードされる。SEQ ID NO: 1 - 14のある断片は、SEQ ID NO: 1
- 14を特異的に同定する固有のアミノ酸配列の領域を含む。例えば、SEQ ID NO
: 1 - 14のある断片は、SEQ ID NO: 1 - 14を特異的に認識する抗体の開発に
おける免疫原性ペプチドとして有用である。ある断片と一致するSEQ ID NO: 1 -
14の正確な断片の長さや領域は、その断片の目的に基づいて当分野で一般的な技術によ
って日常的に決定できる。

20

【0066】

「完全長」ポリヌクレオチド配列とは、少なくとも1つの翻訳開始コドン（例えばメチオ
ニン）、それに続くオープンリーディングフレーム及び翻訳終止コドンを有する配列であ
る。「完全長」ポリヌクレオチド配列は、「完全長」ポリペプチド配列をコードする。

【0067】

「相同性」は、2つ以上のポリヌクレオチド配列間または2つ以上のポリペプチド配列間
の配列類似性である。この配列類似性は配列同一性と言い換えることができる。

【0068】

ポリヌクレオチド配列についての用語「パーセントの同一性」または「%の同一性」とは
、標準化されたアルゴリズムを用いてアラインメントされる、2つ以上のポリヌクレオチ
ド配列間の一致する残基の百分率のことである。このようなアルゴリズムは、標準化され
再現できる方法で、2つの配列間のアラインメントを最適化するべく、配列にギャップを
挿入して、より意味をもつ2つの配列間の比較を行うことができる。

30

【0069】

ポリヌクレオチド配列間の同一性のパーセントは、MEGALIGN version 3
. 12e配列アラインメントプログラムに組込まれるCLUSTAL Vアルゴリズムの
デフォルトパラメータを用いて決定可能である。このプログラムはLASERGENEソ
フトウェアパッケージの一部であり、分子生物学分析プログラム一式(DNASTAR,
Madison WI)である。このCLUSTAL Vは、Higgins, D. G.
及びP. M. Sharp (1989) CABIOS 5: 151 - 153、Higgin
s, D. G. 他 (1992) CABIOS 8: 189 - 191に記載されてい
る。ポリヌクレオチド配列の対のアライメントの場合、デフォルトパラメータは、Ktu
ple = 2、gap penalty = 5、window = 4、「diagonals s
aved」= 4と設定する。「重み付けされた」残基重み付け表が、デフォルトとして選
択された。同一性のパーセントは、アラインメントされたポリヌクレオチド配列間の「類
似性のパーセント」としてCLUSTAL Vによって報告される。

40

【0070】

別法では、National Center for Biotechnology Inf

50

ormation (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410) が提供する、広く用いられている無料の配列比較アルゴリズム一式が、NCBI (Bethesda, MD) を含む幾つかのソース及びインターネット (<http://WWW.ncbi.nlm.nih.gov/BLAST/>) で入手可能である。このBLASTソフトウェア一式には、既知のポリヌクレオチド配列と様々なデータベースの別のポリヌクレオチド配列とのアラインメントに用いられる「blastn」を含む、様々な配列分析プログラムが含まれる。「BLAST 2 Sequences」と呼ばれるツールが入手可能であり、2つのヌクレオチド配列の対を直接比較するために用いられる。「BLAST 2 Sequences」は、<http://WWW.ncbi.nlm.nih.gov/gorf/b12.html> にアクセスして、対話形式で利用できる。「BLAST 2 Sequences」ツールは、blastn 及び blastp (以下に記載) の両方に用いることができる。BLASTプログラムは、一般的には、デフォルトを設定するギャップ及び他のパラメータと共に用いられる。例えば、2つのヌクレオチド配列を比較する場合、ある者は、デフォルトパラメータに設定された「BLAST 2 Sequences」ツール Version 2.0.12 (April-21-2000) でblastnを使用するであろう。そのようなデフォルトパラメータは、例えば、以下のようにする。

10

20

30

40

50

【0071】

```
Matrix : BLOSUM62
Reward for match : 1
Penalty for mismatch : -2
Open Gap : 5 及び Extension Gap : 2 penalties
Gap x drop-off : 50
Expect : 10
Word Size : 11
Filter : on
```

同一性のパーセントは、例えば、特定の配列番号で決められた、所定の配列の全長に対して測定してもよいし、それより短い長さに対して、例えば、ある大きな所定の配列から得られた断片、例えば、連続する少なくとも、20または30、40、50、70、100、200のヌクレオチドの断片の長さに対して測定してもよい。このような長さは単なる例であり、配列表及び表、図面を含む明細書に記載の配列の任意の長さの断片を用いて、同一性のパーセントが測定される長さを示すことができる。

【0072】

高い同一性を示さない核酸配列でも、遺伝子コードの縮重によって類似のアミノ酸配列をコードし得る。縮重を利用して核酸配列を変え、それぞれが実質的に同じタンパク質をコードする様々な核酸配列を作製できることを理解されたい。

【0073】

ポリペプチド配列に用いられる用語「パーセントの同一性」または「%の同一性」とは、標準化されたアルゴリズムを用いてアラインメントされる2つ以上のポリペプチド配列間の一致する残基の百分率のことである。ポリペプチド配列アラインメントの方法は周知である。アラインメント方法の中には、保存的なアミノ酸置換を考慮したものもある。詳細に上述したこのような保存的な置換は、一般に、置換部位の電荷や疎水性が保存され、ポリペプチドの構造(従って機能も)が保存される。

【0074】

ポリペプチド配列間の同一性のパーセントは、MEGALIGN バージョン3.12e 配列アラインメントプログラム(上記)に組込まれるCLUSTAL Vアルゴリズムのデフォルトパラメータを用いて決定可能である。CLUSTAL Vを用いる対方式のポリペプチド配列のアライメントの場合、デフォルトパラメータは、Ktuple = 1、gap penalty = 3、window = 5、及び「diagonals saved」

= 5 と設定する。P A M 2 5 0 マトリクスが、デフォルトの残基重み付け表として選択される。ポリヌクレオチドアラインメントと同様に、アラインメントされたポリペプチド配列の対の同一性のパーセントは、「類似性のパーセント」として C L U S T A L V によって報告される。

【 0 0 7 5 】

別法では、N C B I B L A S T ソフトウェア一式が用いられる。例えば、2 つのポリペプチド配列を対で比較をする場合、ある者は、デフォルトパラメータで設定された「B L A S T 2 S e q u e n c e s」ツール V e r s i o n 2 . 0 . 1 2 (A p r - 2 1 - 2 0 0 0) で b l a s t p を使用するであろう。そのようなデフォルトパラメータは、例えば、以下のようにする。

【 0 0 7 6 】

M a t r i x : B L O S U M 6 2
 O p e n G a p : 1 1 及 び E x t e n s i o n G a p : 1 p e n a l t i e s
 G a p x d r o p - o f f : 5 0
 E x p e c t : 1 0
 W o r d S i z e : 3
 F i l t e r : o n

同一性のパーセントは、例えば、特定の配列番号で決められた、所定のポリペプチド配列の全長に対して測定してもよいし、それより短い長さに対して、例えば、ある大きな所定のポリペプチド配列から得られた断片、例えば、連続する少なくとも 1 5、2 0 または 3 0、4 0、5 0、7 0、1 5 0 の残基の断片の長さに対して測定してもよい。このような長さは単なる例であり、配列表及び表、図面を含む明細書に記載の配列の任意の長さの断片を用いて、同一性のパーセントが測定される長さを示すことができる。

【 0 0 7 7 】

「ヒト人工染色体 (H A C)」は、約 6 k b (キロベース) ~ 1 0 M b のサイズの D N A 配列を含み得る、安定した有糸分裂染色体の分離及び維持に必要な全てのエレメントを含む直鎖状の小染色体である。

【 0 0 7 8 】

用語「ヒト化抗体」は、もとの結合能力を保持しつつよりヒトの抗体に似せるために、非抗原結合領域のアミノ酸配列が変えられた抗体分子を指す。

【 0 0 7 9 】

「ハイブリダイゼーション」とは、所定のハイブリダイゼーション条件下で、ある一本鎖ポリヌクレオチドがある相補的な一本鎖と塩基対を形成するアニーリングのプロセスである。特異的なハイブリダイゼーションとは、2 つの核酸配列が高い相同性を有することを意味する。アニーリングが許容される条件下で、特異的なハイブリダイゼーション複合体が形成され、洗浄過程の後もハイブリダイズしたままである。洗浄過程は、ハイブリダイゼーションプロセスの厳密性即ちストリンジェンシー (s t r i n g e n c y) の決定において特に重要であり、よりストリンジェントな条件では、非特異的な結合、即ち完全には一致しない核酸鎖間の対の結合が減少する。核酸配列間のアニーリングが許容される条件は、当業者によって日常的に決定され、ハイブリダイゼーションの間は一定であるが、洗浄過程は、目的のストリンジェンシーにするためにその最中に条件の変更が可能であり、ハイブリダイゼーション特異性が得られる。アニーリングが許容される条件は、例えば、温度が 6 8 °C で、約 6 × S S C、約 1 % (w / v) の S D S、並びに約 1 0 0 μ g / m l のせん断して変性したサケ精子 D N A が含まれる。

【 0 0 8 0 】

一般に、ハイブリダイゼーションのストリンジェンシーは、洗浄過程を行う際の温度によっても左右される。この洗浄温度は通常、所定のイオン強度と p H における特定の配列の熱融点 (T m) より約 5 ~ 2 0 °C 低く選択される。この T m は、(所定のイオン強度と p H の下) 標的の配列の 5 0 % が完全に一致するプローブとハイブリダイズする温度である。T m を計算する式及び核酸のハイブリダイゼーションの条件は、周知であり、S a m b

10

20

30

40

50

rook, J. 他による, 1989, Molecular Cloning: A Laboratory Manual, 第2版の1-3巻, Cold Spring Harbor Press, Plainview NY; 特に2巻の9章に記載されている。

【0081】

本発明のポリヌクレオチド間の高いストリンジェンシーのハイブリダイゼーションでは、約0.2 x SSC及び約1%のSDSの存在の下、約68 で1時間の洗浄過程を含む。別法では、65、60、55、42 の温度で行う。SSCの濃度は、約0.1%のSDSが存在の下、約0.1~2 x SSCの範囲である。通常は、ブロッキング試薬を用いて非特異的なハイブリダイゼーションを阻止する。このようなブロッキング試薬には、例えば、約100~200 µg/mlの切断され変性したサケ精子DNAが含まれる。約35~50% v/vの濃度のホルムアミドなどの有機溶剤が、例えば、RNAとDNAのハイブリダイゼーションなどの特定の場合に用いることができる。これらの洗浄条件の有用な改変は、当業者には周知である。特に高いストリンジェントな条件でのハイブリダイゼーションは、ヌクレオチド間の進化における類似性を示唆し得る。このような類似性は、それらのヌクレオチド及びコードされたポリペプチドが類似の役割を果たしていることを強く示唆する。

【0082】

用語「ハイブリダイゼーション複合体」は、相補的な塩基間の水素結合によって、形成された2つの核酸配列の複合体を指す。ハイブリダイゼーション複合体は溶液中（例えば、C₀tまたはR₀t分析）で形成されるか、或いは溶液中の1つの核酸配列と固体の支持物（例えば、紙、膜、フィルター、チップ、ピン、或いはスライドガラス、または細胞及びその核酸を固定する任意の適当な基板）に固定されたもう一つの核酸配列とで形成され得る。

【0083】

用語「挿入」或いは「付加」は、1個以上のアミノ酸残基或いはヌクレオチドがそれぞれ追加されるアミノ酸配列或いは核酸配列の変化を指す。

【0084】

「免疫応答」は、炎症性疾患及び外傷、免疫異常、感染症、遺伝病などに関連する症状を指す。これらの症状は、細胞系及び全身防衛系に影響を及ぼすサイトカイン及びケモカイン、別の情報伝達分子などの様々な因子の発現という特徴をもつ。

【0085】

用語「免疫原性断片」は、例えば哺乳動物などの生きている動物に導入すると、免疫反応を引き起こすGVREDのポリペプチド断片またはオリゴペプチド断片を指す。用語「免疫原性断片」はまた、本明細書で開示するまたは当分野で周知のあらゆる抗体生産方法に有用なPRTSのポリペプチド断片またはオリゴペプチド断片を含む。

【0086】

用語「マイクロアレイ」は、基質上の複数のポリヌクレオチド、ポリペプチドまたはその他の化合物の構成を指す。

【0087】

用語「エレメント」または「アレイエレメント」は、マイクロアレイ上に固有の指定された位置を有する、ポリヌクレオチド、ポリペプチドまたはその他の化合物を指す。

【0088】

用語「調節」は、PRTSの活性の変化を指す。例えば、調節によって、PRTSのタンパク質活性、或いは結合特性、またはその他の生物学的特性、機能的特性或いは免疫学的特性の変化が起こる。

【0089】

用語「核酸」及び「核酸配列」は、ヌクレオチド、オリゴヌクレオチド、ポリヌクレオチド、或いはそれらの断片を指し、一本鎖若しくは二本鎖であって、センス鎖或いはアンチセンス鎖であるゲノム起源若しくは合成起源のDNA或いはRNA、ペプチド核酸(PNA)、任意のDNA様物質、及びRNA様物質である。

10

20

30

40

50

【0090】

「機能的に結合した」は、第1の核酸配列と第2の核酸配列が機能的な関係にある状態を指す。例えば、プロモーターがコード配列の転写または発現に影響を与える場合、そのプロモーターはそのコード配列に機能的に結合している。一般に、機能的に結合したDNA配列は、同じ読み枠内で2つのタンパク質をコードする領域が結合する必要がある場合は、非常に近接或いは連続する。

【0091】

「ペプチド核酸(PNA)」は、末端がリシンで終わるアミノ酸残基のペプチド骨格に結合した、少なくとも約5ヌクレオチドの長さのオリゴヌクレオチドを含む、アンチセンス分子または抗遺伝子剤を指す。この末端のリシンにより、この組成物が溶解性となる。PNAは、相補的な一本鎖DNAやRNAに優先的に結合して転写物の伸長を止め、ポリエチレングリコール化して細胞における寿命を延ばし得る。

【0092】

PRTSの「翻訳後修飾」には、脂質化、グリコシル化、リン酸化、アセチル化、ラセミ化、蛋白分解性切断及びその他の当分野で既知の修飾を含まれ得る。これらのプロセスは、合成或いは生化学的に生じ得る。生化学的修飾は、PRTSの酵素環境に依存し、細胞の種類によって異なり得る。

【0093】

「プローブ」とは、同一配列或いはアレル核酸配列、関連する核酸配列の検出に用いる、PRTSやそれらの相補配列、またはそれらの断片をコードする核酸配列のことである。プローブは、検出可能な標識またはレポーター分子が結合され単離されたオリゴヌクレオチドやポリヌクレオチドである。典型的な標識には、放射性アイソトープ及びリガンド、化学発光試薬、酵素がある。「プライマー」とは、相補的な塩基対を形成して標的のポリヌクレオチドにアニーリング可能な、通常はDNAオリゴヌクレオチドである短い核酸である。プライマーがポリヌクレオチドにアニーリングした後、あるDNAポリメラーゼ酵素によって、標的のDNA一本鎖に沿って伸長される。プライマーの組は、例えば、PCR法における核酸配列の増幅(及び同定)に用いることができる。

【0094】

本発明に用いられるプローブ及びプライマーは、既知の配列の少なくとも15の連続するヌクレオチドを含む。特異性を高めるために、より長いプローブ及びプライマーが用いることも可能である。例えば、開示した核酸配列の連続する少なくとも20または25、30、40、50、60、70、80、90、100、150のヌクレオチドを含む。プローブ及びプライマーは、上記した例より相当長いものも用いることができ、本明細書の表及び図面、配列表に示された任意の長さのヌクレオチドも用いることができることを理解されたい。

【0095】

プローブ及びプライマーの準備及び使用方法については、例えば、Sambrook, J. 他による、1989年、名称「Molecular Cloning: A Laboratory Manual」、第2版の1-3巻(Cold Spring Harbor Press, Plainview NY)、またはAusubel, F.M. 他による、1987年、名称「Current Protocols in Molecular Biology」(Greene Publi. Assoc. & Wiley-Intersciences, New York NY)、並びにInnis他による、1990年、名称「PCR Protocols, A Guide to Methods and Applications」(Academic Press, San Diego CA.)を参照されたい。PCR用のプライマーの組は、例えば、Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA)などのそのような目的のためのコンピュータプログラムを用いて、ある既知の配列から引き出すことができる。

【0096】

10

20

30

40

50

プライマーとして用いるオリゴヌクレオチドは、当分野で周知のプライマー選択用のコンピュータプログラムで選択される。例えば、OLIGO 4.06ソフトウェアは、それぞれが最大100ヌクレオチドまでのPCR用のプライマーの対の選択、及び32,000塩基までの入力ポリヌクレオチド配列から最大5,000ヌクレオチドまでの大きなポリヌクレオチド及びオリゴヌクレオチドの分析に有用である。類似のプライマー選択用プログラムには、能力を拡大する追加の機能が含まれている。例えば、PrimOUプライマー選択プログラム(Genome Center at University of Texas South West Medical Center, Dallas TXより入手可能)は、メガベース配列から特定のプライマーを選択できるため、ゲノムワイドスコープ(genome-wide scope)におけるプライマーの設計に有用である。Primer3プライマー選択プログラム(Whitehead Institute/MIT Center for Genome Research, Cambridge MAより入手可能)によって、ユーザーは、プライマー結合部位として避けたい配列を指定できる「非プライミングライブラリ(mispriming library)」を入力できる。また、Primer3は、特にマイクロアレイのオリゴヌクレオチドの選択に有用である(後の方の2つのプライマー選択プログラムのソースコードは、それぞれのソースから得ることができ、ユーザーのニーズを満たすように変更することもできる)。PrimerGenプログラム(UK Human Genome Mapping Project Resource Centre, Cambridge UKより入手可能)は、多数の配列アラインメントに基づいてプライマーを設計するため、アラインメントされた核酸配列の最も保存された領域或いは最も保存されていない領域のどちらかとハイブリダイズするプライマーを選択することができる。従って、このプログラムは、固有及び保存されたオリゴヌクレオチドやポリヌクレオチドの断片の同定に有用である。上記した任意の選択方法で同定されたオリゴヌクレオチドやポリヌクレオチドの断片は、例えば、PCR法やシーケンシングプライマー、マイクロアレイエレメント、或いはサンプルの核酸の完全或いは部分的に相補的なポリヌクレオチドを同定する特定のプローブなどの、ハイブリダイゼーション技術に有用である。オリゴヌクレオチドの選択方法は、上記した方法に制限されるものではない。

【0097】

本明細書における「組換え核酸」は天然の配列ではなく、2つ以上の配列の離れたセグメントを人工的に組み合わせた配列である。この人工の組み合わせは、化学合成によって作られる場合も多いが、前出のSambrookに記載されたような遺伝子工学の技術を用いて核酸の離れたセグメントを人工的に操作する方がより一般的である。この「組換え核酸」には、単に核酸の一部の追加または置換、欠失によって変更された核酸も含む。組換え核酸は、あるプロモーター配列に機能的に結合した核酸配列を含む場合もある。このような組換え核酸は、例えば、ある細胞を形質転換するのに用いられるベクターの一部であり得る。

【0098】

別法では、このような組換え核酸は、この組換え核酸を発現する哺乳動物のワクチン接種に用いると、その哺乳動物の防衛的な免疫応答を誘発する、ワクシニアウイルスに基づいたウイルスベクターの一部であり得る。

【0099】

「調節エレメント」は、通常は遺伝子の非翻訳領域に由来する核酸配列であり、エンハンサー、プロモーター、イントロン及び5'及び3'の非翻訳領域(UTR)を含む。調節エレメントは、転写や翻訳、またはRNAの安定性を調節する宿主またはウイルスタンパク質と相互作用する。

【0100】

「レポーター分子」は、核酸、アミノ酸または抗体の標識に用いられる化学的または生化学的部分である。レポーター分子には、放射性核種、酵素、蛍光剤、化学発光剤、発色剤、基質、補助因子、インヒビター、磁気粒子及びその他の当分野で既知の成分が含まれ

る。

【0101】

本明細書において、DNA配列に対する「RNA等価物」とは、基準となるDNA配列と同じ直鎖の核酸配列から構成されるが、窒素性塩基のチミンがウラシルに置換され、糖鎖の背骨がデオキシリボースではなくリボースからなる。

【0102】

用語「サンプル」は、その最も広い意味で用いられている。PRTS、PRTSをコードする核酸、またはその断片を含むと推定されるサンプルは、体液と、細胞からの抽出物や細胞から単離された染色体や細胞内小器官、膜と、細胞と、溶液中に存在するまたは基板に固定されたゲノムDNA、RNA、cDNAと、組織と、組織プリント等を含み得る。

10

【0103】

用語「特異的結合」及び「特異的に結合する」は、タンパク質若しくはペプチドと、アゴニスト、抗体、アンタゴニスト、小分子、若しくは任意の天然若しくは合成の結合組成物との間の相互作用を指す。この相互作用は、結合する分子によって認識される、例えば、抗原決定基つまりエピトープなどのタンパク質の特定の構造の存在によって左右される。例えば、抗体がエピトープ「A」に対して特異的である場合、結合していない標識した「A」及び抗体を含む反応液に、エピトープAを含むポリペプチド或いは結合していない無標識の「A」が存在すると、抗体と結合する標識Aの量が減少する。

【0104】

用語「実質的に精製された」は、自然の環境から取り除かれてから、単離或いは分離された核酸配列或いはアミノ酸配列であって、自然に結合している組成物が少なくとも約60%除去されたものであり、好ましくは約75%以上の除去、最も好ましくは90%以上除去されたものを指す。

20

【0105】

「置換」とは、一つ以上のアミノ酸またはヌクレオチドをそれぞれ別のアミノ酸またはヌクレオチドに置き換えることである。

【0106】

用語「基板」は、任意の好適な固体或いは半固体の支持物を指し、膜及びフィルター、チップ、スライド、ウエハ、ファイバー、磁気または非磁気ビード、ゲル、チューブ、プレート、ポリマー、微小粒子、毛細管が含まれる。この基板には、壁または塹壕、ピン、チャンネル、細孔などの様々な表面形態があり、そこにポリヌクレオチドやポリペプチドが結合する。

30

【0107】

「転写イメージ」は、所定条件下での所定時間における特定の細胞の種類または組織による集合的遺伝子発現のパターンを指す。

【0108】

「形質転換」とは、外来DNAが受容細胞に導入されるプロセスのことである。形質転換は、当分野で周知の種々の方法により、自然或いは人工の条件下で起こり、原核宿主細胞若しくは真核宿主細胞の中に外来核酸配列を挿入する任意の周知の方法によって行うことができる。この形質転換の方法は、形質転換される宿主細胞のタイプによって選択される。この方法には、バクテリオファージまたはウイルス感染、電気穿孔法（エレクトロポレーション）、リポフェクション、及び微粒照射が含まれるが、これらに限定されるものではない。「形質転換された」細胞には、導入されたDNAが自律的に複製するプラスミドとして或いは宿主染色体の一部として複製可能である安定的に形質転換された細胞が含まれる。さらに、限られた時間に一時的に導入DNA若しくは導入RNAを発現する細胞も含まれる。

40

【0109】

本明細書における「遺伝子組換え生物」とは、当分野で周知の遺伝子組換え技術などを用いて、人間が生物の1つ以上の細胞に異種の核酸を導入した任意の生物であり、動物及び植物を含むが、それらに限定されるものではない。微量注入や組換えウイルスに感染させ

50

るなどの慎重な遺伝子操作によって、細胞の前駆体に直接或いは間接的に異種核酸を細胞に導入する。「遺伝子操作」とは、典型的な交雑育種や *in vitro* での受精ではなく、組換え DNA 分子を導入することである。本発明に従った遺伝子組換え生物には、細菌及びラン藻類、菌類、植物、動物が含まれる。本発明の単離された DNA は、当分野で周知の、例えば、感染、形質移入、形質転換、トランス接合 (trans conjugation) などの方法によって、宿主に導入することができる。本発明の DNA をそのような生物に導入する技術は周知であり、前出の Sambrook 他 (1989) に記載されている。

【0110】

特定の核酸配列の「変異配列」とは、デフォルトパラメータ設定の「BLAST 2 Sequences」ツール Version 2.0.9 (May - 07 - 1999) を用いる *blastn* によって、ある核酸配列のある長さに対する該特定の核酸配列の同一性が、少なくとも 40% と決定された核酸配列のことである。このような核酸の対は、ある長さにおいて、例えば、少なくとも 50% または 60%、70%、80%、85%、90%、95%、98%、或いはそれ以上の同一性を示し得る。ある変異配列は、例えば、「アレル」変異配列 (上述) または「スプライス」変異配列、「種」変異配列、「多型」変異配列と表すことができる。スプライス変異配列は基準分子と同一性が極めて高い可能性があるが、mRNA プロセッシング中のエキソンの択一的スプライシングによってポリヌクレオチドの数が多くなったり、少なくなったりする。対応するポリペプチドは、基準分子に存在する追加の機能ドメインを有したり、基準分子に存在するドメインが欠落したりし得る。種変異配列は、種によって異なるポリヌクレオチド配列である。得られるポリペプチドは、互いに高いアミノ酸同一性を有する。多型変異配列は、所定の種と種における特定の遺伝子のポリヌクレオチド配列が異なる。多型変異配列はまた、ポリヌクレオチド配列の 1 つのヌクレオチドが異なる「1ヌクレオチド多型」(SNP) も含み得る。SNP の存在は、例えば、或る集団、病態、病態の性向を示唆し得る。

【0111】

特定のポリペプチド配列の「変異体」とは、デフォルトパラメータ設定の「BLAST 2 Sequences」ツール Version 2.0.9 (May - 07 - 1999) を用いる *blastp* によって、ある核酸配列のある長さに対する該特定のポリペプチド配列の同一性が、少なくとも 40% と決定されたポリペプチド配列のことである。このようなポリペプチドの対は、ある長さにおいて、例えば、少なくとも 50% または 60%、70%、80%、85%、90%、95%、98%、或いはそれ以上の同一性を示し得る。

【0112】

(発明)

本発明は、新規のヒトプロテアーゼ (PRTS) 及び PRTS をコードするポリヌクレオチドの発見に基づき、これらの組成物を利用した胃腸疾患、心血管疾患、自己免疫/炎症性の疾患、増殖異常、発生障害、上皮の疾患、神経の疾患、および生殖障害の診断、治療、及び予防に関する。

【0113】

表 1 は、本発明のポリヌクレオチド配列及びポリペプチド配列の識別番号を示す。各ポリヌクレオチドおよびそれに対応するポリペプチドは、1 つのインサイトプロジェクト識別番号 (Incyte Project ID) に相関する。各ポリペプチド配列は、記載されているようにポリペプチド配列識別番号 (Polypeptide SEQ ID NO :) およびインサイトポリペプチド配列番号 (Incyte Polypeptide ID) の両方によって示されている。各ポリヌクレオチド配列は、記載されているようにポリヌクレオチド配列識別番号 (Polypeptide SEQ ID NO :) およびインサイトポリヌクレオチドコンセンサス配列番号 (Incyte Polypeptide ID) の両方によって示されている。

【0114】

10

20

30

40

50

表2は、GenBankタンパク質 (genept) データベースにおいてBLAST解析により同定された本発明のポリペプチドに相同性を有する配列を示す。列1および列2はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号 (Polypeptide SEQ ID NO :) およびそれに対応するインサイトポリペプチド配列番号 (Incyte Polypeptide ID) を示す。列3は、GenBankの最も近い相同体のGenBankの識別番号 (Genbank ID NO :) を示す。列4は、各ポリペプチドとそのGenBank相同体との間の一致を表す確率スコアを示す。列5は、GwnBank相同体のアノテーションを示し、更に該当箇所には適当な引用文も示す。これらを引用することを以って本明細書の一部とする。

【0115】

表3は、本発明の各ポリペプチドの様々な構造的特徴を示す。列1および列2はそれぞれ、本発明の各ポリペプチドのポリペプチド配列識別番号 (SEQ ID NO :) およびそれに対応するインサイトポリペプチド配列番号 (Incyte Polypeptide ID) を示す。列3は、各ポリペプチドのアミノ酸残基数を示す。列4および列5はそれぞれ、GCG配列分析ソフトウェアパッケージのMOTIFSプログラム (Genetics Computer Group, Madison WI) によって決定された、潜在的なリン酸化部位および潜在的なグリコシル化部位を示す。列6は、シグネチャ (signature) 配列、ドメイン、およびモチーフを含むアミノ酸残基を示す。列7は、タンパク質の構造/機能の分析のための分析方法を示し、該当箇所にはさらに分析方法に利用した検索可能なデータベースを示す。

【0116】

表4に示されているように、本発明のポリヌクレオチド配列は、cDNA配列、またはゲノムDNA由来のコード (エキソン) 配列、或いはこれらの2種類の配列のあらゆる組み合わせを用いて組み立てた。列1および列2はそれぞれ、本発明の各ポリヌクレオチドのポリヌクレオチド配列識別番号 (Polynucleotide SEQ ID NO :) およびそれに対応するインサイトポリヌクレオチドコンセンサス配列番号 (Incyte Polynucleotide ID) を示す。列3は、塩基対における各ポリヌクレオチド配列の長さを示す。列4は、例えば、SEQ ID NO : 15 - 28を同定するため、或いはSEQ ID NO : 15 - 28と関連するポリヌクレオチド配列とを区別するためのハイブリダイゼーションまたは増幅技術に有用なポリヌクレオチド配列の断片を示す。列5は、cDNA配列、ゲノムDNAから推定されるコード配列 (エキソン)、および/またはcDNAおよびゲノムDNAの両方からなる群に対応する識別番号を示す。これらの配列を用いて本発明のポリヌクレオチド配列を組み立てた。表4の列6および列7はそれぞれ、列5の配列に対応するcDNA配列およびゲノム配列の開始ヌクレオチド (5') 位置および終了ヌクレオチド (3') 位置を示す。

【0117】

表4の列5に示されている識別番号は、具体的には、例えばインサイトcDNAおよびそれらに対応するcDNAライブラリの識別番号を示す。例えば、7032724H1はインサイトcDNA配列の識別番号であり、BRAXTR12はそれが由来するcDNAライブラリの識別番号である。cDNAライブラリが示されていないインサイトcDNAは、プールされているcDNAライブラリ (例えば、70152356V1) に由来した。列5の識別番号は、完全長ポリヌクレオチド配列の組み立てに用いたGenBankのcDNAすなわちEST (例えば、g5364348) の識別番号の場合もある。または、列5の識別番号は、Genscan分析によって推定されるゲノムDNAのコード領域の場合もある。例えば、GNN.g6436155_002.editは、Genscan推定コード配列の識別番号であって、g6436155がGenscan分析によって得られたGenBankの配列の識別番号である。このGenscan推定コード配列は、配列を組み立てる前に編集する場合がある (実施例4を参照)。または、列5の識別番号は、"exon-stitching" アルゴリズムによってcDNAおよびGenscan推定エキソンの両方からなる群の場合もある (実施例5を参照)。または、列5の

10

20

30

40

50

識別番号は、"exon-stretching" アルゴリズムによって cDNA および Genscan 推定エキソンの両方からなる群の場合もある(実施例 5 を参照)。場合によっては、列 5 に示されている配列の範囲と重複するインサイト cDNA の範囲が得られ、最終的なコンセンサス配列が決定されるが、それに相当するインサイト cDNA の識別番号は示されていない。

【0118】

表 5 は、インサイト cDNA 配列を用いて組み立てられたこれらのポリヌクレオチド配列が由来する代表的な cDNA ライブラリを示す。代表的な cDNA ライブラリとは、上記ポリヌクレオチド配列の組み立ておよび決定に用いられたインサイト cDNA 配列を最も多く含むインサイト cDNA ライブラリのことである。表 5 に示されている cDNA ライ
10
ブラリを作製するために用いた組織およびベクターが表 6 に示されている。

【0119】

本発明はまた、PRTS の変異体も含む。好適な PRTS の変異体は、PRTS の機能的
或いは構造的特徴の少なくともどちらか一方を有し、かつ PRTS アミノ酸配列に対して
少なくとも約 80% のアミノ酸配列同一性、或いは少なくとも約 90% のアミノ酸配列同
一性、更には少なくとも約 95% のアミノ酸配列同一性を有する。

【0120】

本発明はまた、PRTS をコードするポリヌクレオチドを提供する。特定の実施例におい
て、本発明は、PRTS をコードする SEQ ID NO: 15 - 28 からなる一群から選
10
択された配列を含むポリヌクレオチド配列を提供する。配列表に示した SEQ ID NO
: 15 - 28 のポリヌクレオチド配列は、窒素系塩基のチミンがウラシルに置換され、糖
鎖の背骨がデオキシリボースではなくリボースからなる等価 RNA 配列を含む。

【0121】

本発明はまた、PRTS をコードするポリヌクレオチド配列の変異配列を含む。詳細には
、このようなポリヌクレオチド配列の変異配列は、PRTS をコードするポリヌクレオチ
ド配列と少なくとも 70% のポリヌクレオチド配列同一性、或いは少なくとも 85% のポ
リヌクレオチド配列同一性、更には少なくとも 95% ものポリヌクレオチド配列同一性を
有する。本発明の特定の実施形態は、SEQ ID NO: 15 - 28 からなる一群から選
10
択された核酸配列と少なくとも 70% のポリヌクレオチド配列同一性、或いは少なく
とも 85% のポリヌクレオチド配列同一性、更には少なくとも 95% ものポリヌクレオチド配
列同一性を有する SEQ ID NO: 15 - 28 からなる一群から選択された配列を含む
ポリヌクレオチド配列の変異配列を提供する。上記したポリヌクレオチド変異配列は何れ
も、PRTS の機能的或いは構造的特徴の少なくとも 1 つを有するアミノ酸配列をコード
する。

【0122】

遺伝暗号の縮重により作り出され得る PRTS をコードする種々のポリヌクレオチド配列
には、既知の自然発生する任意の遺伝子のポリヌクレオチド配列と最小の類似性しか有し
ないものも含まれることを、当業者は理解するであろう。したがって本発明には、可能な
コドン選択に基づいた組み合わせの選択によって作り出され得る可能なポリヌクレオチド
10
配列の変異の全てが含まれ得る。これらの組み合わせは、天然の PRTS のポリヌクレオ
チド配列に適用される標準的なトリプレット遺伝暗号を基に作られ、全ての変異が明確に
開示されていると考慮する。

【0123】

PRTS をコードするヌクレオチド配列及びその変異配列は一般に、好適に選択されたス
トリンジェントな条件下で、天然の PRTS のヌクレオチド配列とハイブリダイズ可能で
あるが、非天然のコドンを含めるなどの実質的に異なった使い方のコドンを含む PRTS
或いはその誘導体をコードするヌクレオチド配列を作るとは有利となり得る。特定の
コドンが宿主によって利用される頻度に基づいてコドンを選択して、ペプチドの発現が特
10
定の真核細胞または原核宿主に発生する割合を高めることが可能である。コードされたア
ミノ酸配列を変えないで、PRTS 及びその誘導体をコードするヌクレオチド配列を実質
40
50

的に変更する別の理由は、天然の配列から作られる転写物より例えば長い半減期など好ましい特性を備えるRNA転写物を作ることにある。

【0124】

本発明はまた、PRTS及びその誘導体をコードするDNA配列またはそれらの断片を完全に合成化学によって作り出すことも含む。作製後にこの合成配列を、当分野で良く知られた試薬を用いて、種々の入手可能な発現ベクター及び細胞系の何れの中にも挿入可能である。更に、合成化学を用いて、PRTSまたはその任意の断片をコードする配列の中に突然変異を導入することも可能である。

【0125】

更に本発明には、種々のストリンジェントな条件下で、請求項に記載されたポリヌクレオチド配列、特に、SEQ ID NO: 15 - 28及びそれらの断片とハイブリダイズ可能なポリヌクレオチド配列が含まれる(例えば、Wahl, G. M. 及びS. L. Berger (1987) *Methods Enzymol.* 152: 399 - 407; and Kimmel, A. R. (1987) *Methods Enzymol.* 152: 507 - 511. を参照)。アニーリング及び洗浄条件を含むハイブリダイゼーションの条件は、「定義」に記載されている。

【0126】

当分野で周知のDNAのシーケンシング方法を用いて、本発明の何れの実施例も実行可能である。この方法には、例えばDNAポリメラーゼIのクレノウ断片、SEQUENASE (US Biochemical, Cleveland OH)、Taqポリメラーゼ (Applied Biosystems)、熱安定性T7ポリメラーゼ (Amersham, Pharmacia Biotech Piscataway NJ)、或いはELONGASE増幅システム (Life Technologies, Gaithersburg MD) にみられるような校正エキソヌクレアーゼとポリメラーゼとの組み合わせなどの酵素が用いられる。好ましくは、MICROLAB2200液体転移システム (Hamilton, Reno, NV)、PTC200 Thermal Cycler 200 (MJ Research, Watertown MA) 及びABI CATALYST 800 (PE Biosystems) などの装置を用いて配列の準備を自動化する。次に、ABI 373 或いは377 DNAシーケンシングシステム (PE Biosystems)、MEGABACE 1000 DNAシーケンシングシステム (Molecular Dynamics, Sunnyvale CA) または当分野で周知の他の方法を用いてシーケンシングを行う。得られた配列を当分野で周知の様々なアルゴリズムを用いて分析する(例えば、Ausubel, F. M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R. A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856 - 853. を参照)。

【0127】

当分野で周知のPCR法をベースにした種々の方法で、部分的なヌクレオチド配列を利用して、PRTSをコードする核酸配列を伸長し、プロモーターや調節エレメントなどの上流にある配列を検出する。例えば制限部位PCR法を利用する1つの方法では、一般的なプライマー及び入れ子プライマー (nested primer) を用いてクローニングベクター内のゲノムDNAから未知の配列を増幅する(例えば、Sarkar, G. (1993) *PCR Methods Applic* 2: 318 - 322 を参照)。逆PCR法を用いる別法では、広範な方向に伸長して環状化した鋳型から未知の配列を増幅するプライマーを用いる。この鋳型は、既知のゲノム遺伝子座及びその周辺の配列を含む制限断片に由来する(例えば、Triglia, T. ら (1988) *Nucleic Acids Res* 16: 8186 を参照)。キャプチャPCR法を用いる第3の方法は、ヒト及び酵母菌人工染色体DNAの既知の配列に隣接するDNA断片のPCR増幅を含む(例えば、Lagerstrom, M. 他 (1991) *PCR Methods Appli* 40 50

c 1 : 1 1 1 - 1 1 9 を参照)。この方法では、多数の制限酵素による消化及びライゲーションを用いて、PCRを行う前に未知の配列の領域の中に組換え二本鎖配列を挿入することが可能である。また、当分野で周知の別の方法を用いて未知の配列を得ることも可能である。(例えば、Parker, J. D. 他 (1991) *Nucleic Acids Res.* 19:3055-3060を参照)。更に、PCR、ネスト化プライマー、PROMOTERFINDERライブラリ(Clontech, Palo Alto CA)を用いれば、ゲノムDNA内の歩行が可能である。この方法ではライブラリをスクリーニングする必要がなく、イントロン/エキソン接合部を探すのに有用である。全てのPCR法をベースにした方法では、プライマーは、市販のOLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) 或いは別の好適なプログラムなどを用いて、長さが22~30ヌクレオチド、GC含量が50%以上、約68~72の温度で鋳型に対してアニーリングするよう設計される。

【0128】

完全長のcDNAをスクリーニングする場合は、大きなcDNAを含むようにサイズが選択されたライブラリを用いるのが好ましい。更に、オリゴd(T)ライブラリが完全な長さのcDNAを産生できない場合は、遺伝子の5'領域を有する配列を含むものが多いランダムに初回抗原刺激を受けたライブラリが有用である。ゲノムライブラリは、5'非転写調節領域への配列の伸長に有用であろう。

【0129】

市販のキャピラリー電気泳動システムを用いて、シーケンシングまたはPCR産物のヌクレオチド配列のサイズの分析、または確認が可能である。詳しくは、キャピラリーシーケンシングには、電気泳動による分離のための流動性ポリマー、及び4つの異なるヌクレオチドに特異的なレーザーで活性化される蛍光色素、放出された波長の検出に利用するCCDカメラを使用することが可能である。出力/光強度は、適切なソフトウェア(例えば、GENOTYPER及びSEQUENCE NAVIGATOR、PE Biosystems)を用いて電気信号に変換され、サンプルのローディングからコンピュータ分析までのプロセス及び電子データ表示がコンピュータ制御可能である。キャピラリー電気泳動法は、特定のサンプルに少量しか存在しない場合もあるDNAの小片のシーケンシングに特に適している。

【0130】

本発明の別の実施例では、PRTSをコードするポリヌクレオチド配列またはその断片を組換えDNA分子にクローニングして、適切な宿主細胞内にPRTS、その断片または機能的等価物を発現させることが可能である。遺伝暗号固有の縮重により、実質的に同じ或いは機能的に等価のアミノ酸配列をコードする別のDNA配列が作られ得り、これらの配列をPRTSのクローン化及び発現に利用可能である。

【0131】

種々の目的でPRTSをコードする配列を変えるために、当分野で一般的に知られている方法を用いて、本発明のヌクレオチド配列を組換えることができる。この目的には、遺伝子産物のクローン化、プロセッシング及び/または発現の調節が含まれるが、これらに限定されるものではない。ランダムな断片によるDNAの混合や遺伝子断片と合成オリゴヌクレオチドのPCR再組み立てを用いて、ヌクレオチド配列の組換えが可能である。例えば、オリゴヌクレオチド媒介性定方向突然変異誘発を利用して、新しい制限部位を生成する突然変異の導入、グリコシル化パターンの変更、コドン選択の変更、スプライスバリアントの作製等が可能である。

【0132】

本発明のヌクレオチドを、MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; 米国特許第5,837,458号; Chang, C.-C. 他 (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. 他 (1999) *Nat. Biotechno*

10

20

30

40

50

l. 17:259-264; Cramer i, A. 他 (1996) Nat. Biotechnol. 14:315-319) などのDNAシャフリング技術を用いてシャフリングして、PRTSの生物学的または酵素的な活性、或いは他の分子や化合物と結合する能力などのPRTSの生物学的特性を変更或いは改良することができる。DNAシャフリングは、PCR法による遺伝子断片の組換えで遺伝子変異体のライブラリを作製するプロセスである。次に、このライブラリを、目的の特性を有する遺伝子変異体を同定するために選択或いはスクリーニングする。これらの好ましい変異体をプールし、DNAシャフリング及び選択/スクリーニングを繰り返す。従って、人工的な育種及び急速な分子の進化によって多様な遺伝子が作られる。例えば、ランダムな位置に変異がある1つの遺伝子の断片を、目的の特性が最適化するまで、組換え及びスクリーニング、シャフリングを実施することもできる。別法では、所定の遺伝子の断片を、同じ或いは異なった種の同じ遺伝子ファミリーの相同な遺伝子の断片で組換え、それによってプロトコルに従った調節可能な方法で、多数の天然遺伝子の遺伝子多様性を最大にすることができる。

【0133】

別の実施例によれば、PRTSをコードする配列は、当分野で周知の化学的方法を用いて、全体或いは一部が合成可能である(例えば、Caruthers, M. H. ら (1980) Nucl. Acids Res. Symp. Ser 7:215-223; 及び Horn, T. 他 (1980) Nucl. Acids Res. Symp. Ser. 225-232を参照)。別法として、化学的方法を用いてPRTS自体またはその断片を合成することが可能である。例えば、ペプチド合成は種々の固相技術を用いて実行可能である(例えば、Creighton, T. (1984) Proteins. Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J. Y. ら (1995) Science 269:202-204を参照)。また、合成の自動化は例えばABI 431Aペプチドシンセサイザー(PE Biosystems)を用いて達成し得る。更にPRTSのアミノ酸配列または任意のその一部は、直接的な合成の際の変更、及び/または化学的方法を用いた他のタンパク質または任意のその一部からの配列との組み合わせにより、天然のポリペプチド配列を有するポリペプチドまたは変異体ポリペプチドを作製することが可能である。

【0134】

このペプチドは、分離用高速液体クロマトグラフィー(例えば、Chiez, R. M. 及び F. Z. Regnier (1990) Methods Enzymol. 182:392-421を参照)を用いて実質的に精製可能である。合成されたペプチドの組成は、アミノ酸分析或いはシーケンシングにより確認することができる(例えば、Creighton、前出、pp 28-53を参照)。

【0135】

生物学的に活性なPRTSを発現させるために、PRTSをコードするヌクレオチド配列またはその誘導体を好適な発現ベクターに挿入する。この発現ベクターは、好適な宿主に挿入されたコーディング配列の転写及び翻訳の調節に必要なエレメントを含む。これらのエレメントには、ベクター及びPRTSをコードするポリヌクレオチド配列におけるエンハンサー、構成型及び発現誘導型のプロモーター、5'及び3'の非翻訳領域などの調節配列が含まれる。このようなエレメントは、その長さ及び特異性が様々である。特定の開始シグナルによって、PRTSをコードする配列のより効果的な翻訳を達成することが可能である。このようなシグナルには、ATG開始コドン及びコザック配列などの近傍の配列が含まれる。PRTSをコードする配列及びその開始コドン、上流の調節配列が好適な発現ベクターに挿入された場合は、更なる転写調節シグナルや翻訳調節シグナルは必要なくなるであろう。しかしながら、コーディング配列或いはその断片のみが挿入された場合は、インフレームのATG開始コドンを含む外来性の翻訳調節シグナルが発現ベクターに含まれなければならない。外来性の翻訳要素及び開始コドンは、自然及び合成の様々なものから得ることが可能である。用いられる特定の宿主細胞系に好適なエンハンサーを含め

ることによって発現の効率を高めることが可能である。(例えば、Scharf, D. 他 (1994) *Results Probl. Cell Differ.* 201-18-162. を参照)。

【0136】

当業者に周知の方法を用いて、PRTSをコードする配列、好適な転写及び翻訳調節エレメントを含む発現ベクターを作製することが可能である。これらの方法には、in vitro組換えDNA技術、合成技術、及びin vivo遺伝子組換え技術が含まれる。(例えば、Sambrook, J. 他. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, 4章及び8章, 及び16-17章; 及び Ausubel, F.M. 他. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9章及び13章1-4章を参照)。

【0137】

種々の発現ベクター/宿主系を利用して、PRTSをコードする配列の保持及び発現が可能である。これらには、限定するものではないが、組換えバクテリオファージ、プラスミド、またはコスミドDNA発現ベクターで形質転換された細菌などの微生物や、酵母菌発現ベクターで形質転換された酵母菌や、ウイルス発現ベクター(例えば、バキュロウイルス)に感染した昆虫細胞系や、ウイルス発現ベクター(例えば、カリフラワーモザイクウイルス、CaMV; タバコモザイクウイルス、TMV)または細菌発現ベクター(例えば、TiまたはpBR322プラスミド)で形質転換された植物細胞系や、動物細胞系などが含まれる(例えば、前出のSambrook、前出のAusubel、Van Heeke, G. および S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509、Engelhard, E.K. 他 (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227、Sandig, V. 他 (1996) *Hum. Gene Ther.* 7:1937-1945、Takamatsu, N. (1987) *EMBOJ.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196、Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659、Harrington, J.J. 他 (1997) *Nat. Genet.* 15:345-355を参照)。レトロウイルス、アデノウイルス、ヘルペスウイルスまたはワクシニアウイルス由来の発現ベクター、または種々の細菌性プラスミド由来の発現ベクターを用いて、ヌクレオチド配列を標的器官、組織または細胞集団へ輸送することができる(Di Nicola, M. 他 (1998) *Cancer Gen. Ther.* 5(6):350-356、Yu, M. 他 (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344、Buller, R.M. 他 (1985) *Nature* 317(6040):813-815; McGregor, D.P. 他 (1994) *Mol. Immunol.* 31(3):219-226、Verma, I.M. and N. Somia (1997) *Nature* 389:239-242等を参照)。本発明は使用される宿主細胞によって限定されるものではない。

【0138】

細菌系では、多数のクローニングベクター及び発現ベクターが、PRTSをコードするポリヌクレオチド配列の使用目的に応じて選択可能である。例えば、PRTSをコードするポリヌクレオチド配列の日常的なクローニング、サブクローニング、増殖には、PBLUESCRIPT (Stratagene, La Jolla CA) または pSPORT1 プラスミド (GIBCO BRL) などの多機能の大腸菌ベクターを用いることができる。ベクターの多数のクローニング部位にPRTSをコードする配列をライゲーションするとlacZ遺伝子が破壊され、組換え分子を含む形質転換された細菌の同定のための比

色スクリーニング法が可能となる。更に、これらのベクターを用いて、クローニングされた配列の *in vitro* での転写、ジデオキシンスクリーニング、ヘルパーファージによる一本鎖の救出、入れ子状態の欠失を作り出すことが可能である（例えば、Van Heeke, G. 及び S. M. Schuster (1989) *J. Biol. Chem.* 264: 5503 - 5509. を参照)。例えば、抗体の産生のためなどに多量の PRTS が必要な場合は、PRTS の発現をハイレベルで誘導するベクターが使用できる。例えば、強力に発現を誘発する T5 または T7 バクテリオファージプロモーターを含むベクターを使用できる。

【0139】

PRTS の発現に酵母の発現系の使用が可能である。因子やアルコールオキシダーゼや P GH プロモーターなどの構成型或いは誘導型のプロモーターを含む多種のベクターが、酵母菌 *Saccharomyces cerevisiae* または *Pichia pastoris* に使用可能である。更に、このようなベクターは、発現したタンパク質の分泌か細胞内への保持のどちらかを誘導し、安定した増殖のために宿主ゲノムの中に外来配列を組み込む。（例えば、Ausubel, 1995, 前出、Bitter, G. A. ら (1987) *Methods Enzymol.* 153: 516 - 544、及び Scorer, C. A. ら (1994) *Bio/Technology* 121: 181 - 184. を参照）。

【0140】

植物系も PRTS の発現に使用可能である。PRTS をコードする配列の転写は、例えば、CaMV 由来の 35S 及び 19S プロモーターなどのウイルスプロモーターが単独で、或いは TMV（例えば、Coruzzi, G. ら (1984) *EMBO J.* 3: 1671 - 1680; Broglie, R. ら (1984) *Science* 224: 838 - 843; および Winter, J. ら (1991) *Results Probl. Cell Differ.* 17: 85 - 105 を参照) 由来のオメガリーダー配列と組み合わせることで促進される。これらの作製物は、直接の DNA 形質転換或いは病原体を介したトランスフェクションによって、植物細胞の中に導入可能である。（例えば、The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill NY, pp. 191 - 196 を参照）。

【0141】

哺乳動物細胞では、多種のウイルスベースの発現系が利用され得る。アデノウイルスが発現ベクターとして用いられる場合、後発プロモーター及び 3 連リーダー配列からなるアデノウイルス転写物 / 翻訳複合体に PRTS をコードする配列を結合し得る。ウイルスのゲノムの非必須の E1 または E3 領域への挿入により、感染した宿主細胞に PRTS を発現する生ウイルスを得ることが可能である（Logan, J. 及び Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81: 3655 - 3659 を参照）。さらに、ラウス肉腫ウイルス (RSV) エンハンサーなどの転写エンハンサーを用いて、哺乳動物宿主細胞における発現を増大させることが可能である。タンパク質を高レベルで発現させるために、SV40 または EBV を基にしたベクターを用いることが可能である。

【0142】

ヒト人工染色体 (HAC) を用いて、プラスミドで発現しそれに含まれているものより大きな DNA の断片を供給可能である。治療のために約 6 kb ~ 10 Mb の HACs を作製し、従来の輸送方法 (リポソーム、ポリカチオンアミノポリマー、またはベシクル) で供給する。（例えば、Harrington, J. J. 他 (1997) *Nat Genet.* 15: 345 - 355. を参照）。

【0143】

哺乳動物系の組換えタンパク質の長期にわたる産生のためには、株化細胞における PRTS の安定した発現が望ましい。例えば、発現ベクターを用いて、PRTS をコードする配列を株化細胞に形質転換することが可能である。このような発現ベクターは、ウイルス起

源の複製及び/または内在性の発現要素や、同じ或いは別のベクターの上の選択マーカー遺伝子を含む。ベクターの導入の後、細胞を選択培地に移す前に、強化培地で約1~2日の間増殖させる。選択マーカーの目的は選択的な媒介物に対する抵抗性を与えるとともに、その存在により導入された配列を確実に発現する細胞の増殖及び回収が可能となる。安定的に形質転換された細胞の耐性クローンは、その細胞型に好適な組織培養技術を用いて増殖可能である。

【0144】

任意の数の選択系を用いて、形質転換された細胞系を回収することが可能である。選択系には、以下のものに限定はしないが、単純ヘルペスウイルスチミジンキナーゼ遺伝子及びアデニンホスホリボシルトランスフェラーゼ遺伝子が含まれ、それぞれ tk^r または ap^r 細胞において使用される。(例えば、Wigler, M. 他 (1977) Cell 11:223-232; 及び Lowy, I. 他 (1980) Cell 22:817-823を参照)。また代謝拮抗物質、抗生物質或いは除草剤への耐性を選択のベースとして用いることができる。例えば dhfr はメトトレキセートに対する耐性を与え、neo はアミノグリコシッドネオマイシン及び G-418 に対する耐性を与え、als 或いは pat はクロルスルフロン (cPRTS sulfuron)、ホスフィンオトリンアセチルトランスフェラーゼ (phosphinotricin acetyltransferase) に対する耐性を与える(例えば、Wigler, M. 他 (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. 他 (1981) J. Mol. Biol. 150:1-14を参照)。さらに選択に利用できる遺伝子、例えば、代謝のために細胞が必要なものを変える trpB 及び hisD が文献に記載されている(例えば、Hartman, S. C. 及び R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51を参照)。アノトシアニン、緑色蛍光タンパク質 (GFP; Clontech)、グルクロニダーゼ及びその基質 GUS, ルシフェラーゼ及びその基質ルシフェリンなどの可視マーカーが用いられる。緑色蛍光タンパク質 (GFP) (Clontech, Palo Alto, CA) も使用できる。これらのマーカーを用いて、トランスフォーマントを特定するだけでなく、特定のベクター系に起因する一過性或いは安定したタンパク質発現を定量することが可能である(例えば、Rhodes, C. A. 他 (1995) Methods Mol. Biol. 55:121-131を参照)。

【0145】

マーカー遺伝子の発現の存在/不在によって目的の遺伝子の存在が示されても、その遺伝子の存在及び発現の確認が必要な場合もある。例えば、PRTSをコードする配列がマーカー遺伝子配列の中に挿入された場合、PRTSをコードする配列を含む形質転換された細胞は、マーカー遺伝子機能の欠落により特定可能である。または、1つのプロモーターの制御下でマーカー遺伝子がPRTSをコードする配列と一列に配置することも可能である。誘導または選択に回答したマーカー遺伝子の発現は、通常タンデム遺伝子の発現も示す。

【0146】

一般に、PRTSをコードする核酸配列を含み、PRTSを発現する宿主細胞は、当業者に周知の種々の方法を用いて特定することが可能である。これらの方法には、DNA-DNA 或いは DNA-RNA ハイブリダイゼーションや、PCR法、核酸或いはタンパク質の検出及び/または数量化のための膜系、溶液ベース、或いはチップベースの技術を含むタンパク質生物学的試験法または免疫学的アッセイが含まれるが、これらに限定されるものではない。

【0147】

特異的なポリクローナル抗体またはモノクローナル抗体のどちらかを用いるPRTSの発現の検出及び計測のための免疫学的方法は、当分野で周知である。このような技法には、酵素に結合したイムノソルベントアッセイ (ELISA)、ラジオイムノアッセイ (R

I A)、蛍光標示式細胞分取器(FACS)などがある。PRTS上の2つの非干渉エピトープに反応するモノクローナル抗体を用いた、2部位のモノクローナルベース免疫アッセイ(two-site, monoclonal-based immunoassay)が好ましいが、競合の結合アッセイも用いることもできる。これらのアッセイ及びその他のアッセイは、当分野では十分に知られている。(例えば、Hampton, R. 他.(1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Sect. IV; Coligan, J. E. 他 Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York, NY; 及び Pound, J. D. (1990) Immunochemical Protocols, Humana Press, Totowa NJ)。

【0148】

種々の標識技術及び結合技術が当業者には周知であり、様々な核酸アッセイおよびアミノ酸アッセイに用いられ得る。PRTSをコードするポリヌクレオチドに関連する配列を検出するための、標識されたハイブリダイゼーションプローブ或いはPCRプローブを生成する方法には、オリゴ標識化、ニックトランスレーション、末端標識化、または標識されたヌクレオチドを用いるPCR増幅が含まれる。別法として、PRTSをコードする配列、またはその任意の断片をmRNAプローブを生成するためのベクターにクローニングすることも可能である。当分野では周知であり市販されているこのようなベクターを、T7, T3, またはSP6などの好適なRNAポリメラーゼ及び標識されたヌクレオチドの追加によって、in vitroでのRNAプローブの合成に用いることができる。これらの方法は、例えば、Amersham Pharmacia Biotech及びPromega (Madison WI)、U.S. Biochemical Corp (Cleveland OH) が市販する種々のキットを用いて行うことができる。容易な検出のために用い得る好適なレポーター分子或いは標識には、基質、コファクター、インヒビター、磁気粒子、及び放射性核種、酵素、蛍光剤、化学発光剤、色素産生剤などが含まれる。

【0149】

PRTSをコードするヌクレオチド配列で形質転換された宿主細胞は、細胞培地でのこのタンパク質の発現及び回収に好適な条件下で培養される。形質転換された細胞から産生されたタンパク質が分泌されるか細胞内に留まるかは、使用されるその配列及び/またはそのベクターによる。PRTSをコードするポリヌクレオチドを含む発現ベクターは、原核細胞膜及び真核細胞膜を透過するPRTSの分泌を誘導するシグナル配列を含むように設計できることは、当業者には理解されよう。

【0150】

更に、挿入した配列の発現調節能力または発現したタンパク質を所望の形にプロセッシングする能力によって宿主細胞株が選択される。このようなポリペプチドの修飾には、アセチル化、カルボキシル化、グリコシル化、リン酸化、脂質化(lipidation)、及びアシル化が含まれるが、これらに限定されるものではない。タンパク質の「プロ」または「プロ」形を切断する翻訳後のプロセッシングを利用して、標的タンパク質、折りたたみ及び/または活性を特定することが可能である。翻訳後の活性のための特定の細胞装置及び特徴のある機構をもつ種々の宿主細胞(例えば、CHO、HeLa、MDCK、MEK293、WI38)がAmerican Type Culture Collection(ATCC; Bethesda, MD)より入手可能であり、外来のタンパク質の正しい修飾及びプロセッシングを確実にするために選択される。

【0151】

本発明の別の実施例では、PRTSをコードする自然或いは変更された、または組換えの核酸配列を上記した任意の宿主系の融合タンパク質の翻訳となる異種配列に結合させる。例えば、市販の抗体によって認識できる異種部分を含むキメラPRTSタンパク質が、P

R T Sの活性のインヒビターに対するペプチドライブラリのスクリーニングを促進し得る。また、異種タンパク質部分及び異種ペプチド部分が、市販の親和性基質を用いて融合タンパク質の精製を促進し得る。このような部分には、グルタチオンSトランスフェラーゼ (G S T)、マルトース結合タンパク質 (M B P)、チオレドキシン (T r x)、カルモジュリン結合ペプチド (C B P)、6 - H i s、F L A G、c - m c、赤血球凝集素 (H A) が含まれるが、これらに限定されるものではない。G S T及びM B P、T r x、C B P、6 - H i sによって、固定されたグルタチオン、マルトース、フェニルアルシン酸化物 (p h e n y l a r s i n e o x i d e)、カルモジュリン、金属キレート樹脂のそれぞれで同族の融合タンパク質の精製が可能となる。F L A G、c - m c、及び赤血球凝集素 (H A) によって、これらのエピトープ標識を特異的に認識する市販のモノクローナル抗体及びポリクローナル抗体を用いた融合タンパク質の免疫親和性の精製ができる。また、P R T Sをコードする配列と異種タンパク質配列との間にあるタンパク質分解切断部位を融合タンパク質が含むように遺伝子操作すると、P R T Sが精製の後に異種部分から切断され得る。融合タンパク質の発現と精製の方法は、A u s u b e l . (1 9 9 5、前出 c h 1 0) . に記載されている。市販されている様々なキットを用いて、融合タンパク質の発現及び精製を促進できる。

10

【 0 1 5 2 】

本発明の別の実施例では、T N Tウサギ網状赤血球可溶化液またはコムギ胚芽抽出系 (P r o m e g a) を用いて i n v i t r o で放射能標識したP R T Sの合成が可能である。これらの系は、T 7またはT 3、S P 6プロモーターと機能的に結合したタンパク質をコードする配列の転写と翻訳をつなげる。翻訳は、例えば、^{3 5} Sメチオニンである放射能標識されたアミノ酸前駆体の存在の下で起こる。

20

【 0 1 5 3 】

本発明のP R T Sまたはその断片を用いて、P R T Sに特異結合する化合物をスクリーニングすることができる。少なくとも1つまたは複数の試験化合物を用いて、P R T Sへの特異的な結合をスクリーニングすることが可能である。試験化合物の例には、抗体、オリゴヌクレオチド、タンパク質 (例えば受容体) または小分子が挙げられる。

【 0 1 5 4 】

一実施例では、このように同定された化合物は、例えばリガンドやその断片などのP R T Sの天然のリガンド、または天然の基質、構造的または機能的な擬態性または自然結合パートナーに密接に関連している (C o l i g a n , J . E . 他 (1 9 9 1) C u r r e n t P r o t o c o l s i n I m m u n o l o g y 1 (2) の5章等を参照)。同様に、化合物は、P R T Sが結合する天然受容体、或いは例えばリガンド結合部位などの少なくとも受容体のある断片に密接に関連し得る。何れの場合も、既知の技術を用いてこの化合物を合理的に設計することができる。一実施例では、このような化合物に対するスクリーニングには、分泌タンパク質或いは細胞膜上のタンパク質の何れか一方としてP R T Sを発現する好適な細胞の作製が含まれる。好適な細胞には、哺乳動物、酵母、大腸菌からの細胞が含まれる。P R T Sを発現する細胞またはP R T Sを含有する細胞膜断片を試験化合物と接触させて、P R T Sまたは化合物の何れかの結合、刺激または阻害を分析する。

30

40

【 0 1 5 5 】

あるアッセイは、単に試験化合物をポリペプチドに実験的に結合させ、結合を、フルオロフォア、放射性同位体、酵素抱合体またはその他の検出可能な標識により検出することができる。例えば、このアッセイは、少なくとも1つの試験化合物を、溶液中の或いは固体支持物に固定されたP R T Sと結合させるステップと、P R T Sとこの化合物との結合を検出するステップを含み得る。別法では、標識された競合物の存在下での試験化合物の結合の検出及び測定を行うことができる。更にこのアッセイでは、細胞遊離剤、化学ライブラリまたは天然の生成混合物を用いて実施することができ、試験化合物は、溶液中で遊離させるか固体支持体に固定させる。

【 0 1 5 6 】

50

本発明の P R T S またはその断片を用いて、P R T S の活性を調整する化合物をスクリーニングすることが可能である。このような化合物には、アゴニスト、アンタゴニスト、或るいは部分的または逆アゴニスト等が含まれる。一実施例では、P R T S が少なくとも 1 つの試験化合物と結合する、P R T S の活性が許容される条件下でアッセイを実施し、試験化合物の存在下での P R T S の活性が試験化合物不在下での P R T S の活性と比較する。試験化合物の存在下での P R T S の活性の変化は、P R T S の活性を調整する化合物の存在を示唆する。別法では、試験化合物を P R T S の活性に適した条件下で P R T S を含む *i n v i t r o* または細胞遊離系と結合させてアッセイを実施する。これらアッセイの何れかにおいて、P R T S の活性を調整する試験化合物は間接的に結合することが可能であり、試験化合物と直接接触する必要がない。少なくとも 1 つから複数の試験化合物をスクリーニングすることができる。

10

【0157】

別の実施例では、胚性幹細胞 (E S 細胞) における相同組換えを用いて動物モデル系内で、P R T S またはその哺乳動物相同体をコードするポリヌクレオチドを「ロックアウト」する。このような技術は当技術分野において周知であり、ヒト疾患動物モデルの作製に有用である (米国特許第 5 , 1 7 5 , 3 8 3 号及び第 5 , 7 6 7 , 3 3 7 号等を参照) 。例えば 1 2 9 / S v J 細胞株等のマウス E S 細胞は初期のマウス胚に由来し、培地で増殖させることができる。この E S 細胞は、ネオマイシンホスホトランスフェラーゼ遺伝子 (*n e o : C a p e c c h i , M . R . (1 9 8 9) S c i e n c e 2 4 4 : 1 2 8 8 - 1 2 9 2*) 等のマーカー遺伝子で破壊した目的の遺伝子を含むベクターで形質転換する。このベクターは、相同組換えにより宿主ゲノムの対応する領域に組み込まれる。別法では、C r e - l o x P 系を用いて相同組換えを行い、組織特異的または発生段階特異的に目的遺伝子をロックアウトする (*M a r t h , J . D . (1 9 9 6) C l i n . I n v e s t . 9 7 : 1 9 9 9 - 2 0 0 2 ; W a g n e r , K . U . 他 (1 9 9 7) N u c l e i c A c i d s R e s . 2 5 : 4 3 2 3 - 4 3 3 0*) 。形質転換した E S 細胞を同定し、例えば C 5 7 B L / 6 マウス系等から採取したマウス細胞胚盤胞に微量注入する。胚盤胞を偽妊娠メスに外科的に導入し、得られるキメラ子孫の遺伝形質を決め、これを交配させてヘテロ接合性系またはホモ接合性系を作製する。このようにして作製した遺伝子組換え動物は、潜在的な治療薬や毒性薬剤で検査することができる。

20

【0158】

P R T S をコードするポリヌクレオチドを *i n v i t r o* でヒト胚盤胞由来の E S 細胞において操作することが可能である。ヒト E S 細胞は、内胚葉、中胚葉及び外胚葉の細胞の種類を含む少なくとも 8 つの別々の細胞系統に分化する可能性を有する。これらの細胞系統は、例えば神経細胞、造血系統及び心筋細胞に分化する (*T h o m s o n , J . A . 他 (1 9 9 8) S c i e n c e 2 8 2 : 1 1 4 5 - 1 1 4 7*) 。

30

【0159】

P R T S をコードするポリヌクレオチドを用いて、ヒト疾患をモデルとした「ロックイン」ヒト化動物 (ブタ) または遺伝子組換え動物 (マウスまたはラット) を作製することが可能である。ロックイン技術を用いて、P R T S をコードするポリヌクレオチドの或る領域を動物 E S 細胞に注入し、注入した配列を動物細胞ゲノムに組み込ませる。形質転換細胞を胞胚に注入し、胞胚を上記のように移植する。遺伝子組換え子孫または近交系について研究し、潜在的な医薬品を用いて処理し、ヒトの疾患の治療に関する情報を得る。別法では、例えば P R T S を乳汁内に分泌するなど P R T S を過剰に発現する哺乳動物近交系は、便利なタンパク質源となり得る (*J a n n e , J . 他 (1 9 9 8) B i o t e c h n o l . A n n u . R e v . 4 : 5 5 - 7 4*) 。

40

【0160】

(治療)

P R T S のある領域とプロテアーゼのある領域との間に、例えば配列及びモチーフの文脈における化学的及び構造的類似性が存在する。更に、P R T S の発現は、胃腸組織、上皮組織、生殖組織、心血管組織、癌組織、および炎症組織に密接に関連し、また、健康な腎

50

臓及び皮膚組織に密接に関連する。従って、PRTSは、胃腸疾患、心血管疾患、自己免疫/炎症性の疾患、増殖異常、発生障害、上皮の疾患、神経の疾患、および生殖障害においてある役割を果たすと考えられる。PRTSの発現若しくは活性の増大に関連する疾患の治療においては、PRTSの発現または活性を低下させることが望ましい。また、PRTSの発現または活性の低下に関連する疾患の治療においては、PRTSの発現または活性を増大させることが望ましい。

【0161】

従って、一実施例において、PRTSの発現または活性の低下に関連した疾患の治療または予防のために、患者にPRTSまたはその断片や誘導体を投与することが可能である。限定するものではないが、このような疾患には胃腸疾患、心血管疾患、自己免疫/炎症性の疾患、細胞増殖異常、発生障害、上皮の疾患、神経の疾患、および生殖障害が含まれ、胃腸疾患の中には、嚥下障害、消化性食道炎、食道痙攣、食道狭窄、食道癌、消化不良、消化障害、胃炎、胃癌、食欲不振、悪心、嘔吐、胃不全麻痺、洞または幽門の浮腫、腹部アングナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリー ヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、及び後天性免疫不全症候群(AIDS)腸症、黄疸、肝性脳症、肝腎症候群、肝炎、肝脂肪症、血色素症、ウィルソン病、*- 1 -*アンチトリプシン欠損症、ライ症候群、原発性硬化性胆管炎、肝梗塞、門脈循環閉塞及び血栓、小葉中心壊死、肝臓紫斑病、肝静脈血栓、肝静脈閉塞症、子癇前症、子癇、妊娠性急性肝脂肪、妊娠性肝臓内胆汁うっ滞と、結節性再生及び腺腫、癌腫を含む肝癌とが含まれ、心血管疾患の中には、動静脈瘻、アテローム硬化、高血圧、脈管炎、レイノー病、静脈奇形、動脈解離、静脈瘤、血栓静脈炎及び静脈血栓、血管の腫瘍、血栓崩壊の合併症、バルーン血管形成術(*balloon angioplasty*)、血管置換術(*vascular replacement*)、大動脈冠動脈バイパス術移植手術(*coronary artery bypass graft surgery*)、うっ血性心不全、虚血性心疾患、狭心症、心筋梗塞、高血圧性心疾患、変性弁膜性心疾患、石灰化大動脈弁狭窄症、先天性2尖大動脈弁、僧帽弁輪状石灰化(*mitral annular calcification*)、僧帽弁脱出、リウマチ熱、リウマチ性心疾患、感染性心内膜炎、非細菌性血栓性心内膜炎、全身性エリテマトーデスの心内膜炎、カルチノイド心疾患、心筋症、心筋炎、心膜炎、腫瘍性心疾患、先天性心臓疾患、先天性肺異常(*congenital lung anomalies*)、肺拡張不全、肺うっ血及び肺水腫、肺動脈塞栓症、肺出血、肺梗塞、肺高血圧症、血管硬化症、閉塞性肺疾患、拘束性肺疾患(*restrictive pulmonary disease*)、慢性閉塞性肺疾患、肺気腫、慢性気管支炎、気管支喘息、細気管支拡張症、細菌性肺炎、ウイルス性肺炎及びマイコプラズマ肺炎、肺膿瘍、肺結核、びまん性間質性疾患(*diffuse interstitial diseases*)、塵肺症、サルコイド症、特発性肺繊維症(*idiopathic pulmonary fibrosis*)、剥離性間質性肺炎、過敏症肺炎(*hypersensitivity pneumonitis*)、肺好酸球増加閉塞性細気管支炎 器質性肺炎(*pulmonary eosinophilia bronchiolitis obliterans - organizing pneumonia*)、びまん性肺出血症候群(*diffuse pulmonary hemorrhage syndromes*)、グッドパスチャー症候群、特発性肺血鉄症、肺併発膠原血管病(*pulmonary involvement in collagen-vascular disorders*)、肺胞たんぱく症、肺腫瘍、炎症性及び非炎症性胸水(*inflammatory and non-inflammatory pleural effusions*)、気胸症、胸膜腫瘍、薬物による肺疾患(*drug-induced lung disease*)、放射線による肺疾患(*radiation-induced lung disease*)及び肺移植の合併症などが含まれ、自己免疫/炎症性の疾患の中には、炎症及び日光性角化症、後天

性免疫不全症候群 (A I D S) 及び副腎機能不全、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、アテローム斑破裂、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー (A P E C E D)、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、関節軟骨の分解、骨粗しょう症、瘰癧、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、細胞増殖異常の中には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病 (M C T D)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、精巣、胸腺、甲状腺、及び子宮の癌が含まれ、発生障害も含まれ、その中には尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ ベッカー型筋ジストロフィー、骨吸収、癩癧、性腺形成異常、W A G R 症候群 (ウィルムス腫瘍、無虹彩症、尿生殖器異常、精神薄弱)、スミス マジェニス症候群 (S m i t h - M a g e n i s s y n d r o m e)、脊髄形成異常症候群、遺伝性粘膜上皮異形成、遺伝性角皮症、シャルコー マリー ツース病及び神経線維腫症などの遺伝性神経病、甲状腺機能低下症、水頭症、Syndenham 舞蹈病 (S y n d e n h a m ' s c h o r e a) 及び脳性小児麻痺などの発作障害、脊髄二分裂、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、加齢性黄斑変性症、感覚神経性聴力損失が含まれ、上皮の疾患の中には、汗疱状湿疹、アレルギー性接触皮膚炎、角化上皮症、黒皮症、白斑、日光性角化症、基底細胞癌、扁平上皮癌、脂漏性角化症、毛包炎、単純ヘルペス、帯状疱疹、水痘、カンジダ症、皮膚糸状菌症、疥癬、昆虫刺症、老年性血管腫、ケロイド、皮膚線維腫、先端線維性軟肛、じんま疹、一過性棘融解性皮膚症、乾燥症、湿疹、アトピー性皮膚炎、接触性皮膚炎、手湿疹、貨幣状湿疹、ピダール苔癬、a s t e a t o t i c e c z e m a、うつ血性皮膚炎、うつ血性潰瘍、脂漏性皮膚炎、乾癬、扁平苔癬、バラ色粧糖疹、膿痂疹、膿瘡、皮膚糸状菌症、なまず、疣贅、尋常性座瘡、赤鼻、尋常性天疱瘡、落葉状天疱瘡、腫瘍随伴性天疱瘡、水疱性類天疱瘡、妊娠性疱疹、疱疹状皮膚炎、線形性 I g A 疾患、後天性表皮水疱症、皮膚筋炎、紅斑性狼瘡、強皮症および限局性強皮症、紅皮症、脱毛症、定形皮膚障害 (f i g u r a t e s k i n l e s i o n s)、血管拡張症、色素沈着低下、色素沈着過度、小疱疹 / 水疱、発疹、皮膚薬物反応、丘疹小結節皮膚障害 (p a p u l o n o d u l a r s k i n l e s i o n s)、慢性非治癒性創傷、感光性疾患、単純表皮水泡症、表皮剥離および非表皮剥離性掌蹠角皮症、ジューメンス水疱魚鱗癬 (i c h t h y o s i s b u l l o s a o f S i e m e n s)、剥脱性魚鱗癬、手掌足底角化症、掌蹠角皮症 (k e r a t o s i s p a l m o p l a n t a r i s または p a l m o p l a n t a r k e r a t o d e r m a)、k e r a t o s i s p u n c t a t a、M e e s m a n n ' s c o r n e a l d y s t r o p h y、先天性爪肥厚症、白色海綿母斑、多発性皮脂嚢腫、表皮性母斑 / 表皮剥離性角化型、m o n i l e t b r i x、t r i c h o t h i o d y s t r o p h y、慢性肝炎 / 特発性肝硬変、結腸直腸過形成が含まれ、神経の疾患の中には、癩癧、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クール

10

20

30

40

50

ー及びクロイツフェルト ヤコブ病、ゲルストマン症候群、Gerstmann-Straussler-Scheinker症候群を含むプリオン病と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫(cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、ダウン症を含む中枢神経系性精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性の障害、分裂病性疾患を含む精神病と、季節性障害(SAD)と、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、带状疱疹後神経痛、及びトゥレット病と、進行性核上麻痺、皮質基部変性(corticobasal degeneration)及び家族性の前頭側頭性健忘症とが含まれ、生殖障害の中には、プロラクチンの産生異常と、卵管病及び排卵異常、子宮内膜症、発情期異常、月経周期異常、多嚢胞卵巣症候群、卵巣過剰刺激症候群、子宮内膜癌及び卵巣癌、子宮筋腫、自己免疫異常、異所性妊娠、奇形発生を含む不妊症と、乳癌、線維嚢胞性乳腺症、乳漏症と、精子形成異常、生理学上の精子異常、精巣癌、前立腺癌、良性の前立腺過形成、前立腺炎、ペーロニー病、インポテンス、男性乳房及び女性化乳房癌とが含まれる。

10

【0162】

別の実施例では、限定するものではないが上に列記した疾患を含むPRTSの発現または活性の低下に関連した疾患の治療または予防のために、PRTSまたはその断片や誘導体を発現し得るベクターを患者に投与することも可能である。

20

【0163】

更に別の実施例では、限定するものではないが上に列記した疾患を含むPRTSの発現または活性の低下に関連した疾患の治療または予防のために、実質的に精製されたPRTSを含む組成物を好適な医薬用担体と共に患者に投与することも可能である。

【0164】

更に別の実施例では、限定するものではないが上に列記した疾患を含むPRTSの発現または活性の低下に関連した疾患の治療または予防のために、PRTSの活性を調節するアゴニストを患者に投与することも可能である。

【0165】

更なる実施例では、PRTSの発現または活性の増大に関連した疾患の治療または予防のために、患者にPRTSのアンタゴニストを投与することが可能である。限定するものではないが、このような疾患の例には、上記した胃腸疾患、心血管疾患、自己免疫/炎症性の疾患、増殖異常、発生障害、上皮の疾患、神経の疾患、および生殖障害が含まれる。一実施態様では、PRTSと特異的に結合する抗体が直接アンタゴニストとして、或いはPRTSを発現する細胞または組織に薬剤を運ぶターゲティング或いは運搬機構として間接的に用いられ得る。

30

【0166】

別の実施例では、限定するものではないが上に列記した疾患を含むPRTSの発現または活性の増大に関連した疾患の治療または予防のために、PRTSをコードするポリヌクレオチドの相補配列を発現するベクターを患者に投与することも可能である。

40

【0167】

別の実施例では、本発明の任意のタンパク質、アンタゴニスト、抗体、アゴニスト、相補的な配列、ベクターを別の好適な治療薬と組み合わせることもできる。当業者は、従来の医薬原理にしたがって併用療法で用いる好適な治療薬を選択可能である。治療薬との組み合わせにより、上に列記した種々の疾患の治療または予防に相乗効果をもたらし得る。この方法を用いて少ない量の各薬剤で医薬効果をあげることが可能であり、広範囲な副作用の可能性を低減し得る。

【0168】

PRTSのアンタゴニストは、当分野で一般的な方法を用いて製造することが可能である

50

。詳しくは、精製された P R T S を用いて抗体を作ったり、治療薬のライブラリをスクリーニングして P R T S と特異的に結合するものを同定が可能である。P R T S の抗体も、当分野で一般的な方法を用いて製造することが可能である。このような抗体には、ポリクローナル抗体、モノクローナル抗体、キメラ抗体、一本鎖、F a b フラグメント、及び F a b 発現ライブラリによって作られたフラグメントが含まれる。但し、これらに限定されるものではない。治療用には、中和抗体（即ち、二量体の形成を阻害するもの）が特に好ましい。

【0169】

抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ヒト及びその他のものを含む種々の宿主が、P R T S または任意の断片、または免疫原性の特性を備えるそのオリゴペプチドの注入によって免疫化され得る。宿主の種に応じて、種々のアジュバントを用いて免疫応答を高めることもできる。このようなアジュバントにはフロイントアジュバント、水酸化アルミニウムなどのミネラルゲルアジュバント、リゾレシチン、プルロニックポリオール、ポリアニオン、ペプチド、油性乳剤、キーホールリンペットヘモシニアン、及びジニトロフェノールなどの界面活性剤が含まれるが、これらに限定されるものではない。ヒトに用いられるアジュバントの中では、B C G (b a c i l l i C a l m e t t e - G u e r i n) 及び C o r y n e b a c t e r i u m p a r v u m が特に好ましい。

10

【0170】

P R T S に対する抗体を誘発するために用いられるオリゴペプチド、ペプチド、または断片は、少なくとも約5個のアミノ酸からなり、一般的には約10個以上のアミノ酸からなるものが好ましい。これらのオリゴペプチド或いはペプチド、またはそれらの断片は、天然のタンパク質のアミノ酸配列の一部と同一であることが望ましい。P R T S アミノ酸の短いストレッチは、K L H などの別のタンパク質の配列と融合し、キメラ分子に対する抗体が産生され得る。

20

【0171】

P R T S に対するモノクローナル抗体は、培地内の連続した細胞株によって、抗体分子を産生する任意の技術を用いて作製することが可能である。これらの技術には、ハイブリドーマ技術、ヒトB細胞ハイブリドーマ技術、及びE B V - ハイブリドーマ技術が含まれるが、これらに限定されるものではない（例えば、K o h l e r , G . ら . (1 9 7 5) Nature 256 : 495 - 497 ; K o z b o r , D . ら . (1 9 8 5) . J . Immunol . Methods 81 - 8 - 42 ; C o t e , R . J . ら . (1 9 8 3) Proc . Natl . Acad . Sci . 80 : 2026 - 2030 ; C o l e , S . P . ら . (1 9 8 4) Mol . Cell Biol . 62 : 109 - 120 を参照）。

30

【0172】

更に、「キメラ抗体」作製のために発達したヒト抗体遺伝子にマウス抗体遺伝子をスプライシングするなどの技術が、好適な抗原特異性及び生物学的活性を備える分子を得るために用いられる（例えば、M o r r i s o n , S . L . 他 . (1 9 8 4) Proc . Natl . Acad . Sci . 81 - 4851 - 4855 ; N e u b e r g e r , M . S . 他 . (1 9 8 4) Nature 312 : 604 - 608 ; T a k e d a , S . ら . (1 9 8 5) Nature 314 : 452 , 454 を参照）。別法では、当分野で周知の方法を用いて、一本鎖抗体の産生のための記載された技術を適用して、P R T S 特異性一本鎖抗体を生成する。関連する特異性を備えるが別のイディオタイプの組成の抗体は、ランダムな組み合わせの免疫グロブリンライブラリから鎖混合によって生成することもできる（例えば、B u r t o n D . R . (1 9 9 1) Proc . Natl . Acad . Sci . 88 : 11120 - 3 を参照）。

40

【0173】

抗体は、リンパ球集団の中の i n v i v o 産生を誘発することによって、または免疫グロブリンライブラリのスクリーニングまたは文献に示されているような、高度に特異的な結合試薬のパネルをスクリーニングすることによって、得ることもできる（例えば、O r

50

landi, R. 他. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. 他. (1991) Nature 349: 293-299を参照)。

【0174】

PRTSに対する特異的な結合部位を含む抗体も得ることができる。例えば、このような断片には、抗体分子のペプシン消化によって生成されるF(ab')に断片と、F(ab')に断片のジスルフィド架橋を減じることによって生成されるFab断片が含まれるが、これらに限定されるものではない。別法では、Fab発現ライブラリを作製することによって、所望の特異性とモノクローナルFab断片の迅速且つ容易な同定が可能となる(例えば、Huse, W.D.ら. (1989) Science 254: 1275-1281を参照)。

10

【0175】

種々のイムノアッセイを用いてスクリーニングし、所望の特異性を有する抗体を同定する。隔離された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる競合的な結合、または免疫放射線活性のための数々のプロトコルが、当分野では周知である。通常このようなイムノアッセイには、PRTSとその特異性抗体との間の複合体調整の計測が含まれる。二つの非干渉性PRTSエピトープに対して反応性のモノクローナル抗体を用いる、2部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合的結合アッセイも利用することができる(Pound、前出)。

【0176】

ラジオイムノアッセイ技術と共にScatchard分析などの様々な方法を用いて、PRTSに対する抗体の親和性を評価する。親和性を結合定数Kaで表すが、このKaは、平衡状態の下でPRTS抗体複合体のモル濃度を遊離抗体と遊離抗原のモル濃度で除して得られる値である。多数のPRTSエピトープに対して親和性が不均一なポリクローナル抗体医薬のKaは、PRTSに対する抗体の平均親和性または結合活性を表す。特定のPRTSエピトープに単一特異的なモノクローナル抗体医薬のKaは、親和性の真の測定値を表す。Ka値が $10^9 \sim 10^{12}$ L/molの高親和性抗体医薬は、PRTS抗体複合体が激しい操作に耐えなければならないイムノアッセイに用いるのが好ましい。Ka値が $10^6 \sim 10^7$ L/molの低親和性抗体医薬は、PRTSが抗体から最終的に活性化状態で解離する必要がある免疫精製(immunopurification)及び類似の処理に用いるのが好ましい。(Catty, D. (1988) Antibodies, Volume I: A Practical Approach. IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

20

30

【0177】

ある下流での適用におけるこのような医薬品の品質及び適性を調べるために、ポリクローナル抗体医薬の抗体価及び結合活性を更に評価する。例えば、少なくとも $1 \sim 2$ mg/mlの特異的な抗体、好ましくは $5 \sim 10$ mg/mlの特異的な抗体を含むポリクローナル抗体医薬は一般に、PRTS抗体複合体を沈殿させなければならない処理に用いられる。様々な適用例における抗体の特異性及び抗体価、結合活性、抗体の品質や使用法の指針は一般に入手可能である。(例えば、Catty、前出、及びColigan他、前出を参照)。

40

【0178】

本発明の別の実施例では、PRTSをコードするポリヌクレオチド、またはその任意の断片や相補配列が、治療目的で使用することができる。ある実施態様では、PRTSをコードする遺伝子のコーディング領域や調節領域に相補的な配列やアンチセンス分子(DNA及びRNA、修飾ヌクレオチド)を設計して遺伝子発現を変更することができる。このような技術は当分野では周知であり、センスまたはアンチセンスオリゴヌクレオチドまたは大きな断片が、PRTSをコードする配列の制御領域から、またはコード領域に沿ったさ

50

さまざまな位置から設計可能である。

【0179】

治療に用いる場合、アンチセンス配列を好適な標的細胞に導入するのに好適な任意の遺伝子送達系を用いることができる。アンチセンス配列は、転写時に標的タンパク質をコードする細胞配列の少なくとも一部に相補的な配列を発現する発現プラスミドの形で細胞内に送達することができる(例えば、Slater, J. E. 他(1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K. J. 他(1995) 9(13):1288-1296.を参照)。また、アンチセンス配列は、例えばレトロウイルスやアデノ関連ウイルスベクター等のウイルスベクターを用いて細胞内に導入することもできる(例えば、Miller, A. D. (1990) *Blood* 76:271; Ausubel, 前出; Ucker, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347を参照)。その他の遺伝子送達機構には、リポソーム系、人工的なウイルスエンベロープ、及び当分野で周知のその他の系が含まれる(Rossi, J. J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R. J. 他(1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M. C. 他(1997) *Nucleic Acids Res.* 25(14):2730-2736.を参照)。

【0180】

本発明の別の実施例では、PRTSをコードするポリヌクレオチドを、体細胞若しくは生殖細胞の遺伝子治療に用いることが可能である。遺伝子治療は、(i) 遺伝子欠損症(例えば、X染色体連鎖遺伝(Cavazzana-Calvo, M. 他(2000) *Science* 288:669-672)によって特徴づけられる重度の複合型免疫欠損(SCID) - X1)、遺伝性アデノシン-デアミナーゼ(ADA)欠損症(Blaese, R. M. 他(1995) *Science* 270:475-480; Bordignon, C. 他(1995) *Science* 270:470-475)に関連する重度の複合型免疫欠損、嚢胞性繊維症(Zabner, J. 他(1993) *Cell* 75:207-216; Crystal, R. G. 他(1995) *Hum. Gene Therapy* 6:643-666; Crystal, R. G. 他(1995) *Hum. Gene Therapy* 6:667-703)、サラセミア(thalassaemia)、家族性高コレステロール血症、第VIII因子若しくは第IX因子欠損による血友病(Crystal, R. G. (1995) *Science* 270:404-410; Verma, I. M. and Somia, N. (1997) *Nature* 389:239-242)を治療したり、(ii) 条件的致死性遺伝子産物(例えば、細胞増殖の制御不能による癌の場合)を発現させたり、及び(iii) 細胞内の寄生虫(例えば、ヒト免疫不全ウイルス(HIV)(Baltimore, D. (1988) *Nature* 335:395-396; Poeschl, E. 他(1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399)や、B型若しくはC型肝炎ウイルス(HBV、HCV)、Candida albicans及びParacoccidioides brasiliensis等の真菌寄生虫、Plasmodium falciparum及びTrypanosoma cruzi等の原虫寄生体)に対する防御機能を有するタンパク質を発現させて行うことができる。PRTSの発現若しくは調節に必要な遺伝子の欠損が疾患を引き起こす場合、導入した細胞の好適な集団からPRTSを発現させて、遺伝子欠損によって起こる症状の発現を緩和することが可能である。

【0181】

本発明の更なる実施例では、PRTSの欠損による疾患や異常症は、PRTSをコードする哺乳動物発現ベクターを作製して、これらのベクターを機械的手段によってPRTS欠損細胞に導入することによって治療する。in vivo 或いは ex vitro の細胞に用いる機械的な導入技術には、(i) 個々の細胞内へのDNAのマイクロインジェクショ

ン、(ii)金粒子の打ち込み、(iii)リポソーム仲介性トランスフェクション、(iv)受容体仲介性遺伝子導入、及び(v)DNAトランスポゾン(Morgan, R. A. and W. F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Recipon (1998) Curr. Opin. Biotechnol. 9:445-450)の使用が含まれる。

【0182】

PRTSの発現に影響を及ぼし得る発現ベクターには、限定するものではないが、PCDNA3.1、EPITAG、PRCCMV2、PREP、PVAXベクター(Invitrogen, Carlsbad CA)、PCMV-SCRIPT、PCMV-TAG、PEGSH/PERV(Stratagene, La Jolla CA)、PTET-OFF、PTET-ON、PTRE2、PTRE2-LUC、PTK-HYG(Clontech, Palo Alto CA)が含まれる。PRTSを発現させるために、(i)恒常的に活性なプロモーター(例えば、サイトメガロウイルス(CMV)、ラウス肉腫ウイルス(RSV)、SV40ウイルス、チミジンキナーゼ(TK)、若しくは -アクチン遺伝子等)、(ii)誘導性プロモーター(例えば、市販されているT-REXプラスミド(Invitrogen)に含まれている、テトラサイクリン調節性プロモーター(Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen, M. 他(1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456)、エクジソン誘導性プロモーター(市販されているプラスミドPVGRXR及びPINDに含まれている:Invitrogen)、FK506/ラパマイシン誘導性プロモーター、またはRU486/ミフェプリストーン誘導性プロモーター(Rossi, F.M.V. and H.M. Blau, 前出)、または(iii)正常な個体に由来するPRTSをコードする内在性遺伝子の天然のプロモーター若しくは組織特異的プロモーターを用いることが可能である。

【0183】

市販のリポソーム形質転換キット(例えば、Invitrogenが販売しているPERFECT LIPID及びTRANSFECTION KIT)を用いれば、当業者は経験にそれほど頼らないでもポリヌクレオチドを培養中の標的細胞に導入することが可能である。別法では、リン酸カルシウム法(Graham, F.L. and A.J. Eb (1973) Virology 52:456-467)若しくは電気穿孔法(Neumann, B. 他(1982) EMBO J. 1:841-845)を用いて形質転換を行う。初代細胞にDNAを導入するためには、これらの標準的な哺乳動物トランスフェクションプロトコルを変更する必要がある。

【0184】

本発明の別の実施例では、PRTSの発現に関連する遺伝子欠損によって起こる疾患や異常症は、(i)レトロウイルス末端反復配列(LTR)プロモーター若しくは独立したプロモーターのコントロール下でPRTSをコードするポリヌクレオチドと、(ii)好適なRNAパッケージングシグナルと、(iii)追加のレトロウイルス・シス作用性RNA配列及び効率的なベクターの増殖に必要なコーディング配列を伴うRev応答性エレメント(RRE)とからなるレトロウイルスベクターを作製して治療することができる。レトロウイルスベクター(例えば、PFB及びPFBNEO)はStratagene社から入手可能であり、公表データ(Riviere, I. 他.(1995) Proc. Natl. Acad. Sci. U.S.A. 92:6733-6737)に基づいている。上記データを引用することをもって本明細書の一部とする。このベクターは、VSVg(Armentano, D. 他(1987) J. Virol. 61:1647-1650; Bender, M.A. 他(1987) J. Virol. 61:

1639 - 1646 ; Adam, M. A. and A. D. Miller (1988) J. Virol. 62:3802 - 3806 ; Dull, T. 他 (1998) J. Virol. 72:8463 - 8471 ; Zufferey, R. 他 (1998) J. Virol. 72:9873 - 9880) 等の乱交雑エンベロープタンパク質若しくは標的細胞上の受容体に対する親和性を有するエンベロープ遺伝子を発現する好適なベクター産生細胞系 (VPCL) において増殖される。RIGGに付与された米国特許第5,910,434号 (「Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant」) において、レトロウイルスパッケージング細胞系を得るための方法が開示されており、引用することをもって本明細書の一部とする。レトロウイルスベクターの増殖、ある細胞集団 (例えば、CD4⁺T細胞) の形質導入、並びに形質導入した細胞を患者に戻す方法は、遺伝子治療の分野では周知であり、多数の文献に記載されている (Ranga, U. 他. (1997) J. Virol. 71:7020 - 7029 ; Bauer, G. 他 (1997) Blood 89:2259 - 2267 ; Bonyhadi, M. L. (1997) J. Virol. 71:4707 - 4716 ; Ranga, U. 他 (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201 - 1206 ; Su, L. (1997) Blood 89:2283 - 2290)。

【0185】

別法では、アデノウイルス系遺伝子治療の送達系を用いて、PRTSの発現に関連する1 あるいは複数の遺伝子異常を有する細胞にPRTSをコードするポリヌクレオチドを送達する。アデノウイルス系ベクターの作製及びパッケージングは当分野では周知である。複製欠損型アデノウイルスベクターは、免疫調節タンパク質をコードする遺伝子を脾臓の無損傷の脾島の中に導入するために可変性であることが証明された (Csete, M. E. 他. (1995) Transplantation 27:263 - 268)。使用できる可能性のあるアデノウイルスベクターが、米国特許第5,707,618号 (「Adenovirus vectors for gene therapy」) に記載されており、引用することをもって本明細書の一部とする。アデノウイルスベクターについてはまた、Antinozzi, P. A. 他 (1999) Annu. Rev. Nutr. 19:511 - 544 ; and Verma, I. M. and N. Somia (1997) Nature 18:389:239 - 242を参照し、引用することをもって本明細書の一部とする。

【0186】

別法では、ヘルペス系遺伝子治療の送達系を用いて、PRTSの発現に関連する1 あるいは複数の遺伝子異常を有する標的細胞にPRTSをコードするポリヌクレオチドを送達する。単純疱疹ウイルス (HSV) 系のベクターは、HSV親和性の中枢神経細胞にPRTSを導入する際に特に重要である。ヘルペス系ベクターの作製及びパッケージングは当分野では周知である。複製適格性の単純疱疹ウイルス (HSV) I型系のベクターは、霊長類の眼にレポーター遺伝子を送達するために用いられてきた (Liu, X. 他 (1999) Exp. Eye Res. 169:385 - 395)。HSV-1ウイルスベクターの作製は、DeLucaに付与された米国特許第5,804,413号 (Herpes simplex virus swains for gene transfer) に記載されており、引用することをもって本明細書の一部とする。米国特許第5,804,413号には、ヒト遺伝子治療を含む目的のために、好適なプロモーターのコントロールの下で、細胞に導入される少なくとも1つの内在性遺伝子を含むゲノムからなる組換えHSV d92についての記載がある。また上記特許には、ICP4、ICP27及びICP22のために除去される組換えHSV株の作製及び使用方法が開示されている。HSVベクターについては、Goins, W. F. 他 (1999) J. Virol. 73:519 - 532 and Xu, H. 他 (1994) Dev. Biol. 163:15

2 - 161を参照し、引用することをもって本明細書の一部とする。クローニングされたヘルペスウイルス配列の操作や、巨大ヘルペスウイルスのゲノムの異なった部分を含む多数のプラスミドをトランスフェクトした後の組換えウイルスの継代、ヘルペスウイルスの成長及び増殖、並びにヘルペスウイルスの細胞への感染は当分野で周知の技術である。

【0187】

別法では、ウイルス（正の一本鎖RNAウイルス）ベクターを用いてPRTSをコードするポリヌクレオチドを標的細胞に送達する。プロトタイプのウイルスであるセムリキ森林熱ウイルス（Semliki Forest Virus, SFV）の生物学的な研究が広範に行われ、遺伝子伝達ベクター（gene transfer vector）がSFVゲノムに基づいていることが分かった（Garoff, H. and K. - J. Li (1998) *Cun. Opin. Biotech.* 9: 464 - 469）。

ウイルスRNAの複製中に、通常はウイルスカプシドタンパク質をコードするサブゲノムRNAが作り出される。このサブゲノムRNAが完全長のゲノムRNAより高いレベルで複製されるため、酵素活性（例えばプロテアーゼ及びポリメラーゼ）を有するウイルスタンパク質に対してカプシドタンパク質が過剰に産生される。同様に、PRTSをコードする配列をウイルスゲノムのカプシドをコードする領域に導入することによって、ベクター導入細胞において多数のPRTSをコードするRNAが産生され、高いレベルでPRTSが合成される。通常はウイルス感染は2～3日以内の細胞溶解に関係するが、シンドビスウイルス（SIN）の変異体を有するハムスターの正常な腎細胞（BHK-21）の持続的な感染を確立する能力は、ウイルスの溶解性の複製が遺伝子治療に適用できるように好適に変更することが可能であることを示唆している（Dryga, S. A. 他. (1997) *Virology* 228: 74 - 83）。様々な宿主にウイルスを導入できることから、様々なタイプの細胞にPRTSを導入することができる。ある集団における細胞のサブセットの特定の形質導入には、形質導入する前に細胞のソーティングを必要とする場合がある。ウイルスの感染性cDNAクローンの操作、ウイルスcDNA及びRNAのトランスフェクション、並びにウイルスの感染方法は当分野で周知である。

【0188】

例えば開始部位から約-10から約+10までの転写開始部位に由来するオリゴヌクレオチドを用いて、遺伝子の発現を阻害することが可能である。同様に、三重らせん塩基対合法を用いて阻害することができる。三重らせん構造は、二重らせんがポリメラーゼ、転写因子、または調節分子の結合のために十分に広がるのを阻止するため有用である。三重式DNAを用いる最近の治療の進歩は文献に記載されている（例えば、Geer, J. E. ら. (1994) *In: Huber, B. E. 及び B. I. Carr, Molecular and Immunological Approaches, Futura Publishing Co., Mt. Kisco, NY*を参照）。相補的な配列またはアンチセンス分子もまた、転写物がリボソームに結合するのを阻止することによってmRNAの翻訳を阻止するように設計できる。

【0189】

酵素性RNA分子であるリボザイムは、RNAの特異的切断を触媒するために用いることができる。リボザイム作用の機構には、相補的な標的RNAへのリボザイム分子の配列特異性ハイブリダイゼーションが含まれ、ヌクレオチド鎖切断が続く。例えば、PRTSをコードする配列のヌクレオチド鎖切断を、特異的且つ効果的に触媒する組換え型のハンマーヘッド型リボザイム分子が含まれる。

【0190】

任意の潜在的RNA標的内の特異的なリボザイム切断部位が、後続の配列GUA、GUU、及びGUCを含むリボザイム切断部位に対して、標的分子をスキヤニングすることによって初めに同定される。一度同定されると、切断部位を含む標的遺伝子の領域に対応する15個から20個のリボヌクレオチドの短いRNA配列を、オリゴヌクレオチドの機能を不全にする二次的な構造の特徴について評価することが可能である。候補標的の適合性も

、リボヌクレアーゼ保護アッセイを用いて、相補的なオリゴヌクレオチドとのハイブリダイゼーションの容易性をテストすることによって評価することが可能である。

【0191】

本発明の相補的なリボ核酸分子及びリボザイムは、当分野で周知の方法を用いて、核酸分子の合成のために作製することができる。これらの方法には、固相ホスホラミダイト化合物などのオリゴヌクレオチドを化学的に合成する方法が含まれる。別法では、RNA分子が *in vitro* 及び *in vivo* で PRTS をコードする DNA 配列の転写によって生成され得る、このような DNA 配列は T7 または SP6 等の好適な RNA ポリメラーゼプロモータを用いて、種々のベクターの中に組み入れることが可能である。別法では、相補的な RNA を構成的または誘導的に合成するこれらの cDNA 作製物は、細胞株、細胞、または組織の中に導入することができる。

10

【0192】

RNA 分子を修飾することによって、細胞内の安定性を高め、半減期を長くすることができる。可能な修飾には、分子の 5' 及び / または 3' 端部でのフランキング配列の追加、または分子のバックボーン内のホスホジエステル結合の代わりにホスホロチオネートまたは 2' Oメチルを用いる修飾が含まれるが、これらに限定されるものではない。PNA の生成に固有のこの概念は、内在性のエンドヌクレアーゼによって容易には認識されないアデニン、シチジン、グアニン、チミン、及びウリジンのアセチル -、メチル -、チオ -、及び同様の修飾形態だけでなく、イノシン、キューエオシン (queosine)、及びワイプトシン (wybutosine) などの従来のものでない塩基を含めることによって、これらの分子の全体に拡大することができる。

20

【0193】

本発明の更なる実施例は、PRTS をコードするポリヌクレオチドの発現の変化に有効な化合物をスクリーニングする方法を含む。特定のポリヌクレオチドの発現の変化に有効な化合物には、限定するものではないが、特定のポリヌクレオチド配列と相互作用可能な非高分子化学物質、オリゴヌクレオチド、アンチセンスオリゴヌクレオチド、三重らせん形成オリゴヌクレオチド、転写因子やその他のポリペプチド転写調節因子が含まれる。有効な化合物は、ポリヌクレオチド発現のインヒビター或いはエンハンサーとして作用し、ポリヌクレオチドの発現を変化させ得る。従って、PRTS の発現または活性の増加に関連する疾患の治療においては、PRTS をコードするポリヌクレオチドの発現を特異的に阻害する化合物が治療上有用であり、PRTS の発現または活性の低下に関連する疾患の治療においては、PRTS をコードするポリヌクレオチドの発現を特異的に促進する化合物が治療上有用であり得る。

30

【0194】

特定のポリヌクレオチドの発現の変化の有効性を調べるために、少なくとも1個から複数個の試験化合物をスクリーニングすることができる。試験化合物は、有効な化合物の化学修飾を含む当分野で周知の任意の方法で得ることができる。このような方法は、ポリヌクレオチドの発現を変化させる場合、一般に市販されている或いは専売の天然または非天然の化合物ライブラリから選択する場合、標的ポリヌクレオチドの化学的及び / または構造的な特性に基づいて化合物を合理的にデザインする場合、更に組合せ的にまたは無作為に生成した化合物のライブラリから選択する場合に有効である。PRTS をコードするポリヌクレオチドを含むサンプルは、少なくとも1つの試験化合物に曝露して得る。サンプルには、例えば無傷細胞、透過化処理した細胞、*in vitro* 細胞遊離系または再構成生化学系が含まれ得る。PRTS をコードするポリヌクレオチドの発現における変化は、当分野で周知の任意の方法でアッセイする。通常、PRTS をコードするポリヌクレオチドの配列に相補的なヌクレオチド配列を有するプローブを用いたハイブリダイゼーションにより、特定のヌクレオチドの発現を検出する。ハイブリダイゼーションの収量を定量し、その値が1或いは複数の試験化合物に曝露される及び曝露されないポリヌクレオチドの発現の比較における基準となり得る。試験化合物に曝露されるポリヌクレオチドの発現の変化が検出される場合は、ポリヌクレオチドの発現の変化に試験化合物が有効であることを

40

50

示している。特定のポリヌクレオチドの発現の変化に有効な化合物を調べるために、例えば Schizosaccharomyces pombe 遺伝子発現系 (Atkins, D. 他 (1999) 米国特許第 5,932,435 号、Arndt, G.M. 他 (2000) *Nucleic Acids Res.* 28:E15) または HeLa 細胞等のヒト細胞株 (Clarke, M.L. 他 (2000) *Biochem. Biophys. Res. Commun.* 268:8-13) を用いてスクリーニングする。本発明の特定の実施例は、特異的ポリヌクレオチド配列に対するアンチセンス活性を調べるための、各オリゴヌクレオチド (デオキシリボヌクレオチド、リボヌクレオチド、ペプチド核酸、及び修飾オリゴヌクレオチド) の組み合わせライブラリのスクリーニングを含む (Bruice, T.W. 他 (1997) 米国特許第 5,686,242 号、Bruice, T.W. 他 (2000) 米国特許第 6,022,691 号)。

【0195】

ベクターを細胞または組織に導入する多数の方法が利用でき、in vivo、in vitro、及び ex vivo での使用に等しく適している。ex vivo での治療の場合、患者から採取された肝細胞の中にベクターを導入して、自家移植で同じ患者に戻すためにクローニング増殖される。トランスフェクション、リボソーム注入またはポリカチオンアミノポリマーによる運搬は、当分野で周知の方法を用いて実行することができる (例えば、Goldman, C.K. 他 (1997) *Nature Biotechnology* 15:462-66:を参照)。

【0196】

上記したいかなる治療方法も、例えば、ヒト、イヌ、ネコ、ウシ、ウマ、ウサギ及びサルなどの哺乳動物を含む、治療が必要な全ての被験者に適用できる。

【0197】

本発明の別の実施例は、上記した全ての治療効果のために、医学上認められる担体と共に医薬品或いは無菌組成物の投与に関連する。このような組成物は、PRTS、PRTS の抗体、擬態、アゴニスト、アンタゴニスト、または PRTS のインヒビターなどからなる。この組成物は、単体で、或いは安定剤などの 1 種類以上の別の薬剤と共に、無菌の生体適合性医薬品担体に投与することができる。このような医薬品担体には、生理食塩水、緩衝食塩水、ブドウ糖、及び水などが含まれるがこれらに限定されるものではない。この組成物は、単独或いは薬物またはホルモンなどの別の薬剤と共に投与することができる。

【0198】

本発明に用いられる組成物は、様々な経路を用いて投与するが可能である。この経路には、経口、静脈内、筋肉内、動脈内、骨髄内、クモ膜下、心室内、経皮、皮下、腹腔内、鼻腔内、腸内、局所、舌下、または直腸が含まれるがこれらに限定されるものではない。

【0199】

肺投与用の組成物は、液状または乾燥粉末状に調製することができる。このような組成物は通常、患者が吸入する直前にエアロゾル化する。小分子 (例えば、従来の低分子量有機薬剤) の場合には、速効製剤のエアロゾル輸送が当分野で周知である。高分子 (例えばより大きなペプチドやタンパク質) の場合には、肺の肺泡領域を介する肺輸送の技術が近年向上したため、インスリン等の薬剤を実際に血中に輸送することが可能となった (Patton, J.S. 他, 米国特許第 5,997,848 号等を参照)。肺輸送は、針注射を用いないで投与できるという点で優れており、潜在的に有毒な浸透エンハンサーが必要でなくなる。

【0200】

本発明に用いる好適な組成物には、目的を達成するため、効果的な量の活性処方成分を含む組成物が含まれる。当業者は、十分に自身の能力で効果的な服用量を定めることができる。

【0201】

PRTS またはその断片を含む高分子を直接細胞内に輸送するべく、特殊な形態に組成物が調製されるのが好ましい。例えば、細胞不透過性高分子を含むリボソーム製剤は、細胞

融合及び高分子の細胞内輸送を促進し得る。別法では、PRTSまたはその断片をHIV Tat-1タンパク質の陽イオンN末端部に結合することもできる。このようにして作製された融合タンパク質は、マウスモデル系の脳を含む全ての組織の細胞に形質導入されることが確認されている(Schwarze, S. R. 他(1999) Science 285:1569-1572)。

【0202】

どのような組成物であっても、治療に効果的な薬用量は、初めは、例えば腫瘍細胞の腫瘍細胞アッセイで、或いは動物モデルのどちらかで推定することができる。通常、動物モデルには、マウス、ウサギ、イヌ、サル、またはブタなどが用いられる。動物モデルはまた、好適な濃縮範囲及び投与の経路を決めるのに用いることができる。このような治療をもとに、ヒトへの有益な薬用量及び投与経路を決定することができる。

10

【0203】

医学的に効果的な薬用量は、症状や容態を回復させる、たとえばPRTSまたはその断片、PRTSの抗体、PRTSのアゴニストまたはアンタゴニスト、インヒビターなどの活性処方成分の量に関連する。薬用有効度及び毒性は、たとえば、ED₅₀ (服用に対して集団の50%に医薬的效果がある用量)またはLD₅₀ (服用に対して集団の50%に致命的である用量)統計を計算するなど、細胞培養または動物実験における標準的な薬剤手法によって決定することができる。毒性効果と治療効果との薬用量比は治療指数であり、LD₅₀/ED₅₀と示すことができる。高い治療指数を示す組成物が望ましい。細胞培養アッセイ及び動物実験から得られたデータが、ヒトへの適用のために、薬用量の範囲を調剤するのに用いられる。このような組成物が含まれる薬用量は、毒性を殆ど或いは全く含まず、ED₅₀を含む血中濃度の範囲であることが望ましい。薬用量は、用いられる投与形態及び患者の感受性、投与の経路によって、この範囲内で様々である。

20

【0204】

正確な薬用量は、治療が必要な患者に関する要素を考慮して、実務者によって決められるであろう。薬用量及び投与は、効果的なレベルの活性成分を与えるため或いは所望の効果を維持するために調節される。薬用量の要素として考慮されるものには、疾患の重症度、患者の一般的な健康状態、年齢、体重、及び患者の性別、投与の時間及び頻度、併用する薬剤、反応感受性、及び治療に対する応答が含まれる。作用期間が長い組成物は、三日か四日に一度、一週間に一度、二週間に一度、特定の製剤の半減期及びクリアランス率によって左右され、投与され得る。

30

【0205】

通常薬用量は投与の経路によって異なるが、約0.1~100,000µgまでの最大約1グラムまでである。特定の薬用量及び運搬の方法に関するガイダンスは文献に記載されており、一般に当分野の実務者はそれを利用することができる。当業者は、タンパク質またはインヒビターとは異なったヌクレオチドの製剤を利用するであろう。同様に、ポリヌクレオチドまたはポリペプチドの運搬は、特定の細胞、状態、位置などに対して特異的であろう。

【0206】

(診断)

別の実施例では、PRTSに特異的に結合する抗体が、PRTSの発現によって特徴付けられる疾患の診断、またはPRTSやPRTSのアゴニストまたはアンタゴニスト、インヒビターで治療を受けている患者をモニターするためのアッセイに用いられる。診断に有用な抗体は、治療のところで記載した方法と同じ方法で製剤される。PRTSの診断アッセイには、抗体及び標識を用いてヒトの体液或いは細胞や組織から採取されたものからPRTSを検出する方法が含まれる。これらの抗体は、修飾をして或いはしないで使用され、レポーター分子の共有結合性或いは非共有結合性の接着によって標識化され得る。当分野で周知の種々のレポーター分子が用いられるが、その内の幾つかは上記した。

40

【0207】

PRTSを測定するためのELISA,RIA,及びFACSを含む種々のプロトコルは

50

、当分野では周知であり、変わった或いは異常なレベルのPRTSの発現を診断する元となるものを提供する。正常或いは標準的なPRTSの発現の値は、複合体の形成に適した条件の下、正常な哺乳動物、例えばヒトなどの被験者から採取した体液または細胞とPRTSに対する抗体とを結合させることによって決定する。標準的な複合体形成の量は、測光法 (photometric) などの種々の方法で定量され得る。被験者のPRTSの発現の量、制御及び疾患、生検組織からのサンプルが基準値と比較される。基準値と被験者との間の偏差が、診断の指標となる。

【0208】

本発明の別の実施例によれば、PRTSをコードするポリヌクレオチドを診断のために用いることもできる。用いられるポリヌクレオチドには、オリゴヌクレオチド配列、相補的なRNA及びDNA分子、及びPNAが含まれる。このポリヌクレオチドを用いて、疾患と相関し得るPRTSを発現する生検組織における遺伝子の発現を検出し定量する。この診断アッセイを用いて、PRTSの存在の有無、更に過剰な発現を調べ、治療中のPRTS値の調節を監視する。

10

【0209】

一実施形態では、PRTSまたは近縁の分子をコードする遺伝子配列を含むポリヌクレオチド配列を検出可能なPCRプローブを用いたハイブリダイゼーションによって、PRTSをコードする核酸配列を同定することが可能である。例えば5'調節領域である高度に特異的な領域か、例えば保存されたモチーフであるやや特異性の低い領域から作られているかのプローブの特異性と、ハイブリダイゼーション或いは増幅のストリンジェントは、

20

【0210】

プローブはまた、関連する配列の検出に利用され、PRTSをコードする任意の配列と少なくとも50%の配列同一性を有し得る。目的の本発明のハイブリダイゼーションプローブには、DNAあるいはRNAが可能であり、SEQ ID NO: 15 - 28の配列、或いはPRTS遺伝子のプロモーター、エンハンサー、イントロンを含むゲノム配列に由来し得る。

【0211】

PRTSをコードするDNAに対して特異的なハイブリダイゼーションプローブの作製方法には、PRTS及びPRTS誘導体をコードするポリヌクレオチド配列をmRNAプローブの作製のためのベクターにクローニングする方法がある。このようなベクターは市販されており、当業者には周知であり、好適なRNAポリメラーゼ及び好適な標識されたヌクレオチドを加えることによって、in vitroでRNAプローブを合成するために用いられる。ハイブリダイゼーションプローブは、例えば³²P或いは³⁵Sなどの放射性核種、或いはアビジン/ビオチン (biotin) 結合系によってプローブに結合されたアルカリホスファターゼなどの酵素標識等の種々のレポーターの集団によって標識され得る。

30

【0212】

PRTSをコードするポリヌクレオチド配列を用いて、PRTSの発現に関連する疾患を診断することが可能である。限定するものではないが、このような疾患には胃腸疾患、心血管疾患、自己免疫/炎症性の疾患、細胞増殖異常、発生障害、上皮の疾患、神経の疾患、および生殖障害が含まれ、胃腸疾患の中には、嚥下障害、消化性食道炎、食道痙攣、食道狭窄、食道癌、消化不良、消化障害、胃炎、胃癌、食欲不振、悪心、嘔吐、胃不全麻痺、洞または幽門の浮腫、腹部アンギナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリー ヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、及び後天性免疫不全症候群 (AIDS) 腸症、黄疸、肝性脳症、肝腎症候群、肝炎、肝脂肪症、血色素症、ウィルソン病、

40

50

1 - アンチトリプシン欠損症、ライ症候群、原発性硬化性胆管炎、肝梗塞、門脈循環閉塞及び血栓、小葉中心壊死、肝臓紫斑病、肝静脈血栓、肝静脈閉塞症、子癇前症、子癇、妊娠性急性肝脂肪、妊娠性肝臓内胆汁うっ滞と、結節性再生及び腺腫、癌腫を含む肝癌とが含まれ、心血管疾患の中には、動静脈瘤、アテローム硬化、高血圧、脈管炎、レイノー病、静脈奇形、動脈解離、静脈瘤、血栓静脈炎及び静脈血栓、血管の腫瘍、血栓崩壊の合併症、バルーン血管形成術 (balloon angioplasty)、血管置換術 (vascular replacement)、大動脈冠動脈バイパス術移植手術 (coronary artery bypass graft surgery)、うっ血性心不全、虚血性心疾患、狭心症、心筋梗塞、高血圧性心疾患、変性弁膜性心疾患、石灰化大動脈弁狭窄症、先天性2尖大動脈弁、僧帽弁輪状石灰化 (mitral annular calcification)、僧帽弁脱出、リウマチ熱、リウマチ性心疾患、感染性心内膜炎、非細菌性血栓性心内膜炎、全身性エリテマトーデスの心内膜炎、カルチノイド心疾患、心筋症、心筋炎、心膜炎、腫瘍性心疾患、先天性心臓疾患、先天性肺異常 (congenital lung anomalies)、肺拡張不全、肺うっ血及び肺水腫、肺動脈血栓症、肺出血、肺梗塞、肺高血圧症、血管硬化症、閉塞性肺疾患、拘束性肺疾患 (restrictive pulmonary disease)、慢性閉塞性肺疾患、肺気腫、慢性気管支炎、気管支喘息、細気管支拡張症、細菌性肺炎、ウイルス性肺炎及びマイコプラズマ肺炎、肺膿瘍、肺結核、びまん性間質性疾患 (diffuse interstitial diseases)、塵肺症、サルコイド症、特発性肺繊維症 (idiopathic pulmonary fibrosis)、剥離性間質性肺炎、過敏症肺炎 (hypersensitivity pneumonitis)、肺好酸球増加閉塞性細気管支炎 器質性肺炎 (pulmonary eosinophilia bronchiolitis obliterans - organizing pneumonia)、びまん性肺出血症候群 (diffuse pulmonary hemorrhage syndromes)、グッドパスチャー症候群、特発性肺血鉄症、肺併発膠原血管病 (pulmonary involvement in collagen-vascular disorders)、肺胞たんぱく症、肺腫瘍、炎症性及び非炎症性胸水 (inflammatory and noninflammatory pleural effusions)、気胸症、胸膜腫瘍、薬物による肺疾患 (drug-induced lung disease)、放射線による肺疾患 (radiation-induced lung disease) 及び肺移植の合併症などが含まれ、自己免疫/炎症性の疾患の中には、炎症及び日光性角化症、後天性免疫不全症候群 (AIDS) 及び副腎機能不全、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、アテローム斑破裂、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー (APECED)、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時的リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、関節軟骨の分解、骨粗しょう症、膵炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ、細胞増殖異常の中には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病 (MCTD)、骨髓線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髓腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髓、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、精巣、胸腺、甲状腺、及び子宮の癌が含まれ、発生障害も含まれ、その中には尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ

10

20

30

40

50

ベッカー型筋ジストロフィー、骨吸収、癩癧、性腺形成異常、WAGR症候群（ウィルムス腫瘍、無虹彩症、尿生殖器異常、精神薄弱）、スミス マジェニス症候群（Smith-Magenis syndrome）、脊髄形成異常症候群、遺伝性粘膜上皮異形成、遺伝性角皮症、シャルコー マリー ツース病及び神経線維腫症などの遺伝性神経病、甲状腺機能低下症、水頭症、Sydenham 舞蹈病（Sydenham's chorea）及び脳性小児麻痺などの発作障害、脊髄二分裂、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、加齢性黄斑変性症、感覚神経性聴力損失が含まれ、上皮の疾患の中には、汗疱状湿疹、アレルギー性接触皮膚炎、角化上皮症、黒皮症、白斑、日光性角化症、基底細胞癌、扁平上皮癌、脂漏性角化症、毛包炎、単純ヘルペス、帯状疱疹、水痘、カンジダ症、皮膚糸状菌症、疥癬、昆虫刺症、老年性血管腫、ケロイド、皮膚線維腫、先端線維性軟朮、じんま疹、一過性棘融解性皮膚症、乾燥症、湿疹、アトピー性皮膚炎、接触性皮膚炎、手湿疹、貨幣状湿疹、ピダール苔癬、asteatotic eczema、うつ血性皮膚炎、うつ血性潰瘍、脂漏性皮膚炎、乾癬、扁平苔癬、バラ色粧糖疹、膿痂疹、膿瘡、皮膚糸状菌症、なまず、疣贅、尋常性座瘡、赤鼻、尋常性天疱瘡、落葉状天疱瘡、腫瘍随伴性天疱瘡、水疱性類天疱瘡、妊娠性疱疹、疱疹状皮膚炎、線形性IgA疾患、後天性表皮水疱症、皮膚筋炎、紅斑性狼瘡、強皮症および限局性強皮症、紅皮症、脱毛症、定形皮膚障害（figurate skin lesions）、血管拡張症、色素沈着低下、色素沈着過度、小疱疹/水疱、発疹、皮膚薬物反応、丘疹小結節皮膚障害（papulonodular skin lesions）、慢性非治癒性創傷、感光性疾患、単純表皮水泡症、表皮剥離および非表皮剥離性掌蹠角皮症、ジューメンス水疱魚鱗癬（ichthyosis bullosa of Siemens）、剥脱性魚鱗癬、手掌足底角化症、掌蹠角皮症（keratosis palmoplantarisまたはpalmoplantar keratoderma）、keratosis punctata、Meesmann's corneal dystrophy、先天性爪肥厚症、白色海綿母斑、多発性皮脂嚢腫、表皮性母斑/表皮剥離性角化型、monilethrix、trichothiodystrophy、慢性肝炎/特発性肝硬変、結腸直腸過形成が含まれ、神経の疾患の中には、癩癧、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクローイツフェルト ヤコブ病、ゲルストマン症候群、Gerstmann-Straussler-Scheinker症候群を含むプリオン病と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫（cerebelloretinal hemangioblastomatosis）、脳3叉神経血管症候群、ダウン症を含む中枢神経系性精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性の障害、分裂病性疾患を含む精神病と、季節性障害（SAD）と、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病と、進行性核上麻痺、皮質基部変性（corticobasal degeneration）及び家族性の前頭側頭性健忘症とが含まれ、生殖障害の中には、プロラクチンの産生異常と、卵管病及び排卵異常、子宮内膜症、発情期異常、月経周期異常、多嚢胞卵巣症候群、卵巣過剰刺激症候群、子宮内膜癌及び卵巣癌、子宮筋腫、自己免疫異常、異所性妊娠、奇形発生を含む不妊症と、乳癌、線維嚢胞性乳腺症、乳漏症と、精子形成異常、生理学上の精子異常、精巣癌、前立腺癌、良性の前立腺過形成、前立腺炎、ペーロニー病、インポテンス、男性乳房及び女性化乳房癌とが含まれる。PRTSをコードするポリヌクレオチド配列は、サザーン法やノーザン法、ドットプロット法、或いはその他の膜系の技術、PCR法、ディップスティック（dipstick）、ピン（pin）、ELISA式アッセイ、及

10

20

30

40

50

び変異 P R T S の発現を検出するために患者から採取した体液或いは組織を利用するマイクロアレイに使用することが可能である。このような質的或いは量的方法は、当分野では周知である。

【0213】

ある実施態様では、P R T S をコードするヌクレオチド配列は、関連する疾患、特に上記した疾患を検出するアッセイにおいて有用であろう。P R T S をコードするヌクレオチド配列は、標準的な方法で標識化され、ハイブリダイゼーション複合体の形成に好適な条件の下、患者から採取した体液或いは組織のサンプルに加えることができるであろう。好適な培養期間の後、サンプルを洗浄し、シグナルを定量して基準値と比較する。患者のサンプルのシグナルの量が、制御サンプルと比べて著しく変わっている場合は、サンプル内の P R T S をコードするヌクレオチド配列の変異レベルにより、関連する疾患の存在が明らかになる。このようなアッセイを用いて、動物実験、臨床試験、或いは個人の患者の治療を監視における、特定の治療効果を推定することが可能である。

10

【0214】

P R T S の発現に関連する疾患の診断の基準となるものを提供するために、発現の正常すなわち標準的なプロファイルが確立される。これは、ハイブリダイゼーション或いは増幅に好適な条件の下、動物或いはヒトの何れかの正常な被験者から抽出された体液或いは細胞と、P R T S をコードする配列或いはその断片とを結合させることにより達成され得る。標準的なハイブリダイゼーションは、正常な被験者から得た値と周知の量の実質的に精製されたポリヌクレオチドが用いられる実験からの値とを比較することによって定量可能である。正常なサンプルから得た標準的な値を、疾患の症状を示す被験者から得た値と比較可能である。基準値と被験者の値との偏差を用いて罹患しているかどうかを決定する。

20

【0215】

疾患の存在が確定され、治療プロトコルが開始されると、ハイブリダイゼーションアッセイを通常ペースで繰り返して、被験者における発現のレベルが正常な患者に示される値に近づき始めたかどうかを推定することが可能である。繰り返し行ったアッセイの結果を、数日から数ヶ月の期間の治療の効果を見るのに用いることができる。

【0216】

癌では、個体からの生体組織における異常な量の転写物が、疾患の発生の素因を示し、また実際に臨床的症状が出る前に疾患を検出する方法を提供することが可能である。この種のより明確な診断により、医療の専門家が予防方法或いは積極的な治療法を早くから利用して、癌の発生または進行を防ぐことが可能となる。

30

【0217】

P R T S をコードする配列から設計されたオリゴヌクレオチドのさらなる診断への利用には、P C R の利用が含まれ得る。このようなオリゴマーは、化学的な合成、酵素を用いた生成、或いは *in vitro* で生成され得る。オリゴマーは、好ましくは P R T S をコードするポリヌクレオチドの断片、或いは P R T S をコードするポリヌクレオチドと相補的なポリヌクレオチドの断片を含み、最適な条件の下、特定の遺伝子や条件を識別するために利用される。また、オリゴマーは、やや緩いストリンジェントな条件の下、近縁の D N A 或いは R N A 配列の検出及び/または定量のため用いることが可能である。

40

【0218】

或る実施態様において、P R T S をコードするポリヌクレオチド配列由来のオリゴヌクレオチドプライマーを用いて、一塩基多型 (S N P) を検出し得る。S N P は、ヒトの先天性または後天性遺伝病の原因となる場合が多いヌクレオチドの置換、挿入及び欠失である。限定するものではないが、S N P の検出方法には、一本鎖立体構造多型 (S S C P) 及び蛍光 S S C P (f S S C P) 法が含まれる。S S C P では、P R T S をコードするポリヌクレオチド配列由来のオリゴヌクレオチドプライマーを用いたポリメラーゼ連鎖反応 (P C R) で D N A を増幅する。この D N A は、例えば病変或いは正常な組織、生検サンプル、体液等に由来し得る。この D N A 内の S N P は、一本鎖形状の P C R 産物の 2 次及び 3 次構造に差異を生じさせる。この差異は非変性ゲル中でのゲル電気泳動法を用いて検出

50

可能である。f S C C Pでは、オリゴヌクレオチドプライマーを蛍光標識することによって、DNAシーケンシング装置などのハイスループット機器でアンプリマー (a m p l i m e r) の検出をすることが可能になる。更に、インシリコSNP (i n s i l i c o S N P : i s S N P) と呼ばれる配列データベース分析法は、共通のコンセンサス配列の構築に用いられる個々の重複するDNA断片の配列を比較することによって、多型を同定することができる。これらのコンピュータベースの方法は、DNA配列クロマトグラムの自動分析及び統計モデルを用いたシーケンシングエラーや研究室でのDNAの調整に起因する配列のばらつきを排除する。別法では、例えばハイスループットのM A S S A R R A Y システム (S e q u e n o m , I n c . , S a n D i e g o C A) を用いた質量分析によりSNPを検出し、特徴付ける。

10

【0219】

P R T S の発現を定量するために用いられ得る方法には、ヌクレオチドの放射標識或いはビオチン標識、調節核酸の相互増幅 (c o a m p l i f i c a t i o n) 、及び標準的な曲線に結果が加えられたものが含まれる (例えば、M e l b y , P . C . ら (1 9 9 3) J . I m m u n o l . M e t h o d s , 1 5 9 : 2 3 5 - 4 4 ; D u p l a a , C . ら (1 9 9 3) A n a l . B i o c h e m . 2 2 9 - 2 3 6 を参照) 。多数のサンプルの定量速度は、ハイスループット型のアッセイを用いることで速くなるであろう。このアッセイでは、目的のオリゴマーやポリヌクレオチドが様々な希釈液中に含まれ、分光光度法或いは非色応答によって定量が迅速である。

【0220】

更に別の実施例では、本明細書に記載した任意のポリヌクレオチド配列に由来するオリゴヌクレオチドまたはより長い断片を、マイクロアレイにおける標的として用いることができる。マイクロアレイを、上記したように多数の遺伝子の相対的な発現レベルを同時にモニタリングする転写イメージング技術に用いることができる。マイクロアレイはまた、遺伝子変異、突然変異及び多型の同定に用いることができる。この情報を用いて、遺伝子機能を決定し、疾患の遺伝的根拠を解明し、疾患を診断し、遺伝子発現に関連する疾病の進行/後退をモニタリングし、疾患の治療における治療薬の開発や活性のモニタリングを行うことができる。特に、患者にとって最適かつ有効な治療法を選択するために、この情報を用いて患者の薬理ゲノムプロファイルを作成することができる。例えば、患者の薬理ゲノムプロファイルに基づいて、患者に対して極めて効果的でありながら副作用を殆ど示さない治療薬を選択することができる。

20

30

【0221】

別の実施例では、P R T S 、 P R T S の断片、P R T S に特異的な抗体をマイクロアレイ上のエレメントとして用いることができる。マイクロアレイを用いて、上記のようにタンパク質間相互作用、薬剤 - 標的相互作用及び遺伝子発現プロファイルをモニタリング及び測定することが可能である。

【0222】

特定の実施例は、或る組織または細胞型の転写イメージを生成する本発明のポリヌクレオチドの使用に関連する。転写イメージは、特定の組織または細胞型により遺伝子発現の包括的パターンを表す。包括的遺伝子発現パターンは、所定の条件下で所定の時間に発現した遺伝子の数及び相対存在量を定量することにより分析される (S e i l l i a m e r 他、米国特許第5,840,484号の " C o m p a r a t i v e G e n e T r a n s c r i p t A n a l y s i s " を参照。この特許に言及することを以って本明細書の一部とする) 。従って、特定の組織または細胞型の転写物または逆転写物の全てに本発明のポリヌクレオチドまたはその相補配列をハイブリダイズすることにより、転写イメージが生成され得る。或る実施例では、本発明のポリヌクレオチドまたはその相補配列がマイクロアレイ上に複数のエレメントのサブセットを構成するハイスループット型でハイブリダイゼーションさせる。結果として得られる転写イメージは、遺伝子活性のプロファイルとなり得る。

40

【0223】

50

転写イメージは、組織、細胞株、生検サンプル、またはその他の生体サンプルから単離した転写物を用いて生成し得る。従って、転写イメージは、組織または生検サンプルの場合には in vivo、または細胞株の場合には in vitro における遺伝子発現を反映する。

【0224】

本発明のポリヌクレオチドの発現プロファイルを示す転写イメージはまた、合成化合物または天然化合物の毒性試験のみならず、in vitroモデル系及び薬剤の前臨床評価に関連して使用され得る。全ての化合物は、作用及び毒性の機構を示唆する、頻繁に分子フィンガープリント若しくは毒性シグネチャ (signature) と称されるような特徴的な遺伝子発現パターンを引き起こす (Nuwaysir, E. F. 他 (1999) Mol. Carcinog. 24:153-159、Steiner, S. and N. L. Anderson (2000) Toxicol. Lett. 112-113:467-471、また言及することを以って本明細書の一部とする)。試験化合物が、毒性を有する既知の化合物のシグネチャと同一のシグネチャを有する場合には、毒性特性を共有している可能性が高い。フィンガープリンまたはシグネチャが、より多くの遺伝子及び遺伝子ファミリーからの発現情報を含んでいれば、より有用かつ正確になる。理想としては、発現のゲノム全域にわたって測定し、最高品質のシグネチャを提供することである。任意の試験化合物によっても発現が変化しない遺伝子も同様に重要である。それは、これらの遺伝子の発現レベルを用いて残りの発現データを標準化することができるためである。標準化処理は、異なる化合物で処理した後の発現データの比較に有用である。毒性シグネチャのエLEMENTへの遺伝子機能を割り当てることは毒性機構の解明に役立つが、毒性の予測につながるシグネチャの統計的な一致には遺伝子機能の知識は必要ではない (例えば2000年2月29日にNational Institute of Environmental Health Sciencesより発行されたPress Release 00-02を参照されたい。これについては<http://www.niehs.nih.gov/oc/news/toxchip.htm>で入手可能である)。従って、毒性シグネチャを用いる毒性スクリーニングにおいて、全ての発現した遺伝子配列を含めることは重要でありまた望ましいことである。

【0225】

一実施例では、試験化合物の毒性は、核酸を含有する生体サンプルをその試験化合物で処理して評価する。処理した生体サンプル中で発現した核酸は、本発明のポリヌクレオチドに特異的な1若しくは複数のプローブでハイブリダイズさせ、それによって本発明のポリヌクレオチドに対応する転写レベルを定量することができる。処理した生体サンプル中の転写レベルを、非処理生体サンプル中のレベルと比較する。両サンプルの転写レベルの差が、処理されたサンプル中で試験化合物が引き起こす毒性反応を示唆する。

【0226】

別の実施例は、本発明のポリペプチド配列を用いて組織または細胞型のプロテオームを分析することに関連する。「プロテオーム」という用語は、或る特定の組織または細胞型におけるタンパク質発現の包括的パターンを指す。プロテオームを構成する各タンパク質は、個々に更なる分析をすることができる。プロテオーム発現パターン即ちプロファイルは、所定の条件下で所定の時間に発現したタンパク質の数及びそれらの相対的な存在量を定量することにより分析する。従って、ある細胞のプロテオームのプロファイルは、特定の組織または細胞型のポリペプチドを分離及び分析することにより作成し得る。或る実施例では、このような分離は2次元ゲル電気泳動によって行う。この2次元ゲル電気泳動法では、まず、1次元の等電点電気泳動によりサンプルからタンパク質を分離し、次に、2次元のドデシル硫酸ナトリウムスラブゲル電気泳動により分子量に従って分離する (前出のSteiner and Anderson)。これらのタンパク質は、通常クーマシーブルーまたはシルバーまたは蛍光染色などの染色剤を用いてゲルを染色して、分散した個別の位置にあるスポットとしてゲル中で可視化される。各タンパク質スポットの光学密度は、通常サンプル中のタンパク質レベルに比例する。異なるサンプル、例えば試験化合物ま

10

20

30

40

50

たは治療薬で処理済みまたは未処理のいずれかの生体サンプルから得られる等位置にあるタンパク質スポットの光学密度を比較し、処理に関連するタンパク質スポット密度の変化を調べる。スポット内のタンパク質は、例えば化学的または酵素的に切断した後、質量分析する標準的な方法を用いて部分的にシークエンシングする。スポット内のタンパク質の同一性は、好適には少なくとも5個の連続するアミノ酸残基であるその部分的な配列を、本発明のポリペプチド配列と比較することにより決定し得る。場合によっては、決定的なタンパク質同定のための更なる配列が得られる。

【0227】

プロテオームのプロファイルは、PRTSに特異的な抗体を用いてPRTS発現レベルを定量することによっても作成可能である。或る実施例では、マイクロアレイ上のエレメントとして抗体を用い、マイクロアレイをサンプルに曝露して各アレイエレメントへのタンパク質結合レベルを検出することによりタンパク質発現レベルを定量する(Lueking, A.ら。(1999) Anal. Biochem. 270:103-111、Mendoze, L.G.ら。(1999) Biotechniques 27:778-788)。検出は当分野で既知の様々な方法で行うことができ、例えば、チオール反応性またはアミノ反応性蛍光化合物を用いてサンプル中のタンパク質を反応させ、各アレイエレメントにおける蛍光結合の量を検出し得る。

10

【0228】

プロテオームレベルでの毒性シグネチャも中毒学的スクリーニングに有用であり、転写レベルでの毒性シグネチャと並行して分析するべきである。或る組織における或るタンパク質では、転写物の存在量とタンパク質の存在量との相関性が低いことがあるため(Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537)、プロテオーム毒性シグネチャは、転写イメージにはそれ程影響しないがプロテオームのプロファイルを変化させる化合物の分析において有用たり得る。更に、体液中での転写の分析は、mRNAが急速に分解するため困難である。したがって、このような場合にはプロテオームのプロファイル作成はより信頼でき、情報価値がある。

20

【0229】

別の実施例では、試験化合物の毒性は、タンパク質を含む生体サンプルをその試験化合物で処理して評価する。処理された生体サンプル中で発現したタンパク質を分離して、各タンパク質の量が定量できるようにする。各タンパク質の量を、未処理生体サンプル中の対応するタンパク質の量と比較する。両サンプル中のタンパク質の量の差は、処理されたサンプル中の試験化合物に対する毒性反応を示唆する。個々のタンパク質は、それらのアミノ酸残基をシークエンシングし、これらの部分配列を本発明のポリペプチドと比較することで同定する。

30

【0230】

別の実施例では、試験化合物の毒性は、タンパク質を含む生体サンプルをその試験化合物で処理することにより評価する。生体サンプルから得たタンパク質を、本発明のポリペプチドに特異的な抗体と共にインキュベートする。その抗体により認識されたタンパク質の量を定量する。処理された生体サンプル中のタンパク質の量を、未処理生体サンプル中のタンパク質の量と比較する。両サンプルのタンパク質量の差が、処理サンプル中の試験化合物に対する毒性反応を示唆する。

40

【0231】

当分野で周知の方法でマイクロアレイを準備して使用し、分析する。(例えば、Brennan, T.M. 他(1995) 米国特許第5,474,796号; Schena, M. 他(1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler 他(1995) PCT出願番号WO95/251116; Shalon, D. 他(1995) PCT出願番号WO95/35505; Heller, R.A. 他(1997) Proc. Natl. Acad. Sci. 94:2150-2155; 及び Heller, M.J. 他(199

50

7) 米国特許第5,605,662号を参照)。様々なタイプのマイクロアレイが周知であり、詳細については、DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, Londonに記載されている。また、この文献を引用することを以って本明細書の一部とする。

【0232】

本発明の別の実施例ではまた、PRTSをコードする核酸配列を用いて、天然のゲノム配列をマッピングするのに有用なハイブリダイゼーションプローブを作製することが可能である。コーディング配列または非コーディング配列の何れかを用いることができるが、或る例では、コーディング配列より非コード配列が好ましい。例えば、多重遺伝子ファミリーのメンバー間にコーディング配列が保存されていることにより、染色体マッピング時に望ましくない交差ハイブリダイゼーションが生じる可能性がある。この配列は、特定の染色体、染色体の特定領域または人工の染色体、例えば、ヒト人工染色体(HAC)、酵母人工染色体(YAC)、細菌人工染色体(BAC)、細菌P1産物、或いは単一染色体cDNAライブラリに対してマッピングされる(Harrington, J. J.ら(1997) Nat Genet. 15:345-355、Price, C. M. (1993) Blood Rev. 7:127-134、Trask, B. J. (1991) Trends Genet. 7:149-154等を参照)。一度マッピングすると、本発明の核酸配列を用いて、例えば病状の遺伝と特定の染色体領域やまたは制限断片長多型(RFLP)の遺伝とが相関するような遺伝子連鎖地図を作成可能である(Lander, E. S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357を参照)。

【0233】

in situ 蛍光ハイブリダイゼーション(FISH)は、他の物理的及び遺伝子地図データと相関し得る(例えば、Heinz-Ulrich, 他による(1995) in Meyers, 前出, pp. 965-968を参照)。遺伝子地図データの例は、種々の科学誌あるいはOnline Mendelian Inheritance in Man(OMIM)のワールドワイドウェブのサイトで見付けることができる。物理的な染色体地図上のPRTSをコードする遺伝子の位置と特定の疾患との相関性、或いは特定の疾患に対する素因が、このような疾患と関連するDNA領域の決定に役立つため、更なる位置を決定するクローニングが行われる。

【0234】

染色体標本のin situハイブリダイゼーション、及び確定した染色体マーカーを用いた結合分析などの物理的マッピング技術を用いて、遺伝子地図を拡張することもできる。マウスなどの別の哺乳動物の染色体上に遺伝子を配置させることにより、たとえ正確なヒト染色体の位置が分かていなくても、関連するマーカーが明らかになる場合が多い。この情報は、位置クローニング或いは別の遺伝子発見技術を用いて遺伝的疾患の研究をしている研究者にとって価値がある。疾患や症候群に關与する1つ或いは複数の遺伝子の位置が、例えば血管拡張性失調症の11q22-23などの特定の遺伝子領域に遺伝子結合によって大まかに決定されると、その領域に対するどの配列マッピングも、さらなる調査のための関連する遺伝子或いは調節遺伝子を表す(例えば、Gatti, R. A. 他による(1988) Nature 336:577-580を参照)。また、目的の本発明のヌクレオチド配列を用いて、正常者、保有者、即ち感染者の間の、転位置、反転などによる染色体位置の違いを検出することもある。

【0235】

本発明の別の実施例では、PRTS、その触媒作用断片或いは免疫原断片またはそのオリゴペプチドを、種々の任意の薬剤スクリーニング技術における化合物のライブラリのスクリーニングに用いることができる。このようなスクリーニングに用いる断片は、溶液に遊離、固体支持物に固定、細胞の表面上に保持、或いは細胞内に存在する。PRTSと検査する薬剤との結合による複合体の形成を測定してもよい。

10

20

30

40

50

【0236】

薬剤スクリーニングに用いる別の方法は、目的のタンパク質に対して、好適な結合親和性を有する化合物のスクリーニング処理能力を高めるために用いられる（例えば、Geysen, 他による（1984）PCT出願番号WO84/03564を参照）。この方法では、相当な数の異なる小さな試験用化合物が、プラスチックピン或いは他の基板の上に合成される。試験用化合物は、PRTS、或いはその断片と反応してから洗浄される。次に、結合されたPRTSが、当分野で周知の方法で検出される。精製されたPRTSはまた、前記した薬剤をスクリーニングする技術に用いられるプレート上で直接被覆することもできる。別法では、非中和抗体を用いて、ペプチドを捕らえ、固体支持物に固定することもできる。

10

【0237】

別の実施例では、PRTSと結合可能な中和抗体がPRTSと結合するため試験用化合物と特に競合する、競合的薬剤スクリーニングアッセイを用いることができる。この方法では、抗体が、PRTSと1つ以上の抗原決定因子を共有するどのペプチドの存在も検出する。

【0238】

別の実施例では、発展途上の分子生物学技術にPRTSをコードするヌクレオチド配列を用いて、限定はされないが、現在知られているトリプレット暗号及び特異的な塩基対相互作用などのヌクレオチド配列の特性に依存する新しい技術を提供することができる。

【0239】

当分野の技術者であれば、更なる説明がなくても前述の説明だけで最大限に本発明を利用できるであろう。したがって、以下に記載する特定の好適な実施例は、例示目的であって本発明を限定するものではない。

20

【0240】

前述した及び以下に記載する全ての特許出願、特許、刊行物、特に米国特許出願第60/172,055号、同第60/177,334号、同第60/178,844号、および同第60/179,903号に言及することをもって本明細書の一部とする。

【0241】

（実施例）

1 cDNAライブラリの作製

インサイトcDNAはLIFESEQ GOLD データベース（Incyte Genomics, Palo Alto CA）に含まれているcDNAライブラリに由来し、表4の列5に示されている。SEQ ID NO: 15として示されているインサイトcDNAは、小腸、卵巣、肺、皮膚、乳房、前立腺epithelium、及び、神経前駆物質に含まれる臍帯血や奇形癌細胞などの混合子宮筋組織から作製されたcDNAライブラリに由来する。SEQ ID NO: 17のインサイトcDNAは、気管支上皮第一株化細胞、皮膚微小血管内皮細胞、膵臓、クローン病に関連する回腸組織、パター症候群に関連する肋骨組織、腎臓、胸部後根神経節、及び陰茎海綿体組織から作製されたcDNAライブラリに由来する。SEQ ID NO: 18のインサイトcDNAは、脳腫瘍組織から作製されたcDNAライブラリに由来する。SEQ ID NO: 19のインサイトcDNAは、副腎、大腸、および乳房組織より作製されたcDNAライブラリに由来する。SEQ ID NO: 20のインサイトcDNAは、T-リンパ球、肺、乳房、陰茎海綿体組織より作製されたcDNAライブラリに由来する。まず、組織の一部をホモジナイズしてグアニジニウムイソチオシアネート溶液に溶解する一方、この組織の別の一部をホモジナイズしてフェノールに溶解するか、或いはTRIZOL（Life Technologies）、グアニジニウムイソチオシアネート及びフェノールの単相溶液などの好適な変性剤の混合液に溶解した。この溶解物を塩化セシウムにおいて遠心分離によって、或いはクロロホルムで抽出した。イソプロパノール或いは酢酸ナトリウムのどちらかとエタノール、或いは別の方法でこの溶解物からRNAを沈殿させた。

30

40

【0242】

50

RNAの純度を高めるためにRNAのフェノールによる抽出及び沈殿を必要な回数繰り返した。場合によっては、DNA分解酵素でRNAを処理する。殆どのライブラリでは、オリゴd(T)連結常磁性粒子(Promega)またはOLIGOTEXラテックス粒子(QIAGEN, Valencia CA)、OLIGOTEX mRNA精製キット(QIAGEN)を用いてポリ(A+)RNAを単離した。別法では、POLY(A)PURE mRNA精製キット(Ambion, Austin TX)などの別のRNA単離キットを用いて組織溶解物から直接単離した。

【0243】

ある場合には、Stratagene社にRNAを提供し、Stratagene社が対応するcDNAライブラリを作製した。そうでない場合は、UNIZAPベクターシステム(Stratagene)またはSUPERSCRIPTプラスミドシステム(Life Technologies)を用いて当分野で周知の推奨方法または類似の方法でcDNAを合成してcDNAライブラリを作製した。(例えば、Ausubel, 1997, 前出, ユニット5.1-6.6を参照)。逆転写は、オリゴd(T)またはランダムプライマーを用いて開始した。合成オリゴヌクレオチドアダプターを二本鎖cDNAに結合させてから、好適な1つの制限酵素或いは複数の制限酵素でcDNAを消化した。殆どのライブラリでは、SEPHACRYL S 1000またはSEPHAROSE CL2B、SEPHAROSE CL4Bカラムクロマトグラフィー(Amersham Pharmacia Biotech)、アガロースゲル電気泳動法によってcDNAの大きさ(300~1000bp)を選択した。PBLUESCRIPTプラスミド(Stratagene)またはpSPORT1プラスミド(Life Technologies)、pcDNA2.1プラスミド(Invitrogen Carlsbad CA)、PBK-CMVプラスミド(Stratagene)、pINCYプラスミド(Incyte Pharmaceuticals, Palo Alto CA)若しくはその誘導体などの好適なプラスミドのポリリンカーの適合性制限酵素部位にcDNAを結合させた。この組換えプラスミドを、Stratagene社のXL1-Blue, XL1-Blue MRF、SOLR、またはLife Technologies社のDH5 またはDH10B、ELECTROMAX DH10Bを含むコンピテント大腸菌細胞に導入し組み込んだ。

【0244】

2 cDNAクローンの単離

上記実施例1に記載したように得たプラスミドを、UNIZAPベクターシステム(Stratagene)或いは細胞溶解を利用した*in vivo*切除によって宿主細胞から回収した。MagicまたはWIZARD Minipreps DNA精製システム(Promega)、及びAGTC Miniprep精製キット(Edge Biosystems, Gaithersburg MD)、QIAGEN社のQIAWELL 8 Plasmid、QIAWELL 8 Plus Plasmid、QIAWELL 8 Ultra Plasmid 精製システム、REAL Prep 96プラスミドキットの内の少なくとも1つを用いてプラスミドを精製した。沈殿させた後、0.1mlの蒸留水に再懸濁して、凍結乾燥して或いは凍結乾燥しないで4℃で保管した。

【0245】

別法では、ハイスループットの直接結合PCR法によって宿主細胞溶解物からプラスミドDNAを増幅した。(Rao, V.B. (1994) Anal. Biochem. 216:1-14)。宿主細胞の溶解及び熱サイクリング過程を単一反応混合液で行った。サンプルを処理してから384-ウェルプレートに移して保管し、増幅したプラスミドDNAの濃度をPICOGREEN色素(Molecular Probes, Eugene OR)及びFluoroskan II蛍光スキャナ(Labsystems Oy, Helsinki, Finland)を用いて蛍光定量的に測定した。

【0246】

3 シークエンシング及び分析

実施例 2 に記載したようにプラスミドから回収したインサイト cDNA を、以下に示すようにシーケンシングした。cDNA のシーケンシング反応は標準的な方法で行うか、または HYDRA マイクロディスペンサー (Robbins Scientific) 或いは MICROLAB 2200 (Hamilton) 液体移送装置と共に ABI CATALYST 800 (PE Biosystems) サーマルサイクラー或いは PTC-200 thermal cycler (MJ Research) などのハイスループット装置を用いて行った。cDNA のシーケンシング反応は、Amersham Pharmacia Biotech 社の試薬、または ABI PRISM BIGDYE Terminator cycle sequencing ready reaction キット (PE Biosystems) などの ABI シーケンシングキットに含まれる試薬を用いて行った。cDNA シーケンシングの反応物の電気泳動的による分離及び標識したポリヌクレオチドの検出は、MEGABACE 1000 DNA シーケンシングシステム (Molecular Dynamics)、標準 ABI プロトコル及び塩基対呼び出しソフトウェアを用いる ABI PRISM 373 または 377 シーケンシングシステム (PE Biosystems)、または当分野で周知のその他の配列解析システムを用いて行った。cDNA 配列内の読み枠は、標準的な方法 (Ausubel, 1997, 前出, unit 7.7) を用いて決定した。cDNA 配列の幾つかを選択して、実施例 8 に記載した方法で配列を伸長した。

10

【0247】

インサイト cDNA に由来する本ポリヌクレオチド配列の確認は、BLAST、動的プログラミング、およびジヌクレオチドの分布による解析 (dinucleotide nearest neighbor analysis) に基づいたプログラム及びアルゴリズムを用いて、ベクター、リンカー、およびポリ A 配列を取り除き、更にあいまいな塩基対をマスクすることで行った。次に、インサイト cDNA 配列およびそれらの翻訳を、公共のデータベースである GenBank の霊長類、げっ歯類、哺乳類、脊椎動物、および真核生物のデータベース、および BLOCKS、PRINTS、DOMO、PRODOM、および PFAM などの隠れマルコフモデル (HMM) を基にしたタンパク質ファミリーのデータベースから選択した配列に対して問合せた (HMM は、遺伝子ファミリーのコンセンサス主構造を分析する確率的手法である。例えば、Eddy, S. R. (1996) Curr. Opin. Struct. Biol. 6: 361-365 を参照) 。このような問合せは、BLAST、FASTA、BLIMPS、および HMMR に基づいたプログラムを用いて行った。インサイト cDNA 配列を組み立てて、完全長ポリヌクレオチド配列を作製した。或いは、GenBank cDNAs、GenBank EST、ステッチ配列 (stitched sequence)、ストレッチ配列 (stretched sequences)、または Genscan-推定コード配列 (実施例 4 および 5 を参照) を用いて、インサイト cDNA 群を完全長の配列に伸長した。配列の組み立ては、Phred、Phrap、および Consed に基づいたプログラムを用いて行い、GeneMark、BLAST、および FASTA に基づいたプログラムを用いて cDNA 群をスクリーニングし、オープンリーディングフレームを決定した。これらの完全長ポリヌクレオチド配列を翻訳して対応する完全長ポリペプチド配列を得た。次に、これらのポリペプチド配列を GenBank タンパク質データベース (genpept)、SwissProt、BLOCKS、PRINTS、DOMO、PRODOM、Prosite、および PFAM などの隠れマルコフモデル (HMM) に基づいたタンパク質ファミリーデータベースに対して問合せて分析した。これらの完全長ポリヌクレオチド配列はまた、MACDNASIS PRO ソフトウェア (Hitachi Software Engineering, South San Francisco CA) および LASER GENE ソフトウェア (DNASTAR) を用いて分析した。ポリヌクレオチド配列およびポリペプチド配列のアラインメントを、アラインメントした配列間のパーセント同一性も計算する MEGALIGN マルチシーケンシングアラインメントプログラム (DNASTAR) に組み込まれた CLUSTAL アルゴリズムによって指定されたデフォルトパラメー

20

30

40

50

タを用いて作成した。

【0248】

表7は、インサイトcDNAの組み立ておよび完全長配列の分析に利用したツール、プログラム、およびアルゴリズム、並びにそれらの説明、引用文献、閾値パラメーターを簡単に示す。表7の列1は用いたツール、プログラム、およびアルゴリズム、列2はそれらの簡単な説明、列3は引用することで本明細書の一部とした引用文献、列4の記載されている部分は2つの配列の一致の程度を評価するために用いたスコア、確率値、およびその他のパラメータを示す(スコアが高くなれば高くなるほど即ち確率値が低ければ低いほど、配列間の相同性が高くなる)。

【0249】

完全長ポリヌクレオチド配列およびポリペプチド配列の組み立て及び分析に用いる上記のプログラムは、SEQ ID NO: 15 - 28のポリヌクレオチド配列断片の同定にも利用できる。ハイブリダイゼーション及び増幅技術に有用である約20～約4000ヌクレオチドの断片を表4の列4に示した。

【0250】

4 ゲノムDNA由来のコード配列の同定および編集

推定プロテアーゼは、公共のゲノム配列データベース(例えば、gbpriやgbhtg)においてGenscan遺伝子同定プログラムを実行して初めに同定された。Genscanは、様々な生物に由来するゲノムDNA配列を分析するための汎用遺伝子同定プログラムである(Burge, C. および S. Karlin (1997) J. Mol. Biol. 268: 78-94, Burge, C. および S. Karlin (1998) Curr. Opin. Struct. Biol. 8: 346-354を参照)。このプログラムは推定エキソンを連結して、メチオニンから停止コドンまで伸長した組み立てcDNA配列を構築する。Genscanにより得られる配列は、FASTAデータベースのポリヌクレオチド配列およびポリペプチド配列になる。Genscanによって一回で解析できる配列の最大長さは30kbに設定されている。これらのGenscan推定cDNA配列の内、どの配列がプロテアーゼをコードするかを決定するために、コードされたポリペプチドをPFAMモデルにおいてプロテアーゼについて問合せて分析した。潜在的なプロテアーゼが、プロテアーゼとしてアノテーションが付けられたインサイトcDNA配列に対する相同性を基に同定された。次に、これらの選択されたGenscan推定配列を、BLAST解析を用いてgenepfおよびgbpri公共データベースの配列と比較した。必要に応じて、Genscan推定cDNA配列を、genepfにおいてBLASTで最もヒットした配列と比較して、Genscan推定配列における余分なエキソンや省いてしまったエキソンなどのエラーを修正し、編集した。BLAST解析を用いてGenscan推定cDNA配列を含むインサイトcDNAまたは公共のcDNAを見つけ出すことにより、転写の証拠が得られる。インサイトcDNAがGenscan推定cDNA配列を含む場合、この情報を用いてGenscan推定配列を修正或いは確認できる。完全長ポリヌクレオチド配列は、実施例3に説明した組み立て方法でGenscan推定コード配列とインサイトcDNAおよび/または公共のcDNA配列を組み立てて作製した。また、完全長ポリヌクレオチド配列は、最終的には、編集および編集されていないGenscan推定配列に由来した。

【0251】

5 ゲノム配列データとcDNA配列データとの組み立て

ステッチ配列(Stitched Sequence)

部分的なcDNA配列を、実施例4に記載したGenscan遺伝子同定プログラムによって推定されたエキソンで伸長した。実施例3に記載されたように組み立てられた部分的なcDNAをゲノムDNAにマッピングし、関連するcDNAおよび1或いは複数のゲノム配列に由来する関連する推定Genscanエキソンを含む複数のクラスターに入れた。各クラスターを、グラフ理論および動的計画法に基づいたアルゴリズムを用いて、cDNAおよびゲノム情報を統合して分析し、後に確認される潜在的なスプライスバリエント

10

20

30

40

50

を生成し、編集或いは伸長して完全長の配列を作製した。或るクラスターの2つ以上の配列に或る区間の全長が存在する配列区間を同定し、推移 (transitivity) により同定した区間を同等と考える。例えば、或る区間が cDNA および 2 つのゲノム配列のそれぞれに存在する場合、これら 3 つ全ての区間を同等と考える。この方法によって、関連しないが連続するゲノム配列を cDNA 配列によって繋ぎ 1 つにする。このようにして同定された区間を、親配列 (parent sequence) に沿って現われるようにステッチアルゴリズムで縫い合わせ、可能な最も長い配列および変異配列を作製する。或るタイプ (cDNA と cDNA、またはゲノム配列とゲノム配列) の親配列に沿って連結される区間と区間との繋ぎ合わせは、親配列のタイプが異なる (cDNA とゲノム配列) 連結より好ましい。得られたステッチ配列を翻訳し、BLAST 解析で genepept および gbprl 公共データベースにおける配列と比較した。Genscan によって推定された不適当なエキソンを、genepept において BLAST で最もヒットした配列と比較して修正する。このような配列を更なる cDNA 配列で伸長し、必要に応じてゲノム DNA で検査した。

10

【0252】

ストレッチ配列 (Stretched Sequence)

部分的な DNA 配列を BLAST 解析に基づいたアルゴリズムで完全長に伸長した。まず、実施例 3 に記載したように組み立てた部分的な cDNA を、BLAST プログラムを用いて GenBank の霊長類、げっ歯類、哺乳類、脊椎動物、および真核生物のデータベースなどの公共のデータベースに対して問い合わせた。次に、GenBank の相同性の最も高いタンパク質を、実施例 4 に記載したインサイト cDNA 或いは Genscan エキソン推定配列の何れかと比較した。得られた複数の高スコアのセグメント対 (HSP) を用いてキメラタンパク質を作製し、GenBank の相同タンパク質上に翻訳した配列をマッピングした。元の GenBank の相同タンパク質に対して、キメラタンパク質に挿入や欠失が起こり得る。公共のヒトゲノムデータベースから相同ゲノム配列を探し出すために、GenBank の相同タンパク質およびキメラタンパク質の両方をプローブとして用いた。このようにして、部分的な DNA 配列を相同ゲノム配列の付加によりストレッチすなわち伸長した。完全な遺伝子を含んでいるか得られたストレッチ配列を検査した。

20

【0253】

6 PRTS をコードするポリヌクレオチドの染色体マッピング

SEQ ID NO: 15 - 28 を組み立てるために用いた配列を、BLAST 及び Smith-Waterman アルゴリズムを用いて、インサイト LIFESEQ データベース及び公共のドメインデータベースの配列と比較した。SEQ ID NO: 15 - 28 と一致するこれらのデータベースの配列を、Phrap (表 7) などの構築アルゴリズムを使用して、連続及び重複した配列のクラスターに組み入れた。Stanford Human Genome Center (SHGC)、Whitehead Institute for Genome Research (WIGR) 及び Genethon などの公共の情報源から入手できる放射線ハイブリッド (radiation hybrid) 及び遺伝子マッピングのデータを用いて、クラスター化した配列がすでにマッピングされているかを調べる。クラスターにマッピングされた配列が含まれている場合は、そのクラスターの全ての配列 (特定の SEQ ID NO を含む) をそのマッピング位置に割り当てた。

30

40

【0254】

遺伝子地図の位置は、範囲、区間、またはヒト染色体によって表される。センチモルガンで示したマッピング位置の範囲は、染色体の短腕 (p) の末端から測定した (センチモルガン (cM) は、同一染色体上の遺伝子間の乗換え率に基づいた距離を表す単位である。平均すると、1 cM はヒトの染色体の 1 メガベースに概ね等しいが、組換え率の高い部分と低い部分があるため、大きく変化し得る)。距離 cM は、配列がそれぞれのクラスターに含まれている放射線ハイブリッドマーカーの境界を検出できる Genethon によってマッピングされた遺伝子マーカーに基づいている。NCBI 「GeneMap99」

50

(<http://www.ncbi.nlm.nih.gov/genemap>)などの公衆が入手可能なヒト遺伝子マップおよびその他の情報源を用いて、上記した区間が既に同定されている疾患遺伝子マップ内若しくは近傍に位置するかを決定できる。

【0255】

7 ポリヌクレオチド発現の分析

ノーザン分析は、遺伝子の転写物の存在を検出するために用いられる実験用技術であり、特定の細胞種或いは組織からのRNAが結合されている膜への標識されたヌクレオチド配列のハイブリダイゼーションを伴う(例えば、Sambrook, 前出, 7章; 及び Ausubel, F.M., 他, 前出, 4章及び16章を参照)。

【0256】

BLASTに用いる類似のコンピュータ技術を用いて、GenBank或いはLIFESEQ (Incyte Pharmaceuticals)のようなcDNAデータベース内の同一或いは関連する分子を検索する。この分析は多くの膜系ハイブリダイゼーションより非常に速度が速い。さらにコンピュータ検索の感度を変更して、任意の特定の一致が、厳密な一致或いは相等的一致の何れかとして分類されるかを確定することができる。検索の基準は、

【0257】

【数1】

(BLASTスコア×配列一致率)

$5 \times (\text{長さ(配列1)}, \text{長さ(配列2)})$ の最小値

として定義される積スコアである。積スコアは、0~100の標準化された値であり、以下のように求める。BLASTスコアにヌクレオチド配列の一致率を乗じ、その積を2つの配列の短い方の長さの5倍で除する。高スコアのセグメントの対(HSP)において一致する各塩基に+5のスコアを割り当て、各不適性塩基対に-4を割り当てることにより、BLASTスコアを計算する。2つの配列は、2以上のHSPを共有し得る(ギャップにより離隔される)。2以上のHSPがある場合には、最高BLASTスコアの塩基対を用いて積スコアを計算する。積スコアは、BLASTアラインメントの断片的重複と質とのバランスを表す。例えば積スコア100は、比較した2つの配列の短い方の長さ全体にわたって100%一致する場合にのみ得られる。積スコア70は、100%の一致で一端が70%重畳しているか、或いは88%一致で他端が100%重畳しているかの何れかの場合である。積スコア50は、100%の一致で一端が50%重畳しているか、或いは79%の一致で他端が100%重畳しているかの何れかの場合である。

【0258】

或いは、PRTSをコードするポリヌクレオチド配列は、由来する組織に対して分析する。例えば、ある完全長の配列は、少なくとも部分的にインサイトcDNA配列をオーバーラップさせて組み立てられる(実施例3を参照)。各cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。各ヒト組織は、心血管系、結合組織、消化系、胚構造、内分泌系、外分泌腺、女性生殖器、男性生殖器、生殖細胞、血液および免疫系、肺、筋骨格系、神経系、脾臓、呼吸器系、感覚器官、皮膚、顎口腔系、分類不能/混合、または尿管などの1つの生物/組織のカテゴリーに分類される。各カテゴリーにおけるライブラリの数をカウントし、その合計数を全カテゴリーのライブラリ数で除す。同様に、各ヒト組織は、癌、細胞系、発生、炎症、神経、外傷、心血管、プール(pool)などの1つの疾患/症状のカテゴリーに分類され、各カテゴリーにおけるライブラリの数をカウントし、その合計数を全カテゴリーのライブラリ数で除す。得られるパーセンテージは、PRTSをコードするcDNAの疾患特異的な発現を反映する。cDNA配列およびcDNAライブラリ/組織の情報は、LIFESEQ GOLD データベース(Incyte Genomics, Palo Alto CA)から得ることができる。

【0259】

10

20

30

40

50

8 P R T S をコードするポリヌクレオチドの伸長

完全長のポリヌクレオチド配列は、完全長分子の好適な断片から設計したオリゴヌクレオチドプライマーを用いてその完全長分子の好適な断片を伸長して作製した。一方のプライマーは既知の断片の 5' の伸長を開始するために合成し、他方のプライマーは既知の断片の 3' の伸長を開始するために合成した。開始プライマーは、OLIGO 4.06 ソフトウェア (National Biosciences) 或いは他の適切なプログラムを用いて、約 22 個から約 30 個のヌクレオチドの長さで約 50% 以上の GC 含量を有し、かつ約 68 ~ 72 の温度で標的配列にアニールするように設計した。ヘアピン構造及びプライマー-プライマー二量体が生じないようにヌクレオチドを伸長した。

【0260】

選択されたヒト cDNA ライブラリを用いてこの配列を伸長した。2 段階以上の伸長が必要な場合、若しくは望ましい場合は、追加或いはネスト化プライマーの組を設計する。

【0261】

当分野で既知の方法を利用した PCR 法で高い忠実度で増幅した。PCR は PTC-200 thermal cycler (MJ Research, Inc.) 用いて 96 ウェルブロックプレートで行った。反応混合液は、鋳型 DNA 及び 200 nmol の各プライマー、 Mg^{2+} と $(NH_4)_2SO_4$ と β -メルカプトエタノールを含むバッファー、Taq DNA ポリメラーゼ (Amersham Pharmacia Biotech)、ELONGASE 酵素 (Life Technologies)、Pfu DNA ポリメラーゼ (Stratagene) を含む。プライマーの組、PCI A と PCI B に対して以下のパラメーターで増幅を行った。

ステップ 1	94	で 3 分間
ステップ 2	94	で 15 秒
ステップ 3	60	で 1 分間
ステップ 4	68	で 2 分間
ステップ 5	ステップ 2、3、及び 4 を 20 回繰り返す	
ステップ 6	68	で 5 分間
ステップ 7	4	で保管

別法では、プライマーの組、T7 と SK+ に対して以下のパラメーターで増幅を行った。

ステップ 1	94	で 3 分間
ステップ 2	94	で 15 秒
ステップ 3	57	で 1 分間
ステップ 4	68	で 2 分間
ステップ 5	ステップ 2、3、及び 4 を 20 回繰り返す	
ステップ 6	68	で 5 分間
ステップ 7	4	で保管。

【0262】

各ウェルの DNA 濃度は、1X TE 及び 0.5 μ l の希釈していない PCR 産物に溶解した 100 μ l の PICO GREEN 定量試薬 (0.25% (v/v) PICO GREEN; Molecular Probes, Eugene OR) を不透明な蛍光光度計プレート (Coming Costar, Acton MA) の各ウェルに分配して DNA が試薬と結合できるようにして測定する。このプレートを Fluoroskan II (Labsystems Oy, Helsinki, Finland) でスキャンして、サンプルの蛍光を計測して DNA の濃度を定量化する。反応混合物の 5 ~ 10 μ l のアリコットを 1% のアガロースミニゲル上での電気泳動によって解析し、何れの反応物が配列を伸長することに成功したかを決定する。

【0263】

伸長したヌクレオチドを脱塩及び濃縮してから 384 ウェルプレートに移し、CviJI コレラウィルスエンドヌクレアーゼ (Molecular Biology Research, Madison WI) で消化し、pUC18 ベクター (Amersham Ph

10

20

30

40

50

armacia Biotech) に再連結する前に音波処理またはせん断を行った。シヨットガンシーケンシングのために、消化したヌクレオチドを低濃度 (0.6 ~ 0.8 %) のアガロースゲル上に分離して断片を切断し、寒天を Agar ACE (Promega) で消化した。T4 リガーゼ (New England Biolabs, Beverly MA) を用いて伸長したクローンを pUC 18 ベクター (Amersham Pharmacia Biotech) に再連結し、Pfu DNA ポリメラーゼ (Stratagene) で制限部位の伸び出しを処理してコンピテント大腸菌細胞に形質移入した。形質移入した細胞を選択して抗生物質を含む培地に移し、それぞれのコロニーを切りとって LB / 2 X カルベニシリン培養液の 384 ウェルプレートに 37 で一晩培養した。

【0264】

細胞を溶解して、Taq DNA ポリメラーゼ (Amersham Pharmacia Biotech) 及び Pfu DNA ポリメラーゼ (Stratagene) を用いて以下の手順で DNA を PCR 増幅した。

ステップ 1 94 で 3 分間

ステップ 2 94 で 15 秒

ステップ 3 60 で 1 分間

ステップ 4 72 で 2 分間

ステップ 5 ステップ 2、3、及び 4 を 29 回繰り返す

ステップ 6 72 で 5 分間

ステップ 7 4 で保管。

上記したように PICO GREEN 試薬 (Molecular Probes) で DNA を定量化した。DNA 回収率の悪いサンプルは、上記した条件で再び増幅した。サンプルを 20 % のジメチルサルホサイド (dimethyl sulphoxide) (1 : 2, v/v) で希釈し、DYENAMIC DIRECT キット (Amersham Pharmacia Biotech) または ABI PRISM BIGDYE Terminator cycle sequencing ready reaction キット (Applied Biosystems) を用いてシーケンシングした。

【0265】

同様に上述の手順で、完全長のポリヌクレオチド配列を検査したり、或いは完全長のポリヌクレオチド配列を利用して、この伸長のために設計したオリゴヌクレオチドと好適なゲノムライブラリを用いて 5 調節配列を得た。

【0266】

9 個々のハイブリダイゼーションプローブの標識化及び使用法

SEQ ID NO : 15 - 28 から導き出されたハイブリダイゼーションプローブを用いて、cDNA、mRNA、またはゲノム DNA をスクリーニングする。約 20 塩基対からなるオリゴヌクレオチドの標識について特に記すが、より大きな cDNA フラグメントの場合でも基本的に同じ手順を用いる。オリゴヌクレオチドを、OLIGO 4.06 ソフトウェア (National Bioscience) のような最新式のソフトウェアを用いてデザインし、50 pmol の各オリゴマーと、250 μ Ci の [³²P] アデニン三リン酸 (Amersham, Chicago, IL) 及び T4 ポリヌクレオチドキナーゼ (DuPont NEN, Boston MA) とを組み合わせて用いることにより標識する。標識されたオリゴヌクレオチドを、SEPHADEX G-25 超精細排除デキストランビードカラム (Amersham Pharmacia Biotech) を用いて実質的に精製する。毎分 10⁷ カウントの標識されたプローブを含むアリコットを、次のエンドヌクレアーゼ、Ase I、Bgl II、Eco RI、Pst I、Xba I 或いは Pvu II (DuPont NEN) の 1 つを用いて切断したヒトゲノム DNA の典型的な膜ベースのハイブリダイゼーション解析において用いる。

【0267】

各切断物からの DNA を、0.7 % アガロースゲル上で分画して、ナイロン製メンブラン (Nytran Plus, Schleicher & Schuell, Durham

10

20

30

40

50

NH)に転写する。ハイブリダイゼーションは40で16時間かけて行う。非特異的シグナルを取り除くため、例えば、最大0.1×クエン酸ナトリウム食塩水及び0.5%ドデシル硫酸ナトリウムの条件の下、プロットを順次室温にて洗浄する。ハイブリダイゼーションパターンをオートラジオグラフィー或いは別のイメージ化手段で視覚化して比較する。

【0268】

10 マイクロアレイ

マイクロアレイ上のアレイエレメントの連結または合成は、フォトリソグラフィ、ピエゾプリント(インクジェットプリンター、前出のBaldeschweiler等を参照)、機械的マイクロスポッティング技術及びこれらから派生したものをを用いて達成することが可能である。上記各技術において基板は、均一な非多孔性の固体とするべきである(Schena(1999)、前出)。推奨する基板には、シリコン、シリカ、スライドガラス、ガラスチップ及びシリコンウエハがある。別法では、ドットプロット法またはスポットプロット法に類似のアレイを利用して、熱や紫外線、または化学的或いは機械的な結合手段で基板の表面にエレメントを配置して結合させることができる。通常のアレイは利用可能な方法や機械を用いて作製でき、任意の適正な数のエレメントを含めることができる(Schena, M. 他(1995) Science 270:467-470、Shalon, D. 他(1996) Genome Res. 6:639-645、Marshall, A. and J. Hodgson(1998) Nat. Biotechnol. 16:27-31.を参照)。

10

20

【0269】

完全長cDNA、発現遺伝子配列断片(EST)、或いはそれらの断片やオリゴマーが、マイクロアレイのエレメントとなり得る。ハイブリダイゼーションに好適な断片やオリゴマーを、LASERGENEソフトウェア(DNA STAR)などの当分野で周知のソフトウェアを用いて選択することが可能である。このアレイエレメントを、生体サンプル中のポリヌクレオチドとハイブリダイズさせる。生体サンプル中のポリヌクレオチドは、検出を容易にするために蛍光標識またはその他の分子タグに結合する。ハイブリダイゼーションの後、生体サンプルからハイブリダイズしなかったヌクレオチドを除去し、蛍光スキャナを用いて各アレイエレメントにおけるハイブリダイゼーションを検出する。別法では、レーザー脱離及び質量スペクトロメトリーを用いてもハイブリダイゼーションを検出し得る。マイクロアレイ上のエレメントにハイブリダイズする各ポリヌクレオチドの相補性の程度及び相対的存在量は、算定することができる。一実施例におけるマイクロアレイの調整及び使用について、以下に詳述する。

30

【0270】

組織または細胞サンプルの調製

グアニジウムチオシアネート法を用いて組織サンプルから全RNAを単離し、オリゴ(dT)セルロース法を用いてポリ(A)⁺RNAを精製する。各ポリ(A)⁺RNAサンプルは、MMLV逆転写酵素、0.05 pg/μlのオリゴ(dT)プライマー(21mer)、1×第1鎖緩衝液、0.03単位/μlのRNアーゼインヒビター、500 μM dATP、500 μM dGTP、500 μM dTTP、40 μM dCTP、40 μM dCTP-Cy3(BDS)またはdCTP-Cy5(Amersham Pharmacia Biotech)を用いて逆転写する。この逆転写反応は、GEMBRIGHTキット(Incyte)を用いて、200 ngのポリ(A)⁺RNAを含む25 ml容量で行う。特異的なコントロールポリ(A)⁺RNAは、in vitro転写により非コーディング酵母ゲノムDNAから合成する。370で2時間インキュベートした後、各反応サンプル(一方はCy3標識、他方はCy5標識)は、2.5 mlの0.5 M 水酸化ナトリウムで処理し、850で20分間インキュベートし、反応を停止させてRNAを変性する。サンプルは、2つの連続するCHROMA SPIN 30ゲル濾過スピナラム(CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA)を用いて精製する。結合後、2つの反応サンプルを、1

40

50

mlのグリコーゲン(1mg/ml)、60mlの酢酸ナトリウム及び300mlの100%エタノールを用いてエタノール沈殿させる。サンプルは次に、Speed VAC (Savant Instruments Inc., Holbrook NY)を用いて乾燥して仕上げ、14µl 5xSSC/0.2% SDS中で再懸濁する。

【0271】

マイクロアレイの準備

本発明の配列を用いて、アレイエレメントを作製する。各アレイエレメントは、クローン化cDNA挿入断片を含むベクターを含有する細菌性細胞から増幅する。PCR増幅は、cDNA挿入断片に隣接するベクター配列に相補的なプライマーを用いる。30サイクルのPCRによって、1~2ngの初期量から5µgを超える最終量までアレイエレメントを増幅する。増幅されたアレイエレメントは、SEPHACRYL-400 (Amersham Pharmacia Biotech)を用いて精製する。

10

【0272】

精製したアレイエレメントを、ポリマーコートされたスライドガラス上に固定する。顕微鏡スライドガラス(Corning)は、処理中及び処理後に大量の蒸留水での洗浄と、0.1%のSDS及びアセトン中で超音波による洗浄を行う。スライドガラスは、4%フッ化水素酸(VWR Scientific Products Corporation (VWR), West Chester PA)中でエッチングし、蒸留水中で広範囲にわたって洗浄し、95%エタノール中の0.05%アミノプロピルシラン(Sigma)でコーティングする。コーティングしたスライドガラスは、110℃の天火で硬化させる。

20

【0273】

米国特許第5,807,522号に記載されている方法を用いて、コーティングしたガラス基板にアレイエレメントを付加する。この特許に引用することを以って本明細書の一部とする。平均濃度が100ng/µlのアレイエレメントDNA1µlを高速機械装置により開放型キャピラリープリンティングエレメント(open capillary printing element)に充填する。次にこの装置が、スライド毎に約5nlのアレイエレメントサンプルを分注する。

【0274】

マイクロアレイには、STRATALINKER UVクロスリンカー(Stratagene)を用いてUV架橋する。マイクロアレイは、室温において0.2%SDSで1回洗浄し、蒸留水で3回洗浄する。非特異的な結合部位は、リン酸緩衝生理食塩水(PBS)(Tropix, Inc., Bedford MA)における0.2%カゼイン中で60℃で30分間マイクロアレイをインキュベートし、その後上述したように0.2%SDS及び蒸留水で洗浄することによってブロックする。

30

【0275】

ハイブリダイゼーション

ハイブリダイゼーション反応液は、5xSSC、0.2%SDSハイブリダイゼーション緩衝液にCy3及びCy5標識したcDNA合成産物を各0.2µg含む9µlのサンプル混合体を含めたものである。サンプル混合液を、65℃で5分間加熱し、マイクロアレイ表面上に一定量分注してから1.8cm²のカバーガラスで覆う。このアレイを、顕微鏡スライドより僅かに大きいキャピティを有する防水チャンパーに移す。チャンパーの角に140µlの5xSSCを加えて、チャンパー内を湿度100%に保持する。このアレイを含むチャンパーを、60℃で約6.5時間インキュベートする。アレイは、第1洗浄緩衝液中(1xSSC, 0.1%SDS)において45℃で10分間、第2洗浄緩衝液中(0.1xSSC)において45℃で10分間それぞれ3回洗浄し、その後乾燥させる。

40

【0276】

検出

レポーター標識されたハイブリダイゼーション複合体は、Cy3を励起するための488

50

nm、及びCy3を励起するための632nmのスペクトル線を生成し得るInnova 70混合ガス10Wレーザー(Coherent, Inc., Santa Clara CA)を備えた顕微鏡で検出する。20倍の顕微鏡対物レンズ(Nikon, Inc., Melville NY)を用いて、アレイ上に励起レーザー光を集中させる。このアレイを含むスライドを顕微鏡のコンピュータ制御X-Yステージに置き、対物レンズを通してラスタスキャンする。本実施例で用いた1.8cm×1.8cmのアレイは、20μmの解像度でスキャンする。

【0277】

2つの異なるスキャンにおいて、混合ガスマルチラインレーザーは2つの蛍光体を連続的に励起する。放射された光は、波長に基づいて2つの蛍光体に対応する2つの光電子増倍管検出器(PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ)に分割される。アレイと光電子増倍管との間に配設された好適なフィルターを用いて信号をフィルタリングする。用いる蛍光体の最大発光は、Cy3では565nm、Cy5では650nmである。装置は両方の蛍光体からのスペクトルを同時に記録できるが、レーザー源に好適なフィルターを用いて、蛍光体1つにつき1回スキャンし、各アレイを通常2回スキャンする。

【0278】

スキャンの感度は通常、既知濃度のサンプル混合体に添加されるcDNAコントロール種により生成されるシグナル強度を用いて較正する。アレイ上の特定の位置には相補的DNA配列を含め、その位置におけるシグナルの強度がハイブリダイズする種の重量比1:100,000に相関するようにする。異なる試料(例えば検査細胞及びコントロール細胞を代表する)からの2つのサンプルを、各々異なる蛍光体で標識し、他と異なって発現する遺伝子を同定するために単一のアレイにハイブリダイズさせる場合には、較正は2つの蛍光体を有する較正するcDNAのサンプルを標識して、ハイブリダイゼーション混合液に各々等量を加えて行う。

【0279】

光電子増倍管の出力は、IBMコンパチブルPCコンピュータにインストールされた12ビットRTI-835Hアナログ-デジタル(AID)変換ボード(Analog Devices, Inc., Norwood MA)を用いてデジタル化される。デジタル化されたデータは、リニア20色変換を用いてシグナル強度が青色(低シグナル)から赤色(高シグナル)までの擬似カラー範囲にマッピングされるイメージとして表示される。データはまた、定量的に分析される。2つの異なる蛍光体を同時に励起して測定する場合には、各蛍光体の発光スペクトルを用いて、先ずデータは蛍光体間の光学的漏話(重複発光スペクトルに起因する)に対して補正される。

【0280】

グリッドを蛍光シグナルイメージ上に重畳して、各スポットからのシグナルがグリッドの各エレメントに中央に位置するようにする。各エレメント内の蛍光シグナルを統合し、シグナルの平均強度に対応する数値を得る。シグナル分析に用いるソフトウェアは、GEMTOOLS遺伝子発現分析プログラム(Incyte)である。

【0281】

1.1 相補的ポリヌクレオチド

PRTSをコードする配列或いはその任意の一部に対して相補的な配列は、天然のPRTSの発現を低下させるため即ち阻害するために用いられる。約15~約30個の塩基対を含むオリゴヌクレオチドの使用について記すが、より小さな或いはより大きな配列の断片の場合でも本質的に同じ方法を用いることができる。Oligo4.06ソフトウェア(National Biosciences)及びPRTSのコーディング配列を用いて、適切なオリゴヌクレオチドを設計する。転写を阻害するためには、最も独特な5'配列から相補的なオリゴヌクレオチドを設計し、これを用いてプロモーターがコーディング配列に結合するのを阻害する。翻訳を阻害するためには、相補的なオリゴヌクレオチドを設計して、リボソームがPRTSをコードする転写物に結合するのを阻害する。

10

20

30

40

50

【0282】

12 PRTSの発現

PRTSの発現及び精製は、細菌若しくはウイルスを基にした発現系を用いて行うことができる。細菌でPRTSが発現するために、抗生物質耐性及びcDNAの転写レベルを高める誘導性のプロモーターを含む好適なベクターにcDNAをサブクローニングする。このようなプロモーターには、lacオペレーター調節エレメントに関連するT5またはT7バクテリオファージプロモーター及びtrp-lac(tac)ハイブリッドプロモーターが含まれるが、これらに限定されるものではない。組換えベクターを、BL21(DE3)などの好適な細菌宿主に形質転換する。抗生物質耐性をもつ細菌が、イソプロピル-Dチオガラクトピラノシド(IPTG)で誘発されるとPRTSを発現する。真核細胞でのPRTSの発現は、昆虫細胞株または哺乳動物細胞株に一般にバキュロウイルスとして知られているAutographica californica核多面性ウイルス(ACMNPV)を感染させて行う。バキュロウイルスの非必須ポリヘドリン遺伝子を、相同組換え或いは転移プラスミドの媒介を伴う細菌の媒介による遺伝子転移のどちらかによって、PRTSをコードするcDNAと置換する。ウイルスの感染力は維持され、強いポリヘドリンプロモーターによって高いレベルのcDNAの転写が行われる。組換えバキュロウイルスは、多くの場合はSpodoptera frugiperda(Sf9)昆虫細胞に感染に用いられるが、ヒト肝細胞の感染にも用いられることもある。後者の感染の場合は、バキュロウイルスの更なる遺伝的変更が必要になる。(例えば、Engelhard, E. K. 他(1994) Proc. Natl. Acad. Sci. U S A 91:3224-3227; Sandig, V. 他(1996) Hum. Gene Ther. 7:1937-1945. を参照)。

【0283】

殆どの発現系では、PRTSが、例えばグルタチオンSトランスフェラーゼ(GST)、またはFLAGや6-Hisなどのペプチドエピトープ標識で合成された融合タンパク質となるため、未精製の細胞溶解物からの組換え融合タンパク質の親和性ベースの精製が素早く1回で行うことができる。Schistosoma japonicumからの26キロダルトンの酵素GSTによって、タンパク質の活性及び抗原性を維持した状態で固定されたグルタチオンで融合タンパク質の精製が可能となる(Amersham Pharmacia Biotech)。精製の後、GST部分を特定の操作部位でPRTSからタンパク質的に切断できる。アミノ酸8個のペプチドであるFLAGで、市販のモノクローナル及びポリクローナル抗FLAG抗体(Eastman Kodak)を用いた免疫親和性の精製が可能となる。6個の連続するヒスチジン残基のストレッチである6-Hisによって、金属キレート樹脂(QIAGEN)で精製が可能となる。タンパク質の発現及び精製の方法は、Ausubel(1995, 前出, ch 10, 16)に記載されている。これらの方法で精製したPRTSを直接用いて以下の実施例16、17、18、及び19のアッセイを行うことができる。

【0284】

13 機能のアッセイ

PRTSの機能は、哺乳動物細胞培養系において生理学的に高められたレベルでのPRTSをコードする配列の発現によって評価する。cDNAを、cDNAを高いレベルで発現する強いプロモーターを含む哺乳動物発現ベクターにサブクローニングする。このようなベクターには、pCMV SPORTTM(Life Technologies.)及びpCR3.1(Invitrogen, Carlsbad, CA)が含まれ、どちらもサイトメガロウイルスプロモーターを含んでいる。5~10µgの組換えベクターを、例えば内皮由来か造血由来のヒト細胞株にリポソーム製剤或いは電気穿孔法によって一時的に形質移入する。更に、標識タンパク質をコードする配列を含む1~2µgのプラスミドを同時に形質移入する。標識タンパク質の発現により、形質移入された細胞と形質移入されていない細胞とを区別できる。また、標識タンパク質の発現によって、cDNAの組換えベクターからの発現を正確に予想できる。このような標識タンパク質には、緑色蛍

光タンパク質 (GFP; Clontech)、及びCD64またはCD64-GFP融合タンパク質が含まれる。レーザー光学に基づいた技術を利用した自動流動細胞計測法 (FCM) を用いて、GFPまたはCD64-GFPを発現する形質移入された細胞を同定し、その細胞のアポトーシス状態や他の細胞特性を評価する。また、FCMで、先行した或いは同時の細胞死の現象を診断する蛍光分子の取り込みを検出して計量する。これらの現象には、プロピジウムヨウ化物でのDNAの染色によって計測される核DNA内容物の変化と、プロモデオキシウリジンの取り込み量の低下によって計測されるDNA合成の下方調節と、特異的な抗体との反応性によって計測される細胞表面及び細胞内のタンパク質の発現の変化と、蛍光複合アネキシンVタンパク質の細胞表面への結合によって計測される原形質膜組成の変化とが含まれる。流動細胞計測法は、Ormerod, M. G. による (1994) Flow Cytometry Oxford, New York, NY. に記載されている。 10

【0285】

遺伝子発現におけるPRTSの影響は、PRTSをコードする配列とCD64またはCD64-GFPのどちらかが形質移入された高度に精製された細胞集団を用いて評価することができる。CD64またはCD64-GFPは形質転換された細胞表面で発現し、ヒト免疫グロブリンG (IgG) の保存された領域と結合する。形質転換された細胞と形質転換されない細胞とは、ヒトIgGがCD64に対する抗体のどちらかで被覆された磁気ビードを用いて分離することができる (DYNAL, Lake Success, NY)。mRNAは、当分野で周知の方法で細胞から精製することができる。PRTS及び目的の他の遺伝子をコードするmRNAの発現は、ノーザン分析やマイクロアレイ技術で分析することができる。 20

【0286】

1.4 PRTSに特異的な抗体の作製

ポリアクリルアミドゲル電気泳動法 (PAGE; 例えば、Harrington, M. G. (1990) Methods Enzymol. 1816-3088-495を参照) または他の精製技術で実質的に精製されたPRTSを用いて、標準的なプロトコルでウサギを免疫化して抗体を作り出す。

【0287】

別法では、PRTSアミノ酸配列をLASERGENEソフトウェア (DNASTAR) を用いて解析して免疫原性の高い領域を決定し、対応するオリゴペプチドを合成してこれを用いて当業者に周知の方法で抗体を生産する。C末端付近の、或いは隣接する親水性領域内のエピトープなどの適切なエピトープの選択については、当分野で周知である (例えば、前出のAusubel, 1995, 11章を参照)。 30

【0288】

通常、約15残基の長さのオリゴペプチドを、Applied BiosystemsのABI 431Aペプチドシンセサイザー (PE Biosystems) を用いてfmoc法のケミストリにより合成し、N-マレイミドベンゾイル-N-ヒドロキシスクシンイミドエステル (MBS) を用いた反応によりKLH (Sigma-Aldrich, St. Louis MO) に結合させて、免疫原性を高める (例えば、前出のAusubel, 1995を参照)。フロイントの完全アジュバントにおいてオリゴペプチド-KLH複合体を用いてウサギを免疫化する。得られた抗血清の抗ペプチド活性及び抗PRTS活性を検査するには、ペプチドまたはPRTSを基板に結合し、1%BSAを用いてブロック処理し、ウサギ抗血清と反応させて洗浄し、さらに放射性ヨウ素標識されたヤギ抗ウサギIgGと反応させる。 40

【0289】

1.5 特異的な抗体を用いる天然PRTSの精製

天然PRTS或いは組換えPRTSを、PRTSに特異的な抗体を用いるイムノアフィニティークロマトグラフィにより実質的に精製する。イムノアフィニティークラムは、CNBr-活性化SEPHAROSE (Amersham Pharmacia Biotec 50

h) のような活性化クロマトグラフィー用レジンと抗 P R T S 抗体とを共有結合させることにより形成する。結合の後、そのレジンを製造者の使用説明書に従ってブロッキング処理し洗浄する。

【0290】

P R T S を含む培養液をイムノアフィニティーカラムに通し、P R T S を優先的に吸着できる条件で（例えば、界面活性剤の存在下において高イオン強度のバッファーで）そのカラムを洗浄する。そのカラムを、抗体と P R T S との結合を切るような条件で（例えば、pH 2 ~ 3 のバッファー、或いは高濃度の尿素またはチオシアン酸塩イオンのようなカオトロピックイオンで）溶出させ、P R T S を回収する。

【0291】

10

1.6 P R T S と相互作用する分子の同定

P R T S または生物学的に活性なその断片を、¹²⁵I ボルトンハンター試薬（例えば、Bolton A. E. 及び W. M. Hunter (1973) Biochem. J. 133: 529 を参照）で標識する。マルチウェルプレートに予め配列しておいた候補の分子を、標識した P R T S と共にインキュベートし、洗浄して、標識した P R T S 複合体を有する全てのウェルをアッセイする。様々な P R T S 濃度で得られたデータを用いて、候補分子と結合した P R T S の数量及び親和性、会合についての値を計算する。

【0292】

別法では、P R T S と相互作用する分子を、Fields, S. 及び O. Song (1989, Nature 340: 245 - 246) に記載の酵母 2 - ハイブリッドシステム (yeast two-hybrid system) や MATCHMAKER システム (Clontech) などの 2 - ハイブリッドシステムに基づいた市販のキットを用いて分析する。

20

【0293】

P R T S はまた、ハイスルーブット型の酵母 2 ハイブリッドシステムを使用する PATH CALLING プロセス (CuraGen Corp., New Haven CT) に用いて、遺伝子の 2 つの大きなライブラリによってコードされるタンパク質間の全ての相互作用を決定することができる (Nandabalan, K. 他 (2000) 米国特許第 6,057,101 号)。

【0294】

30

1.7 P R T S の活性の実証

プロテアーゼ活性は、様々な色素分子と共役された適切な合成ペプチド基質加水分解によって測定され、色素分子内部では、加水分解の程度が放出発色団の分光光度法 (蛍光) 吸収 (Beynon, R. J. 及び J. S. Bond (1994)、Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 25 - 55) で定量される。ペプチド基質は、エンドペプチダーゼ (セリン、システイン、アスパラギン酸プロテアーゼ、若しくはメタロプロテアーゼ)、アミノペプチダーゼ (ロイシンアミノペプチダーゼ)、若しくはカルボキシペプチダーゼ (カルボキシペプチダーゼ A 及び B、プロコラーゲン C - プロテイナーゼ) としての、プロテアーゼ活性カテゴリーに従ってデザインされる。通常、用いられたクロモゲン は、2 - ナフチルアミン、4 - ニトロアニリン、フリルアクリル酸である。アッセイは外界温度で実行され、好適なバッファ内の酵素及び適切な基質のアリコットを含んでいる。反応は蛍光キュベット内で実行され、ペプチド基質の加水分解の間に遊離された色素原の吸光度における増加 / 減少が測定される。吸光度に於ける変化は、アッセイ中の酵素活性に比例する。

40

【0295】

ユビキチン加水分解酵素活性のための代替アッセイは、ユビキチン前駆体の加水分解を測定する。アッセイは外界温度で実行され、適当なバッファ中に適切な基質及び P R T S のアリコットを含んでいる。化学的に合成されたヒトユビキチン - バリンは基質として用いられる。基質よりの C - 末端バリン残基の切断は、毛細管電気泳動 (Franklin,

50

K.ら(1997) Anal. Biochem. 247 : 305 - 309) でモニターされる。

【0296】

別法では、プロテアーゼ活性のためのアッセイが蛍光共鳴エネルギー変換(FRET)を利用する。FRETは、一つのドナー及び一つのアクセプターフルオロフォアが適切なスペクトルの重なりを有して近接している場合に発生する。PRTSに特異的な切断部位を有するフレキシブルなペプチドリンカーは、緑蛍光タンパク質の、赤シフト(red-shifted variant)変異体(RSGFP4)及び青変異体間で融着する。

【0297】

この融着タンパク質は、エネルギー移動をBFP5からRSGFP4へと発生させるスペクトル特性を有する。融着タンパク質がPRTSと共にインキュベートされる場合、基質が切断され二つの蛍光タンパク質が分離される。このことは、PRTSの付加前後における発行スペクトルを比較すること(Mitra, R. 他、(1996) Gene 173 : 13 - 17)で定量される、エネルギー移動中の顕著な減少を伴う。アッセイは生細胞内でも行われ得る。このような場合は、蛍光基質タンパク質は細胞中で恒常的に発現され、またPRTSが誘導ベクター上に提供され、よってFRETがPRTSの存在及び不在という形でモニターされうる(Sagot, Iら : (1999) FEBS Lett. 447 : 53 - 57)。

【0298】

18 PRTS基質の同定

ファージディスプレイライブラリが、PRTSの光学基質配列を識別するのに用いられてもよい。リンカー及び既知の抗体エピトープに続くランダムな六量体は繊維状ファージライブラリ中の遺伝子IIIのN-末端延長としてクローン化される。遺伝子IIIコードが外殻蛋白質をコードし、またエピトープが各々のファージ分子表面上に表示されてもよい。ライブラリはタンパク分解性条件下で、PRTSとインキュベートされ、エピトープはPRTSのための六量体コードが部位を切断する場合取り除かれてもよい。エピトープを認識する抗体が、固定化タンパク質Aと共に添加される。エピトープを有している切断されていないファージは遠心分離によって取り除かれる。表面のファージは次に増幅され、スクリーニングのさらなる複数のラウンドが生じる。個々のファージクローンは次に単離され、配列決定される。これらペプチド基質に対する速度論は、実施例17に於けるアッセイを用いて研究されてもよく、また最適切断配列が得られうる(Ke, S. H.ら(1997) J. Biol. Chem 272 : 16603 - 16609)。

【0299】

in vivo PRTS基質のスクリーンのため、この方法は、ファージ分子(T7 SELECTTM 10-3 ファージディスプレイベクター、Novagen, Madison, WI)、若しくは酵母菌(pYD1 酵母ディスプレイベクターキット、Invitrogen, Carlsbad, CA)の表面上にディスプレイされたcDNA発現ライブラリをスクリーンするべく拡張されてもよい。このような場合、全体のcDNAは、遺伝子IIIと適切なエピトープとの間で融着される。

【0300】

19 PRTS阻害剤の同定

試験される化合物は、実施例17のアッセイで記述されているように、適切なバッファ及び基質に沿って濃度を変化させ、マルチウェルプレート上に配列される。PRTS活性は各々の穴で測定され、PRTS活性を抑制するために各化合物の各化合物の能力が決定されてもよい。用量 応答の速度論も同様である。アッセイはPRTSの活性を向上させる分子を同定することに用いられてもよい。

【0301】

別法では、ファージディスプレイライブラリがペプチドPRTS阻害剤をスクリーンするために用いられてもよい。候補はプロテアーゼに強く結合されたペプチド中に見られた。この場合、マルチウェルプレートの各ウェルは、PRTSでコートされ、ランダムなペプ

10

20

30

40

50

チドファージディスプレイライブラリ若しくは周期的なペプチドライブラリでインキュベートされる (Koivunen, E.ら、(1999) Nat. Biotechnol. 17: 768-774)。結合していないファージは洗い流され、選択されたファージが複数のさらなるラウンドで増幅され、再スクリーニングされる。候補は、実施例17に記載のアッセイを用いてPRTSの阻害活性のために試験される。

【0302】

当業者は、本発明の範囲及び精神から逸脱することなく本発明に記載した方法及びシステムの種々の改変を行うことができるであろう。特定の好適な実施例に基づいて本発明を説明したが、本発明の範囲が、そのような特定の実施例に不当に制限されるべきではないことを理解されたい。実際に、分子生物学或いは関連する分野の専門家には明らかな、本明細書に記載の本発明の実施例の様々な改変は、特許請求の範囲に含まれる。

10

【0303】

(表の簡単な説明)

表1は、本発明のポリヌクレオチド配列及びポリペプチド配列に対する系統的な名称を示す。

【0304】

表2は、本発明の各ポリペプチドに最も近いGenBankの相同体のGenBankの識別番号およびアノテーションを示す。各ポリペプチドとそのGenBankの相同体との間の一致を表す確率値スコアも示す。

【0305】

表3は、推定上のモチーフおよびドメインを含む各ポリペプチド配列の構造的な特徴、並びに各ポリペプチドの分析に用いた方法、アルゴリズム、および検索可能なデータベースを示す。

20

【0306】

表4は、各ポリヌクレオチド配列の組み立てに用いたcDNA断片およびゲノムDNA断片のリスト、並びにポリヌクレオチド配列の選択された断片のリストを示す。

【0307】

表5は、本発明の各ポリヌクレオチドの代表的なcDNAライブラリを示す。

【0308】

表6は、表5に示すcDNAライブラリの作製に用いた組織およびベクターを示す付録である。

30

【0309】

表7は、本発明のポリヌクレオチド及びポリペプチドの分析に用いたツール、プログラム、及びアルゴリズム、並びにその説明、引用文献、閾値パラメーターを示す。

【表1】

表1

インサイトプロジェクトID	ポリペプチド SEQ. ID NO.:	インサイトポリペプチドID	ポリヌクレオチド SEQ. ID NO.:	インサイト-ポリヌクレオチドID
1714846	1	1714846CD1	15	1714846CB1
1856589	2	1856589CD1	16	1856589CB1
2617672	3	2617672CD1	17	2617672CB1
2769104	4	2769104CD1	18	2769104CB1
4802789	5	4802789CD1	19	4802789CB1
60116897	6	60116897CD1	20	60116897CB1
1866356	7	1866356CD1	21	1866356CB1
1872095	8	1872095CD1	22	1872095CB1
2278688	9	2278688CD1	23	2278688CB1
4043361	10	4043361CD1	24	4043361CB1
3937958	11	3937958CD1	25	3937958CB1
7257324	12	7257324CD1	26	7257324CB1
7472038	13	7472038CD1	27	7472038CB1
7472041	14	7472041CD1	28	7472041CB1

10

20

30

40

【表2】

表2

ポリペプチド SEQ ID NO:	インサイト ポリペプチドID	GenBank ID NO:	確率スコア	GenBank 相同体
1	1714846	g6941890	0.0	ユビキチン特異プロセッシングプロテアーゼ [ハツカネズミ] (Valero, R. 他(1999) Genomics 62:395-408)
2	1856589	g1143194	1.2e-45	プロスタシン[ヒト] (Yu, J.X. 他(1994) J Biol.Chem. 269:18843-18848)
3	2617672	g4929827	8.0e-118	尿管管間質性腎炎抗原TIN-Ag [ハツカネズミ]
4	2769104	g179644	3.3e-28	ヒトの種体Or [ヒト]
5	4802789	g4454565	4.1e-30	ユビキチンプロセッシングプロテアーゼ [ヒト] (Cai, S. 他 Proc. Natl. Acad. Sci. USA (1999) 96:2828-2833)
6	60116897	g9886747	0.0	VEGF誘導アミノペプチダーゼ [ハツカネズミ]
7	1866356CD1	g2088823	1.5e-68	ペプチダーゼファミリーA2類似 [線虫]
8	1872095CD1	g2347100	1.7e-22	ユビキチン特異プロテアーゼ [シロイヌナズナ]
9	2278688CD1	g1184161	0.0	アミノペプチダーゼ [ハツカネズミ]
10	4043361CD1	g9843781	2.6e-104	推定ピログルタルミルペプチダーゼ1 [ハツカネズミ]
11	3937958CD1	g180950	5.7e-16	カルボキシエステラーゼ [ヒト]
12	7257324CD1	g2116650	1.1e-78	α -1-アンチトリプシン [オナガザル] (Colau, B. 他(1984) DNA 3:327-330; Yoshida, K. 他(1989) J. Biochem. Mol. Biol. Biophys. 3:59-63)
13	7472038CD1	g293230	4.0e-106	アスパラギン酸プロテアーゼ [ネッタイシマカ]
14	7472041CD1	g3088553	1.1e-14	シスタチン関連糖鎖の精子形成タンパク質 [ヒト] (Cornwall, G.A., Hsia, N., と Sutton, H.G. Biochem. J. (1989) 340(Pt 1):85-93)

10

20

30

40

【表3】

表 3-1

SEQ ID NO:	インサイト ホリペプチド ID	アミノ酸 残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ(signature)配列, ドメイン及びモチーフ	分析方法及び データベース
1	1714846CD1	1055	S85 T95 S109 T123 S136 T147 T286 S357 S375 S467 S489 T541 S546 T557 S631 S632 S745 T796 T824 S835 S892 S945 S1021 S1032 T1050 T55 S113 T235 T267 S354 S460 S513 S582 S719 Y575 Y872 Y873	N31 N256 N561 N646 N833	ユビキチンC-末端ヒドロラーゼ: L232-W703, K823-F899 ユビキチンC-末端ヒドロラーゼ シグネチャ1: V169-Y200 ユビキチンC-末端ヒドロラーゼ シグネチャ2: G170-L187, S258-I267, P590-D614, E617-R638, I587-N656, Y591-Y608, N173-N408, D562-G601	HMMER-PFAM BLIMPS- BLOCKS BLAST- PRODOM BLAST-DOMO MOTIFS
2	1856589CD1	358	S47 T188 T5 S105 T143 Y247	N150	キモトリプシンファミリー: G115-G130, F173-V187, E277-A289 トリプシンファミリー: W100-1327, G114-G130, N278-Y301, P314-1327, W100-M331 トリプシンファミリーヒステジン活性部位: V125-G130, L106-N150 トリプシンファミリーセリン活性部位: V265-K310 クリングルドメイン: G114-S131, I196-S217, G286-1327 リンゴドメイン: G116-P148, V187-Q221, I270-W304, E305-N633	HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS ProfileScan BLAST- PRODOM BLAST-DOMO MOTIFS
3	2617672CD1	467	T80 T117 T126 T169 T205 S296 T411 T180 S210 S239 S401 T417	N78 N161	シグナルペプチド: M1-A19, M1-G21 パバインファミリープロテアーゼ: D222-W456, Q223-F232, Q267-L275, T399-G408, Y420-H436, Q223-A238, H400-E410, Y420-S426, D222-R441, F76-G457, D145-V455 システインプロテアーゼヒステジン活性部位: G398-G408 尿細管間質性腎炎抗原: G45-I193	HMMER SPScan HMMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS PRODOM BLAST-DOMO MOTIFS

10

20

30

40

【表 4】

冊 3-2

SEQ ID NO:	インサイト ポリペプチド ID	アミノ酸 残基数	潜在的リン酸化部位	潜在的 グリコシル化部位	シグネチャ(signature)配列, ドメイン及びモチーフ	分析方法及び データベース
4	2769104CD1	187	S67 T162 S131 S134 T138	N147	CUBドメイン(糖体系タンパク質で見つかった 細胞外ドメイン): G40-Y160 補体C1/C1s反産: C36-Y163, Q51-Y160, M24-Y160 シグナルペプチド: M1-A35	HMER-PFAM BLAST-DOMO MOTIFS
5	4802789CD1	289	T18 S28 S109 T213 S236 S261 S17 S102 S108 S188 S225 T265 S271	N119 N186	膜貫通ドメイン: W25-L52 ユピキチンC-末端ヒドロラーゼ シグネチャ1: G191-L208	SPScan HMER HMER BLIMPS- BLOCKS HMER-PFAM MOTIFS
6	60116897CD1	960	S225 S483 T57 T87 S124 T197 S321 T343 S357 T407 S502 S607 S701 S738 S744 S817 S906 S926 T933 S10 S94 T183 S221 T256 S303 S359 S432 S486 S558 S740 S781 T830 T951 Y312 Y622 Y679 Y885	N85 N103 N119 N219 N294 N405 N431 N650 N714 N879	シグナルペプチド: M1-A44 Zn金属ペプチダーゼファミリー-M1: L69-G458 Zn膜アラニルジペプチターゼ: R205-F220, F253-Y268, F331-L341, V367-I382, R386-Y398 中性Znプロテアーゼ: W64-S500, G529-L837, T521-S889, W64-T902, P54-D555, K553-L837, V849-L956 中性Znプロテアーゼ-Zn結合領域: V367-W376, V367-F377	SPScan HMER-PFAM BLIMPS- BLOCKS BLIMPS- PRINTS BLAST- PRODOM BLAST-DOMO MOTIFS
7	1866356CD1	525	S82 S90 T159 T174 S288 S290 T311 T356 S397 T479 S522 S107 S122 S165 S228		シグナルペプチド: M1-C35 シグナルペプチド: M1-S26 ペプチダーゼファミリー-A2 PD138963類似: F157-G422	SPScan SPScan BLAST- PRODOM

10

20

30

40

【表 5】

SEQ ID NO:	インサイト ポリペプチド ID	アミノ酸 残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
8	1872095CD1	795	S274 S279 S522 Y523 T693 T251 S274 S314 S332 T337 S377 S378 S383 S392 S470 T472 S555 S557 S580 T582 T619 S620 T621	N171 N381 N443 N448 N536 N617 N670 N436 N711 N712 N720 N788	ユビキチンカルボキシル末端 ヒドロラーゼ1モチーフ: G199-1213 ユビキチンカルボキシル末端 ヒドロラーゼ2モチーフ: I599-4610 ユビキチンカルボキシル末端 ヒドロラーゼファミリー-UCH-1: T198-L229 ユビキチンカルボキシル末端 ヒドロラーゼファミリー-UCH-2 K589-K701 プロテアーゼ, ユビキチンヒドロラーゼ, ユビキチン特異酵素, ユビキチンリターネーティング (diubiquitinating)カルボキシル末端 チオールエステラーゼ, プロセッシング, 接合: PD017412, S470-L541	MOTIFS MOTIFS HMMER-PFAM HMMER-PFAM BLAST- PRODOM BLAST-DOMO
					ユビキチンカルボキシル末端 ヒドロラーゼファミリー: DM00659 (P40818) 782-1103: L203-D386	

10

20

30

40

【表 6】

表3-4

SEQ ID NO:	インサイトポリペプチドID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ(s)配列、ドメイン及びモチーフ	分析方法及びデータベース
9	2278688CD1	919	T177 T325 Y326 S379 S427 S547 T548 S549 S632 T633 S667 T669 T721 T758 T759 S32 S33 T143 T325 Y326 S341 S342 S486 S522	N62 N484 N648	膜アラニルペプチターゼ: PR00756: R186-F200, F295-V290, F313-L323, V349-T364, W368-W380 亜鉛, アミノペプチターゼ, 金属ペプチターゼ, 中性: DM00700(P164606)80-887: R53-Y842 亜鉛プロテアーゼ: V349-Q357 ロイシンジッパー: L3-P23 シグナルペプチド: M1-S39 ペプチターゼファミリー: L54-G441 アミノペプチターゼ, ヒドロラーゼ, 金属プロテアーゼ, 亜鉛, N-糖タンパク質, 膜内外, シグナル, アンカー, 膜: PD001134: R53-S486	BLIMPS- PRINTS BLAST-DOMO MOTIFS MOTIFS HMMER HMMER-PFAM BLAST- PRODOM
10	4043361CD1	209	S118	N22	亜鉛, アミノペプチターゼ, 金属ペプチターゼ, 中性: DM00700(P37898)1-794: E52-6845 ピログルタミルペプチターゼ: K6-L182 ピロリドンカルボキシリペプチターゼ: PR00140: T11-L31, S66-E85 (PK0.0041) ペプチターゼ, カルボキシレート, ピロリドン, ピログルタミン酸: DM03107(P42673)1-212: K6-G145 カルボキシエステラーゼドメイン: E4-W62	BLAST-DOMO HMMER-PFAM BLIMPS- PRINTS BLAST-DOMO
11	3937958CD1	77	Y35 T47 S68		カルボキシエステラーゼドメイン: E4-W62 エステラーゼ, ヒドロラーゼ, プレカーサー, シグナル, 糖タンパク質, セリン, カルボキシエステラーゼファミリー: PD000169: K3-W62 コリンエステラーゼ: DM00390(00479)355-538: K3-W62 B型カルボキシエステラーゼ: W15-W25	HMMER-PFAM BLAST- PRODOM BLAST-DOMO BLIMPS- BLOCKS

10

20

30

40

【表7】

表 8 - 5

SEQ ID NO.	インサイト ポリペプチド ID	アミノ酸 残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
12	7257324CID1	414	S93 T94 T223 T258 T16 S26 T124 S182 S235 S300 S346 S396 Y118	N221 N233 N267	セルピタンパク質シグネチャ BL00284: N71-T94, A173-I193, T200-M241, Y306-F332, D387-P411 セルピタンシグネチャ: G364-K414 セルピタン, セリンプロテアーゼ抑制因子, シグナル, プレカイン, タンパク質, 原形質, タンパク質分解酵素, P0000192: A44-P411 セルピタン: DM00112 (P01009) 47-413: D54-N410 シグナルペプチド: M1-619 セルピタン(セリンプロテアーゼ抑制因子): A45-P411	BLIMPS- BLOCKS ProfileScan BLAST- PRODOM BLAST- DOMO HMMER SPSCAN HMMER-PFAM
13	7472038CID1	397	S127 T166 S317 T381 S337 Y340 S16 S31 T90 T154 S252	N156 N166 N169 N178 N190 N195 N245 N298 N245 N298	ペプシン(A1)アスパラギン酸プロテアーゼ シグネチャ PRO0792A: I84-V104, G230-T243, V278-L289, W369-D384 アスパルチルプロテアーゼ, ヒドロラーゼ プレカイン, シグナル, チモージェン, タンパク質, 多量遺伝子: P69-S307 真核細胞及びウイルスのアスパルチル プロテアーゼ: DM00126 (Q03168) 19-385: R23-A395 アスパルチルプロテアーゼ: V93-V104, V278-L289 真核細胞アスパルチルプロテアーゼ: P69-A395 真核細胞及びウイルスのアスパラギン酸 プロテアーゼ BL00141: F91-S106, D184-S195, G235-G244, V278-L287, I370-A383	BLIMPS- PRINTS BLAST- PRODOM BLAST- DOMO MOTIFS HMMER-PFAM BLIMPS- BLOCKS

【表 8】

10

20

30

40

表 3-6

SEQ ID NO.	インサイトポリペプチド ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ(signature)配列, ドメイン及びモチーフ	分析方法及びデータベース
14	747204ICDI	145	T76 S13 S19 S37 T83 S105	N42 N54 N57 N94 N98 N131 N132	システインプロテアーゼ, 抑制因子: DM00182(P01035) I10: G30-G134 システインプロテアーゼ抑制因子: R66-T89 シグナルペプチド: M1-623 システインドメイン: G30-S133 システインプロテアーゼ抑制因子 シグネチャ: N53-S100	ELAST-DOMO BLIMPS-BLOCKS HMMEER SPScan HMMEER-PFAM ProfileScan

10

20

30

40

【表 9】

表4-1

ポリヌクレオチド SEQ ID NO.	インサイト ポリヌクレオチドID	配列長さ	選択された断片	配列断片	5'位置	3'位置
15	1714846CB1	4028	1349-1416, 1-199, 1903-3217	6831476H1 (SINTNCR01)	1	499
				6773219J1 (OVARDIR01)	650	1271
				6426758H1 (LUNGNON07)	998	1685
				1870084F6 (SKINBIT01)	1575	1995
				898127H1 (BRSTNOT05)	1964	2210
				643334H1 (LUNGNON07)	1999	2596
				4442573H1 (SINTNCR22)	2572	2868
				6286315H2 (EPIPUNA01)	2586	3110
				1714846F6 (UCMCNCR02)	3058	3631
				257076T6 (HMT2RAT01)	3300	3988
				6487217H1 (MIXDUNB01)	3635	4028
				95364348	385	839
				70152356V1	1	569
				70161001V1	359	824
				70157441V1	686	1218
16	1856589CB1	1422	539-570, 324-395, 1-214, 756-933	60106256B2	976	1422
				548654H1 (BEPINOT01)	1	268
				2170381F6 (ENDCNCR03)	150	675
				1437060F1 (PANCNCR08)	476	1031
				70098221V1	774	1352
				1428845H1 (SINTBST01)	1170	1408
				3290066H1 (BONRFET01)	1318	1569
				2994130H1 (KIDNFFET02)	1423	1715
				3601537H1 (DRGTNCR01)	1644	1850
				3702672H1 (PENCNCR07)	1760	1911
				754098R1 (BRAITUT02)	1	386
				70186361V1	143	847
				70186120V1	432	854
				3494839F6 (ADRETTUT07)	1	685
				70005795D1	660	1266
17	2617672CB1	1911	1-619	2630625T6 (COLNTPUT15)	708	1364
				605612H1 (BRSTTUT01)	1198	1385
18	2769104CB1	854	1-176			
19	4802789CB1	1386	1-23, 343-503			

10

20

30

40

【表10】

表4-2

ポリスクレオチド SEQ ID NO:	インサイト ポリスクレオチドID	配列長さ	選択された断片	配列断片	5' 位置	3' 位置
20	60116897CB1	3323	2502-2610, 1-735 1122-1879	3154611F6 (TLYMXTT02)	1	834
				60116918U1	740	1236
				2832568F6 (TLYMNOT03)	1119	1657
				2830930F7 (TLYMNOT03)	1610	2078
				6510679H1 (LUNGTA01)	1877	2180
				2849992F6 (BRSTTUT13)	2135	2641
				3200003F6 (PENCNOT02)	2368	2862
				2849992T6 (BRSTTUT13)	2792	3323
				3201617F6 (PENCNOT02)	1219	1713
				824817R1 (PROSNOT06)	1	551
				3257810H1 (PROSTUS08)	2004	2123
				5746464H1 (UTRSTUT05)	244	904
				3739625T6 (MENTNOT01)	1669	2075
21	1866356	2123	1-1590	258590R6 (HNT2RAI01)	642	1073
				6157882H1 (MONOTXN05)	1821	2092
				6289726H1 (BRAJFEN03)	1046	1705
				4570803H1 (GBLADIT02)	1	249
				267175H1 (HNT2NOT01)	555	927
				1442881T6 (THYRNOT03)	2234	2893
				1388162H1 (CARGDIT02)	1368	1619
				4662176H2 (BRSTTUT20)	1545	1809
				1344669H1 (PROSNOT11)	1714	1962
				SXEC01873V1	1898	2461
				449756R6 (TLYMNOT02)	219	734
				449756T6 (TLYMNOT02)	797	1459
				SXEC00314V1	1950	2515
22	1872095	2893	584-1266, 1-56, 2839-2893	267175H1 (HNT2NOT01)	555	927
				1442881T6 (THYRNOT03)	2234	2893

10

20

30

40

表4-3

ポリヌクレオチド SEQ ID NO:	インサイト ポリヌクレオチドID	配列長さ	選択された断片	配列断片	5'位置	3'位置
23	2278688	4170	1-245, 3069-3624, 1149-1809	2254713H1 (OVARTUT01) 097483R1 (PTUNOR01) 3271744H1 (BRAJNOT0) 4422961H1 (BRAPDIT01) 1378162H1 (LUNGNOT10) 3076825H1 (BONEUNT01) 3556490H1 (LUNGNOT31) 1368447H1 (SCORNON02) 4662177H2 (BRSTTUT20) 1 271 3853790H1 (BRAITUT12) 1877059H1 (LEUKNOT03) 2062 2330 1349282T1 (LATRTUT02) 72 578 4289627F6 (BRAEDIR01) 1289505T1 (BRAINOT11) 3530 4149 2373989F6 (ISLTNOT01) 1570 2094 097483F1 (PTUNOR01) 2933 3639 2698679H1 (UTRSNOT12) 1817 2134 2110561H1 (BRAITUT03) 669 950 3011419H1 (MUSCNOT07) 2369 2623 1394210H1 (THYRNCT03) 1001 1296 4880281H1 (UTRMNT01) 524 767 4043361F6 (LUNGNOT35) 1 593 6777288J1 (OVARDIR01) 436 1216 6121924H1 (BRAHNON05) 1022 1538 7032724H1 (BRAXTDR12) 1 480 4692968T6 (BRAENOT02) 636 1265 1871340F6 (SKINBIT01) 1256 1497 3429631T6 (SKINNOT04) 416 1476 7257324H1 (SKIRTC01) 1 474 GNN.96436155_002.edi 1 1194 GNN.95830433_004.edi 1 438	453 2611 1433 1258 1110 2127 759 3998 1 2411 2062 3814 72 3530 1570 2933 1817 669 2369 1001 524 1 436 1022 1 636 1256 416 1 1 1	710 3220 1667 1500 1309 2391 1064 4170 271 2705 2330 4157 578 4149 2094 3639 2134 950 2623 1296 767 593 1216 1538 480 1265 1497 1476 474 1194 438
24	4043361	767	1-66			
25	3937958	1538	385-506, 1-78, 1293-1538			
26	7257324	1497	651-770, 67-206			
27	7472038	1194	1-29, 788-1194			
28	7472041	438	1-27			

10

20

30

40

表5

ポリヌクレオチドSEQ ID NO:	インターネットプロジェクトID	代表的ライブラリ
15	1714846CB1	LUNGN07
16	1856589CB1	PROSN018
17	2617672CB1	PANCN008
18	2769104CB1	COLAN002
19	4802789CB1	ADRET007
20	60116897CB1	TYMNO03
21	1866356CB1	HNT2RAT01
22	1872095CB1	THYRNO03
23	2278688CB1	LATRT02
24	4043361CB1	LUNGN035
25	3937958CB1	KIDMNO05
26	7257324CB1	SKIMNO04

10

20

30

40

【表 1 3】

ライブラリ	ベクター	ライブラリの説明
LUNGNON07	pINCY	この標準化された肺組織ライブラリは、肺組織ライブラリから単離した RNA を用いて作製した。このライブラリは、著しく長い(48 時間/回) アニリングハイドリダイゼーションが用いられたの以外は Soares 他(1994)Proc. Natl. Acad. Sci. USA 91:9228-9232 及び Bonaldo 他(1996)Genome Res. 6:791 の条件に従って 2 回標準化した。
PROSN0T18	pINCY	ライブラリは、58 歳白人男性の根治膀胱切除及び根治前立腺切除、胃フィステル形成の際に採取した病変前立腺組織から単離した RNA を用いて作製した。病理学的には、腺線維筋腫性過形成であり、この組織はグレード 3 の移行性細胞癌に関連する。患者の病歴には、口峽炎及び気腫があった。家族歴には、急性心筋梗塞及びアテローム硬化性冠動脈疾患、II 型糖尿病があった。
PANCNOT08	pINCY	ライブラリは、65 歳の白人女性から根治膵臓癌全切除の際に採取した膵臓組織から単離した RNA を用いて作製した。関連する腫瘍組織は病理学的にはグレード 2 浸潤腺癌を示していた。患者の病歴には、II 型糖尿病、変形性関節症、心血管疾患、大腸の良性新生物、及び白内障が含まれる。過去の手術には、全脾切除術、胆嚢摘除術及び腹式子宮切開術が含まれる。家族歴には、心血管疾患、II 型糖尿病、及び胃癌が含まれる。
COLANOT02	pINCY	ライブラリは、25 歳白人女性の大腸の多数区域切除の際に病変上行結腸組織から単離した RNA を用いて作製した。病理学的には、全結腸切除検体を含み 2 cm の付着回腸を残した中程度から重度の活動性潰瘍性大腸炎を示していた。検体は、直腸近位からの連続関与を示し、著しい粘膜萎縮はあったが跳躍病変は見られなかった。検体には、顕微鏡的に密で優勢な粘膜炎症及び腺窩膿瘍がみられた。患者の病歴には、良性の大腸腫瘍があった。
ADRETUT07	pINCY	ライブラリは、43 歳の白人女性から片側副腎摘出の際に採取した副腎腫瘍組織から単離した RNA を用いて作製した。病理学的には、褐色細胞腫を示していた。

【表 1 4】

【表 15】

ライブラリ	ベクター	ライブラリの説明
TYMNOT03	pINCY	ライブラリは、活性でない Tn1 細胞から単離したポリ A RNA を用いて作製した。これらの細胞は、IL12 及び B7 が導入された COS 細胞で質的 CD4T 細胞から分化した。
HNT2RAT01	PBLUESCRIPT	ライブラリは、Stratagene (STR937231) において (関係付けられた神経性前駆体の性状特性を示すようなヒト奇形癌由来の) hNT2 細胞株から単離した RNA を用いて作製した。細胞は、レチン酸で 24 時間処理した。
LATRUT02	pINCY	ライブラリは、43 歳の白人男性から輸状成形術の際に左心房から採取した粘液腫から単離した RNA を用いて作製した。病理学的には、心房の粘液腫を示していた。患者の病歴には、肺不全、急性心筋梗塞、アテローム硬化性冠動脈疾患、高脂血症及び喫煙が含まれる。家族歴には、良性高血圧症、急性心筋梗塞、アテローム硬化性冠動脈疾患及び I 型糖尿病が含まれる。
LUNGNOT35	pINCY	ライブラリは、62 歳の白人女性より取り除いた肺組織より単離した RNA を用いて作製された。関連組織の病理学的には小結節を形成するグレード 1 の紡錘細胞カルチノイドを示した。病歴には、うつ病、血栓静脈炎、及び高脂血症が有った。家族歴には、脳血管障害、アテローム性冠動脈疾患、乳癌、結腸癌、2 型糖尿病、および悪性の皮膚黒色腫が有った。
THYRN0T03	pINCY	ライブラリは、28 歳の白人女性から完全甲状腺切除の際に左甲状腺から採取した甲状腺組織から単離した RNA で作製した。病理学的には、左甲状腺に存在する腺腫状過形成の小結節を示していた。関連腫瘍組織は、病理学的には左甲状腺によく被包化された腫瘍集塊を形成している優性小胞状腺腫を示していた。
KIDNNOT05	PSPORT1	このライブラリは、脳無酸素症で死亡した生後 2 日のヒスパニック系の女児の腎臓組織から単離した RNA を用いて作製した。家族歴には、先天性の心疾患があった。
SKINN0T04	pINCY	ライブラリは、70 歳の白人女性の乳房生検及び切除術の歳に採取された乳房皮膚組織から単離した RNA を用いて作製した。

表7-1

プログラム名	説明	引用文献	パラメーター閾値
ABI FACTURA	核酸配列においてベクター配列を除去して不特定の塩基をマスクするプログラム。	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	Fast Data Finder は、アミノ酸または核糖配列の比較及び注釈付け (annotation) に有用である。	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	不一致<50%
ABI AutoAssembler	核酸配列を構築するプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	Basic Local Alignment Search Tool は、アミノ酸及び核糖配列の配列類似性検索に有用であり、blastp 及び blastn、blastx、tblastn、tblastx の5つのファンクションがある。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. 他 (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: 確率値=1.0E-8 以下 完全長配列: 確率値=1.0E-10 以下
FASTA	Pearson 及び Lipman アルゴリズムは、問合わせの配列と同種の配列群との類似性を検索する。FASTA は、fasta 及び tfasta、fastx、tfastx、ssearch の少なくとも5つのファンクションを含む。	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E 値=1.06E-6 構築された ESTs: fasta 同一性=95%以上、一致長さ=200 塩基以上、fastx E 値=1.0E-8 以下 完全長配列: fastx スコア=100 以上
BLIMPS	Blocks IMPROVED Searcher は、BLOCKS 及び PRINTS、DOMO、PRODOM、PFAM データベースにおける配列に対して、ある配列の一致性を調べ、遺伝子ファミリー及び配列相同性、構造的アイデンティティ領域を探索する。	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.C: Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; Artwood, T.K. 他 (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	確率値=1.0E-3 以下
HMMER	PFAM などのタンパク質ファミリーコンセンサス配列の隠れマルコフモデル(HMM)に基づいたデータベースに対して問合せ配列を検索するアルゴリズム。	Krogh, A. 他 (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. 他 (1998) Nucleic Acids Res. 26:320-322; Durbin, R. 他 (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM ヒット: 確率値=1.0E-3 以下 シグナルペプチドヒット: スコア=0 以上

【表 16】

表7-2

プログラム名	説明	引用文献	パラメーター閾値
ProfileScan	Prosite で定義された配列パターンと一致するタンパク質配列における構造及び配列のモチーフを検索するアルゴリズム。	Gribskov, M. 他 (1988) CABIOS 4:61-66; Gribskov, 他 (1989) Methods Enzymol. 183:146-159; Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221	標準化された質のスコア≧特定の "HIGH" 値 通常、スコア=1.4-2.1
Phred	高い感度及び確率で自動配列決定機のトレースを調べる塩基読出しアルゴリズムである。	Ewing, B. 他 (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的なインプレメンテーションに基づく SWAT や GrossMatch を含む Phrap's Revised Assembly プログラムであって、配列相同性の検索及び DNA 配列の構築に有用である。	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; Green, P., University of Washington, Seattle, WA.	スコア=120 以上 一致長さ=56 以上
Consed	Phrap で構築したものの表示及び編集をするためのグラフィックツールである。	Gordon, D. 他 (1998) Genome Res. 8:195-202.	
SPScan	タンパク質配列をスキヤンして分泌シグナルペプチドの存在を調べる重み付けマトリクス解析プログラムである。	Nielson, H. 他 (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	スコア=3.5 以上
TMAP	蛋白配列での膜貫通セグメントの描写及び配向の決定のための、加重マトリクスを用いたプログラムである。	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	蛋白配列での膜貫通セグメントの描写及び配向の決定のための、隠された Markov モデル (HMM) を用いたプログラムである。	Sonnhammer, E.L. 他 (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et 他., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	Prosite で定義された配列と一致したパターンについてアミノ酸配列を検索するプログラムである。	Bairoch 他, (1997) Nucleic Acid Res. 25: 217-221; Wisconsin Package Program Manual, version 9, page M61-59, Genetics Computer Group, Madison, WI.	

【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
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number
WO 01/46443 A2

(51) International Patent Classification: C12N 15/57,
94B, A01K 67/00, C07K 16/40, C12Q 1/68, A61K 38/48,
G01N 33/50

83, Sunnyvale, CA 94086 (US); YUE, Henry (US/US),
826 Lois Avenue, Sunnyvale, CA 94087 (US); NGUYEN,
Daniel B. (US/US); 1403 Ridgewood Drive, San Jose,
CA 95118 (US); YANG, Y., Tom (CN/US); 4230 Sanwick
Court, San Jose, CA 95138 (US); YAO, Montague G.
(US/US); 111 Frederick Court, Mountain View, CA 94033
(US); LALL, Preeti (IN/US); 2382 Loss Drive, Santa Clara,
CA 95054 (US).

(21) International Application Number: PCT/US00/34811

(22) International Filing Date:
19 December 2000 (19.12.2000)

(25) Filing Language: English

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(26) Publication Language: English

(30) Priority Data:
60/172,053 23 December 1999 (23.12.1999) US
60/177,334 21 January 2000 (21.01.2000) US
60/178,884 28 January 2000 (28.01.2000) US
60/179,903 2 February 2000 (02.02.2000) US

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



(71) Applicant (for all designated States except US): INCYTE
GENOMICS, INC. (US/US); 3160 Porter Drive, Palo
Alto, CA 94304 (US).

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

(72) Inventors; and
(75) Inventors/Applicants (for US only): YANG, Jun-
ming (CN/US); 7125 Bark Lane, San Jose, CA 95129
(US); BAUGHN, Moriah, R. (US/US); 14244 Santiago
Road, San Leandro, CA 94577 (US); BURFORD, Neil
(GB/US); 105 Wildwood Circle, Durham, CT 06422
(US); AU-YOUNG, Janice (US/US); 233 Golden Eagle
Lane, Brea, CA 94005 (US); LEE, Dyoung, Aina, M.
(US/US); 233 Coy Drive, San Jose, CA 95123 (US);
REDDY, Roopa (IN/US); 1233 W. McKinley Avenue,

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 01/46443 A2

(54) Title: PROTEASIS

(57) Abstract: The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

WO 01/46443

PCT/US00/34811

PROTEASES**TECHNICAL FIELD**

This invention relates to nucleic acid and amino acid sequences of proteases and to the use of these sequences in the diagnosis, treatment, and prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

BACKGROUND OF THE INVENTION

Proteases cleave proteins and peptides at the peptide bond that forms the backbone of the protein or peptide chain. Proteolysis is one of the most important and frequent enzymatic reactions that occurs both within and outside of cells. Proteolysis is responsible for the activation and maturation of nascent polypeptides, the degradation of misfolded and damaged proteins, and the controlled turnover of peptides within the cell. Proteases participate in digestion, endocrine function, and tissue remodeling during embryonic development, wound healing, and normal growth. Proteases can play a role in regulatory processes by affecting the half life of regulatory proteins. Proteases are involved in the etiology or progression of disease states such as inflammation, angiogenesis, tumor dispersion and metastasis, cardiovascular disease, neurological disease, and bacterial, parasitic, and viral infections.

Proteases can be categorized on the basis of where they cleave their substrates. Exopeptidases, which include aminopeptidases, dipeptidyl peptidases, tripeptidases, carboxypeptidases, peptidyl-dipeptidases, dipeptidases, and omega peptidases, cleave residues at the termini of their substrates. Endopeptidases, including serine proteases, cysteine proteases, and metalloproteases, cleave at residues within the peptide. Four principal categories of mammalian proteases have been identified based on active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

Serine Proteases

The serine proteases (SPs) are a large, widespread family of proteolytic enzymes that include the digestive enzymes trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of macromolecules within the cell and in the extracellular matrix. Most of the more than 20 subfamilies can be grouped into six clans, each with a common ancestor. These six clans are hypothesized to have descended from at least four evolutionarily distinct ancestors. SPs are named for the presence of a serine residue found in the active

WO 01/46443

PCT/US00/34811

catalytic site of most families. The active site is defined by the catalytic triad, a set of conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates substrate binding. Other residues outside the active site form an oxyanion hole that stabilizes the tetrahedral transition intermediate formed during catalysis. SPs have a wide range of substrates and can be subdivided into subfamilies on the basis of their substrate specificity. The main subfamilies are named for the residue(s) after which they cleave: trypsinases (after arginine or lysine), asparaginases (after aspartate), chymotrypsinases (after phenylalanine or leucine), methioninases (methionine), and serinases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) *Meth. Enzymol.* 244:19-61).

Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by proteolysis. For example, trypsinogen is converted to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

The two largest SP subfamilies are the chymotrypsin (S1) and subtilisin (S8) families. Some members of the chymotrypsin family contain two structural domains unique to this family. Kringle domains are triple-looped, disulfide cross-linked domains found in varying copy number. Kringles are thought to play a role in binding mediators such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity (PROSITE PDOC00020). Apple domains are 90 amino-acid repeated domains, each containing six conserved cysteines. Three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines (PROSITE PDOC00376). Apple domains are involved in protein-protein interactions. S1 family members include trypsin, chymotrypsin, coagulation factors IX-XII, complement factors B, C, and D, granzymes, kallikrein, and tissue- and urokinase-plasminogen activators. The subtilisin family has members found in the eubacteria, archaeobacteria, eukaryotes, and viruses. Subtilisins include the proprotein-processing endopeptidases kexin and furin and the pituitary prohormone convertases PC1, PC2, PC3, PC6, and PACE4 (Rawlings and Barrett, *supra*).

SPs have functions in many normal processes and some have been implicated in the etiology or treatment of disease. Enterokinase, the initiator of intestinal digestion, is found in the intestinal brush border, where it cleaves the acidic propeptide from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:7588-7592). Prolylcarboxypeptidase, a lysosomal serine peptidase that cleaves peptides such as angiotensin II and III and [des-Arg⁹] bradykinin, shares sequence homology with members of both the serine carboxypeptidase and prolylendopeptidase families

WO 01/46443

PCT/US00/34811

(Tan, F. et al. (1993) *J. Biol. Chem.* 268:16631-16638). The protease neuropsin may influence synapse formation and neuronal connectivity in the hippocampus in response to neural signaling (Chen, Z.-L. et al. (1995) *J. Neurosci.* 15:5088-5097). Tissue plasminogen activator is useful for acute management of stroke (Zivin, J.A. (1999) *Neurology* 53:14-19) and myocardial infarction (Ross, A.M. (1999) *Clin. Cardiol.* 22:165-171). Some receptors (PAR, for proteinase-activated receptor), highly expressed throughout the digestive tract, are activated by proteolytic cleavage of an extracellular domain. The major agonists for PARs, thrombin, trypsin, and mast cell tryptase, are released in allergy and inflammatory conditions. Control of PAR activation by proteases has been suggested as a promising therapeutic target (Vergnolle, N. (2000) *Aliment. Pharmacol. Ther.* 14:257-266; Rice, K.D. et al. (1998) *Curr. Pharm. Des.* 4:381-396). Prostate-specific antigen (PSA) is a kallikrein-like serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. Serum PSA is elevated in prostate cancer and is the most sensitive physiological marker for monitoring cancer progression and response to therapy. PSA can also identify the prostate as the origin of a metastatic tumor (Brawer, M.K. and P.H. Lange (1989) *Urology* 33:11-16).

15 The signal peptidase is a specialized class of SP found in all prokaryotic and eukaryotic cell types that serves in the processing of signal peptides from certain proteins. Signal peptides are amino-terminal domains of a protein which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing, e.g., glycosylation or phosphorylation, activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals. The canine signal peptidase complex is composed of five subunits, all associated with the microsomal membrane and containing hydrophobic regions that span the membrane one or more times (Shelness, G.S. and G. Blobel (1990) *J. Biol. Chem.* 265:9512-9519). Some of these subunits serve to fix the complex in its proper position on the membrane while others contain the actual catalytic activity.

20 Another family of proteases which have a serine in their active site are dependent on the hydrolysis of ATP for their activity. These proteases contain proteolytic core domains and regulatory ATPase domains which can be identified by the presence of the P-loop, an ATP/GTP-binding motif (PROSITE PDOC00803). Members of this family include the eukaryotic mitochondrial matrix proteases, Clp protease and the proteasome. Clp protease was originally found in plant chloroplasts but is believed to be widespread in both prokaryotic and eukaryotic cells. The gene for early-onset torsion dystonia encodes a protein related to Clp protease (Ozelius, L.J. et al. (1998) *Adv. Neurol.* 78:93-105).

25 The proteasome is an intracellular protease complex found in some bacteria and in all eukaryotic cells, and plays an important role in cellular physiology. Proteasomes are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins of all

WO 01/46443

PCT/US00/34811

types, including proteins that function to activate or repress cellular processes such as transcription and cell cycle progression (Ciechanover, A. (1994) *Cell* 79:13-21). In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin, a small heat stable protein. The ubiquitinated protein is then recognized and degraded by the proteasome. The resultant ubiquitin-peptide complex is hydrolyzed by a ubiquitin carboxyl terminal hydrolase, and free ubiquitin is released for reutilization by the UCS. Ubiquitin-proteasome systems are implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes (p53), cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, *supra*). This pathway has been implicated in a number of diseases, including cystic fibrosis, Angelman's syndrome, and Liddle syndrome (reviewed in Schwartz, A.L. and A. Ciechanover (1999) *Annu. Rev. Med.* 50:57-74). A murine proto-oncogene, Ubp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells. The human homologue of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) *Oncogene* 10:2179-2183). Ubiquitin carboxyl terminal hydrolase is involved in the differentiation of a lymphoblastic leukemia cell line to a non-dividing mature state (Maki, A. et al. (1996) *Differentiation* 60:59-66). In neurons, ubiquitin carboxyl terminal hydrolase (PGP 9.5) expression is strong in the abnormal structures that occur in human neurodegenerative diseases (Lowe, J. et al. (1990) *J. Pathol.* 161:153-160). The proteasome is a large (~2000 kDa) multisubunit complex composed of a central catalytic core containing a variety of proteases arranged in four seven-membered rings with the active sites facing inwards into the central cavity, and terminal ATPase subunits covering the outer port of the cavity and regulating substrate entry (for review, see Schmidt, M. et al. (1999) *Curr. Opin. Chem. Biol.* 3:584-591).

Cysteine Proteases

Cysteine proteases (CPs) are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Nearly half of the CPs known are present only in viruses. CPs have a cysteine as the major catalytic residue at the active site where catalysis proceeds via a thioester intermediate and is facilitated by nearby histidine and asparagine residues. A glutamine residue is also important, as it helps to form an oxyanion hole. Two important CP families include the papain-like cysteases (C1) and the calpains (C2). Papain-like family members are generally lysosomal or secreted and therefore are synthesized with signal peptides as well as propeptides. Most members bear a conserved motif in the propeptide that may have structural significance (Karrer, K.M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:3063-3067). Three-dimensional structures of papain family members show a bilobed molecule with the catalytic site located between the two lobes. Papains include cathepsins B, C, H, L, and S, certain plant allergens and dipeptidyl peptidase (for a review, see

WO 01/46443

PCT/US00/34811

Rawlings, N.D. and A.J. Barrett (1994) *Meth. Enzymol.* 244:461-486).

Some CPs are expressed ubiquitously, while others are produced only by cells of the immune system. Of particular note, CPs are produced by monocytes, macrophages and other cells which migrate to sites of inflammation and secrete molecules involved in tissue repair. Overabundance of these repair molecules plays a role in certain disorders. In autoimmune diseases such as rheumatoid arthritis, secretion of the cysteine peptidase cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. Bone weakened by such degradation is also more susceptible to tumor invasion and metastasis. Cathepsin L expression may also contribute to the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium

10 (Keyszer, G.M. (1995) *Arthritis Rheum.* 38:976-984).

Calpains are calcium-dependent cytosolic endopeptidases which contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain. Calpain is expressed as a proenzyme heterodimer consisting of a catalytic subunit unique to each isoform and a regulatory subunit common to different isoforms. Each subunit bears a calcium-binding EF-hand domain. The regulatory subunit also contains a hydrophobic glycine-rich domain that allows the enzyme to associate with cell membranes. Calpains are activated by increased intracellular calcium concentration, which induces a change in conformation and limited autolysis. The resultant active molecule requires a lower calcium concentration for its activity (Chan, S.L. and M.P. Mattson (1999) *J. Neurosci. Res.* 58:167-190). Calpain expression is predominantly neuronal, although it is present in other tissues. Several chronic neurodegenerative disorders, including ALS, Parkinson's disease and Alzheimer's disease are associated with increased calpain expression (Chan and Mattson, *supra*). Calpain-mediated breakdown of the cytoskeleton has been proposed to contribute to brain damage resulting from head injury (McCracken, E. et al. (1999) *J. Neurotrauma* 16:749-761). Calpain-3 is predominantly expressed in skeletal muscle, and is responsible for limb-girdle muscular dystrophy type 2A (Minami, N. et al. (1999) *J. Neurol. Sci.* 171:31-37).

Another family of third proteases is the caspases, which are involved in the initiation and execution phases of apoptosis. A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues. Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention)

WO 01/46443

PCT/US00/34811

and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer. Endogenous caspase inhibitors (inhibitor of apoptosis proteins, or IAPs) also exist. All these interactions have clear effects on the control of apoptosis (reviewed in Chan and Mattson, *supra*; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

Caspases have been implicated in a number of diseases. Mice lacking some caspases have severe nervous system defects due to failed apoptosis in the neuroepithelium and suffer early lethality. Others show severe defects in the inflammatory response, as caspases are responsible for processing IL-1 β and possibly other inflammatory cytokines (Chan and Mattson, *supra*). Cowpox virus and baculoviruses target caspases to avoid the death of their host cell and promote successful infection. In addition, increases in inappropriate apoptosis have been reported in AIDS, neurodegenerative diseases and ischemic injury, while a decrease in cell death is associated with cancer (Salveson and Dixit, *supra*; Thompson, C.B. (1995) Science 267:1456-1462).

Aspartyl proteases

Aspartyl proteases (APs) include the lysosomal proteases cathepsins D and E, as well as chymosin, renin, and the gastric pepsins. Most retroviruses encode an AP, usually as part of the pol polyprotein. APs, also called acid proteases, are monomeric enzymes consisting of two domains, each domain containing one half of the active site with its own catalytic aspartic acid residue. APs are most active in the range of pH 2-3, at which one of the aspartate residues is ionized and the other neutral. The pepsin family of APs contains many secreted enzymes, and all are likely to be synthesized with signal peptides and propeptides. Most family members have three disulfide loops, the first ~5 residue loop following the first aspartate, the second 5-6 residue loop preceding the second aspartate, and the third and largest loop occurring toward the C terminus. Retropepsins, on the other hand, are analogous to a single domain of pepsin, and become active as homodimers with each retropepsin monomer contributing one half of the active site. Retropepsins are required for processing the viral polyproteins.

APs have roles in various tissues, and some have been associated with disease. Renin mediates the first step in processing the hormone angiotensin, which is responsible for regulating electrolyte balance and blood pressure (reviewed in Crews, D.E. and S.R. Williams (1999) Hum. Biol. 71:475-503). Abnormal regulation and expression of cathepsins are evident in various inflammatory disease

WO 01/46443

PCT/US00/34811

states. Expression of cathepsin D is elevated in synovial tissues from patients with rheumatoid arthritis and osteoarthritis. The increased expression and differential regulation of the cathepsins are linked to the metastatic potential of a variety of cancers (Chambers, A.F. et al. (1993) *Crit. Rev. Oncol.* 4:95-114).

5 Metalloproteases

Most zinc-dependent metalloproteases share a common sequence in the zinc-binding domain. The active site is made up of two histidines which act as zinc ligands and a catalytic glutamic acid C-terminal to the first histidine. Proteins containing this signature sequence are known as the metzincins and include aminopeptidase N, angiotensin-converting enzyme, neurilysin, the matrix metalloproteases and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which

10 all three conserved residues – two histidines and a glutamic acid – are involved in zinc binding.

A number of the neutral metalloendopeptidases, including angiotensin converting enzyme and the aminopeptidases, are involved in the metabolism of peptide hormones. High aminopeptidase B activity, for example, is found in the adrenal glands and neurohypophyses of hypertensive rats (Prieto, I. et al. (1998) *Horm. Metab. Res.* 30:246-248). Oligopeptidase M/neurilysin can hydrolyze

15 bradykinin as well as neurotensin (Serizawa, A. et al. (1995) *J. Biol. Chem.* 270:2092-2098). Neurotensin is a vasoactive peptide that can act as a neurotransmitter in the brain, where it has been implicated in limiting food intake (Tritos, N.A. et al. (1999) *Neuropeptides* 33:339-349).

The matrix metalloproteases (MMPs) are a family of at least 23 enzymes that can degrade

20 components of the extracellular matrix (ECM). They are Zn²⁺ endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain which can bind to substrate molecules in the ECM or to inhibitors produced by the tissue (TIMPs, for tissue inhibitor of metalloprotease; Campbell, I.L. et al. (1999) *Trends Neurosci.* 22:285). The presence of fibronectin-like repeats, transmembrane domains, or C-terminal hemopexinase-like domains can be used

25 to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In the inactive form, the Zn²⁺ ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the Zn²⁺-cysteine interaction, or "cysteine switch," exposing the active site. This partially activates the enzyme, which then cleaves off its propeptide and becomes fully active. MMPs are often activated by the serine proteases plasmin and furin. MMPs are often regulated by

30 stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is very important in tissue homeostasis (reviewed in Young, V.W. et al. (1998) *Trends Neurosci.* 21:75).

MMPs are implicated in a number of diseases including osteoarthritis (Mitchell, P. et al. (1996) *J. Clin. Invest.* 97:761), atherosclerotic plaque rupture (Sukhova, G.K. et al. (1999) *Circulation* 99:2503), aortic aneurysm (Schneiderman, J. et al. (1998) *Am. J. Path.* 152:703), non-healing wounds

WO 01/46443

PCT/US00/34811

- (Saarialho-Kere, U.K. et al. (1994) *J. Clin. Invest.* 94:79), bone resorption (Blavier, L. and J.M. Delaisse (1995) *J. Cell Sci.* 108:3649), age-related macular degeneration (Stecu, B. et al. (1996) *Invest. Ophthalmol. Vis. Sci.* 39:2194), emphysema (Finlay, G.A. et al. (1997) *Thorax* 52:502), myocardial infarction (Rohde, L.E. et al. (1999) *Circulation* 99:3063) and dilated cardiomyopathy (Thomas, C.V. et al. (1998) *Circulation* 97:1708). MMP inhibitors prevent metastasis of mammary carcinoma and experimental tumors in rat, and Lewis lung carcinoma, hemangioma, and human ovarian carcinoma xenografts in mice (Eccles, S.A. et al. (1996) *Cancer Res.* 56:2815; Anderson et al. (1996) *Cancer Res.* 56:715-718; Volpert, O.V. et al. (1996) *J. Clin. Invest.* 98:671; Taraboleti, G. et al. (1995) *J. NCJ* 87:293; Davies, B. et al. (1993) *Cancer Res.* 53:2087). MMPs may be active in Alzheimer's disease.
- 10 A number of MMPs are implicated in multiple sclerosis, and administration of MMP inhibitors can relieve some of its symptoms (reviewed in Yong, *supra*).

Another family of metalloproteases is the ADAMs, for A Disintegrin and Metalloprotease Domain, which they share with their close relatives the adamalysins, snake venom metalloproteases (SVMs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing

15 a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

- 20 ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlönörff, J. and C.P. Blobel (1999) *J. Cell. Sci.* 112:3603-3617). The Kazbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), activating the program for lateral inhibition in *Drosophila* neural development. Two ADAMs, TACE
- 25 (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlönörff and Blobel, *supra*). TACE has also been identified as the TNF activating enzyme (Black, R.A. et al. (1997) *Nature* 385:729). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound
- 30 pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules.

The ADAMTS sub-family has all of the features of ADAM family metalloproteases and contain an additional thrombospondin domain (TS). The prototypic ADAMTS was identified in mouse, found to be expressed in heart and kidney and upregulated by proinflammatory stimuli (Kuno, K. et al.

WO 01/46443

PCT/US00/34811

(1997) *J. Biol. Chem.* 272:556). To date eleven members are recognized by the Human Genome Organization (HUGO; <http://www.gene.ucl.ac.uk/users/niester/adants.html#Approved>). Members of this family have the ability to degrade aggrecan, a high molecular weight proteoglycan which provides cartilage with important mechanical properties including compressibility, and which is lost during the development of arthritis. Enzymes which degrade aggrecan are thus considered attractive targets to prevent and slow the degradation of articular cartilage (See, e.g., Tortorella, M.D. (1999) *Science* 284:1664; Abbaszade, I. (1999) *J. Biol. Chem.* 274:23443). Other members are reported to have antiangiogenic potential (Kuno et al., *supra*) and/or procollagen processing (Colige, A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2374).

10 Protease inhibitors

Protease inhibitors and other regulators of protease activity control the activity and effects of proteases. Protease inhibitors have been shown to control pathogenesis in animal models of proteolytic disorders (Murphy, G. (1991) *Agents Actions Suppl.* 35:69-76). Low levels of the cystatins, low molecular weight inhibitors of the cysteine proteases, correlate with malignant progression of tumors (Calkins, C. et al. (1995) *Biol. Biochem. Hoppe Seyler* 376:71-80). Serpins are inhibitors of mammalian plasma serine proteases. Many serpins serve to regulate the blood clotting cascade and/or the complement cascade in mammals. Sp32 is a positive regulator of the mammalian acrosomal protease, acrosin, that binds the proenzyme, proacrosin, and thereby aids in packaging the enzyme into the acrosomal matrix (Baba, T. et al. (1994) *J. Biol. Chem.* 269:10133-10140). The Kunitz family of serine protease inhibitors are characterized by one or more "Kunitz domains" containing a series of cysteine residues that are regularly spaced over approximately 50 amino acid residues and form three intrachain disulfide bonds. Members of this family include aprotinin, tissue factor pathway inhibitor (TFPI-1 and TFPI-2), inter- α -trypsin inhibitor, and bikunin. (Marlor, C.W. et al. (1997) *J. Biol. Chem.* 272:12202-12208.) Members of this family are potent inhibitors (in the nanomolar range) against serine proteases such as kallikrein and plasmin. Aprotinin has clinical utility in reduction of perioperative blood loss.

The discovery of new proteases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

SUMMARY OF THE INVENTION

WO 01/46443

PCT/US00/34811

The invention features purified polypeptides, proteases, referred to collectively as "PRTS" and individually as "PRTS-1," "PRTS-2," "PRTS-3," "PRTS-4," "PRTS-5," "PRTS-6," "PRTS-7," "PRTS-8," "PRTS-9," "PRTS-10," "PRTS-11," "PRTS-12," "PRTS-13," and "PRTS-14." In one aspect, the invention provides an isolated polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-14.

The invention further provides an isolated polynucleotide encoding a polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-14. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:15-28.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c)

WO 01/46443

PCT/US00/34811

a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said

WO 01/46443

PCT/US00/34811

target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

10 The invention further provides a composition comprising an effective amount of a polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group
15 consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need
20 of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the
25 group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound
30 identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting
35 of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally

WO 01/46443

PCT/US00/34811

occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method
5 comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment
10 the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group
15 consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the
20 polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide 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:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the
25 group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test
30 compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in
35 altering expression of a target polynucleotide, wherein said target polynucleotide comprises a

WO 01/46443

PCT/US00/34811

sequence selected from the group consisting of SEQ ID NO:15-28, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

25

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for each polypeptide of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of each polypeptide sequence, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of each polypeptide.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble each

35

WO 01/6443

PCT/US00/34811

polynucleotide sequence, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for each polynucleotide of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

5 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

10 Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

15 It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

20 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in
25 connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PRTS" refers to the amino acid sequences of substantially purified PRTS obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and
30 human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PRTS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

WO 01/46443

PCT/US00/34811

An "allelic variant" is an alternative form of the gene encoding PRTS. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PRTS include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PRTS or a polypeptide with at least one functional characteristic of PRTS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRTS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PRTS. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRTS. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PRTS is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PRTS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS

WO 01/46443

PCT/US00/34811

participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant.

Antibodies that bind PRTS polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PRTS, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

WO 01/46443

PCT/US00/34811

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PRTS or fragments of PRTS may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Pfrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
25	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
30	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
35	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
40	Val	Ile, Leu, Thr

WO 01/46443

PCT/US00/34811

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

5 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

15 A "fragment" is a unique portion of PRTS or the polynucleotide encoding PRTS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 20 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the 25 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:15-28 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:15-28, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:15-28 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish 30 SEQ ID NO:15-28 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:15-28 and the region of SEQ ID NO:15-28 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-14 is encoded by a fragment of SEQ ID NO:15-28. A fragment 35 of SEQ ID NO:1-14 comprises a region of unique amino acid sequence that specifically identifies

WO 01/46443

PCT/US00/34811

SEQ ID NO:1-14. For example, a fragment of SEQ ID NO:1-14 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-14. The precise length of a fragment of SEQ ID NO:1-14 and the region of SEQ ID NO:1-14 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at

<http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to

WO 01/46443

PCT/US00/34811

compare two nucleotide sequences, one may use Blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April 21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

5 *Penalty for mismatch: -2*

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

10 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktupic=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

WO 01/46443

PCT/US00/34811

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

5 *Open Gap: 11 and Extension Gap: 1 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

10 Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length
15 supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

20 The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific
25 hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for
30 annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

WO 01/46443

PCT/US00/34811

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY, specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0f analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PRTS which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of

WO 01/46443

PCT/US00/34811

PRTS which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PRTS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRTS.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PRTS may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PRTS.

"Probe" refers to nucleic acid sequences encoding PRTS, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous

WO 01/46443

PCT/US00/34811

nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

5 Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR
10 Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such
15 purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South
20 West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriiming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for
25 microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences.
30 Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to

WO 01/46443

PCT/US00/34811

those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PRTS, nucleic acids encoding PRTS, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will

WO 01/46443

PCT/US00/34811

reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which
5 they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic heads, gels, tubing, plates, polymers,
10 microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient
15 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells"
20 includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor
25 of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria,
30 cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989),
35 supra.

WO 01/6443

PCT/US00/34811

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 5 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may 10 possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in 15 which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 20 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

25 The invention is based on the discovery of new human proteases (PRTS), the polynucleotides encoding PRTS, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide 30 sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO.) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO.) and an 35 Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

WO 01/6443

PCT/US00/34811

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpep) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO.) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the GenBank identification number (Genbank ID NO.) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of each of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO.) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO.) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:15-28 or that distinguish between SEQ ID NO:15-28 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7032724H1 is the identification number of an Incyte cDNA sequence, and BRAXTDR 12 is the cDNA library from which

WO 01/46443

PCT/US00/34811

it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70152356V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5364348) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6436155_002.edit is the identification number of a Genscan-predicted coding sequence, with g6436155 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PRTS variants. A preferred PRTS variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PRTS amino acid sequence, and which contains at least one functional or structural characteristic of PRTS.

The invention also encompasses polynucleotides which encode PRTS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28, which encodes PRTS. The polynucleotide sequences of SEQ ID NO:15-28, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PRTS. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PRTS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28 which has at

WO 01/46443

PCT/US00/34811

least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:15-28. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRTS.

5 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRTS, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRTS, and all such variations are to be considered as being specifically disclosed.

10 Although nucleotide sequences which encode PRTS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PRTS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRTS or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PRTS and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

15 The invention also encompasses production of DNA sequences which encode PRTS and PRTS derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRTS or any fragment thereof.

20 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:15-28 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

25 Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of

WO 01/46443

PCT/US00/34811

DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PRTS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sackar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been

WO 01/66443

PCT/US00/34811

size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

5 Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire
10 process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which
15 encode PRTS may be cloned in recombinant DNA molecules that direct expression of PRTS, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PRTS.

The nucleotide sequences of the present invention can be engineered using methods generally
20 known in the art in order to alter PRTS-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites,
25 alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,453; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Cramer, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or
30 improve the biological properties of PRTS, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of
35 DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial"

WO 01/6443

PCT/US00/34811

breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby
5 maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PRTS may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively,
10 PRTS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp.55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid
15 sequence of PRTS, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Rogner (1990) *Methods Enzymol.* 182:392-421.)
20 The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active PRTS, the nucleotide sequences encoding PRTS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a
25 suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PRTS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PRTS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where
30 sequences encoding PRTS and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural
35 and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate

WO 01/46443

PCT/US00/34811

for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PRTS and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PRTS. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Hecke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Butler, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PRTS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PRTS can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PRTS into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Hecke, G. and S.M. Schuster (1989) *J. Biol. Chem.*

WO 01/46443

PCT/US00/34811

264:5503-5509.) When large quantities of PRTS are needed, e.g. for the production of antibodies, vectors which direct high level expression of PRTS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PRTS. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of PRTS. Transcription of sequences encoding PRTS may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRTS may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PRTS in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PRTS in cell lines is preferred. For example, sequences encoding PRTS can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous

WO 01/46443

PCT/US00/34811

expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorosulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β -glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

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

In general, host cells that contain the nucleic acid sequence encoding PRTS and that express PRTS may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PRTS using either

WO 01/46443

PCT/US00/34811

specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PRTS is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) *Immunocytochemical Protocols*, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRTS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PRTS, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radioisotopes, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRTS may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PRTS may be designed to contain signal sequences which direct secretion of PRTS through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and W138) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing

WO 01/46443

PCT/US00/34811

of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRTS may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PRTS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PRTS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PRTS encoding sequence and the heterologous protein sequence, so that PRTS may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PRTS may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

PRTS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PRTS. At least one and up to a plurality of test compounds may be screened for specific binding to PRTS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PRTS, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PRTS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PRTS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E.*

WO 01/46443

PCT/US00/34811

coli. Cells expressing PRTS or cell membrane fractions which contain PRTS are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PRTS or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PRTS, either in solution or affixed to a solid support, and detecting the binding of PRTS to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PRTS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PRTS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PRTS activity, wherein PRTS is combined with at least one test compound, and the activity of PRTS in the presence of a test compound is compared with the activity of PRTS in the absence of the test compound. A change in the activity of PRTS in the presence of the test compound is indicative of a compound that modulates the activity of PRTS. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising PRTS under conditions suitable for PRTS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PRTS may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PRTS or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the

WO 01/46443

PCT/US00/34811

resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PRTS may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PRTS can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PRTS is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PRTS, e.g., by secreting PRTS in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PRTS and proteases. In addition, the expression of PRTS is closely associated with gastrointestinal, epithelial, reproductive, cardiovascular, cancerous, and inflamed tissues, and with normal kidney and normal skin tissues. Therefore, PRTS appears to play a role in gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders. In the treatment of disorders associated with increased PRTS expression or activity, it is desirable to decrease the expression or activity of PRTS. In the treatment of disorders associated with decreased PRTS expression or activity, it is desirable to increase the expression or activity of PRTS.

Therefore, in one embodiment, PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel

WO 01/46443

PCT/US00/34811

syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome
 (AIDS) encephalopathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis,
 hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing
 cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis
 5 hepatitis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of
 pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias,
 adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis,
 hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins,
 thrombophlebitis and phlebotrombosis, vascular tumors, and complications of thrombolysis, balloon
 10 angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure,
 ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative
 valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular
 calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis,
 nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart
 15 disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease,
 and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired
 immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies,
 ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture,
 autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-
 20 ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic
 dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with
 lymphocytotoxicity, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,
 Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypercosinophilia, irritable
 bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation,
 25 osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis,
 Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis,
 systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis,
 Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral,
 bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative
 30 disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed
 connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia
 vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia,
 lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal
 gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract,

WO 01/46443

PCT/US00/34811

heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAOR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata,

25 Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic

30 lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Strausler-Scheinker syndrome, fatal familial

WO 01/6443

PCT/US00/34811

insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, cephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, cataplexy, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PRTS in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PRTS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those listed above.

In a further embodiment, an antagonist of PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS. Examples of such disorders include, but are not limited to, those gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PRTS may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PRTS.

WO 01/46443

PCT/US00/34811

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PRTS may be produced using methods which are generally known in the art. In particular, purified PRTS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PRTS. Antibodies to PRTS may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRTS or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PRTS have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PRTS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PRTS may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kohbor, D. et al. (1985) *J.*

WO 01/46443

PCT/US00/34811

Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PRTS-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PRTS may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PRTS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRTS epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PRTS. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of PRTS-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a , determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PRTS epitopes, represents the average affinity, or avidity, of the antibodies for PRTS. The K_a , determined for

WO 01/46443

PCT/US00/34811

a preparation of monoclonal antibodies, which are nonspecific for a particular PRTS epitope, represents a true measure of affinity. High-affinity antibody preparations with K_d ranging from about 10^8 to 10^{12} L/mole are preferred for use in immunoassays in which the PRTS-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_d ranging from about 10^6 to 10^9 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PRTS, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PRTS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PRTS, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PRTS. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PRTS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PRTS may be used for

WO 01/6443

PCT/US00/34811

somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PRTS expression or regulation causes disease, the expression of PRTS from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PRTS are treated by constructing mammalian expression vectors encoding PRTS and introducing these vectors by mechanical means into PRTS-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of PRTS include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, FEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, FTK-HYG (Clontech, Palo Alto CA). PRTS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998)

WO 01/46443

PCT/US00/34811

Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the doxycycline-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PRTS from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PRTS expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PRTS under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PHBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Arraentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PRTS to cells which have one or more genetic abnormalities with respect to

WO 01/46443

PCT/US00/34811

the expression of PRTS. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Ann. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PRTS to target cells which have one or more genetic abnormalities with respect to the expression of PRTS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PRTS to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PRTS to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Caroff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g.,

WO 01/46443

PCT/US00/34811

protease and polymerase). Similarly, inserting the coding sequence for PRTS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PRTS-coding RNAs and the synthesis of high levels of PRTS in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of PRTS into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.J. and B.J. Carr, *Molecular and Immunologic Approaches*, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PRTS.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozymic cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for

WO 01/46443

PCT/US00/34811

chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding PRTS. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize
5 complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

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

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PRTS. Compounds
15 which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of
20 polynucleotide expression. Thus, in the treatment of disorders associated with increased PRTS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PRTS may be therapeutically useful, and in the treatment of disorders associated with decreased PRTS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PRTS may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in
25 altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound
30 based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PRTS is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PRTS are assayed by
35 any method commonly known in the art. Typically, the expression of a specific nucleotide is detected

WO 01/46443

PCT/US00/34811

by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PRTS. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PRTS, antibodies to PRTS, and mimetics, agonists, antagonists, or inhibitors of PRTS.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting

WO 01/66443

PCT/US00/34811

formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle
5 injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of
10 macromolecules comprising PRTS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PRTS or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et
15 al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route
20 of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PRTS or fragments thereof, antibodies of PRTS, and agonists, antagonists or inhibitors of PRTS, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by
25 calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD₅₀/ED₅₀ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is
30 preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active

WO 01/46443

PCT/US00/34811

moieity or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy.

5 Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their
10 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PRTS may be used for the diagnosis of disorders characterized by expression of PRTS, or in assays to monitor patients being treated with
15 PRTS or agonists, antagonists, or inhibitors of PRTS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PRTS include methods which utilize the antibody and a label to detect PRTS in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter
20 molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PRTS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PRTS expression. Normal or standard values for PRTS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PRTS under
25 conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PRTS expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRTS may be used for
30 diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PRTS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PRTS, and to monitor regulation of PRTS levels during therapeutic intervention.

WO 01/46443

PCT/US00/34811

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PRTS or closely related molecules may be used to identify nucleic acid sequences which encode PRTS. The specificity of the probe, whether it is made from a highly specific region, e.g., the 3' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PRTS, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PRTS encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:15-28 or from genomic sequences including promoters, enhancers, and introns of the PRTS gene.

Means for producing specific hybridization probes for DNAs encoding PRTS include the cloning of polynucleotide sequences encoding PRTS or PRTS derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ³²P or ³⁵S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PRTS may be used for the diagnosis of disorders associated with expression of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula,

WO 01/46443

PCT/US00/34811

atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebotrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an

10 autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxicity, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hyperesinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's

20 syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal

25 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder,

30 such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucocutaneous dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism.

WO 01/46443

PCT/US00/34811

hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyslidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma,

10 dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions,

15 chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix,

20 trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases,

25 bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental

30 retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic

WO 01/46443

PCT/US00/34811

disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of 5 prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, 10 impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PRTS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRTS expression. Such qualitative or quantitative methods are well known in the art.

15 In a particular aspect, the nucleotide sequences encoding PRTS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PRTS may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard 20 value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRTS in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

25 In order to provide a basis for the diagnosis of a disorder associated with expression of PRTS, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRTS, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with 30 values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated,

WO 01/46443

PCT/US00/34811

hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

5 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further
10 progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRTS may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding PRTS, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRTS, and will be
15 employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are
20 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal
25 tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplicons in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP
30 (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high

WO 01/46443

PCT/US00/34811

throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PRTS include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PRTS, fragments of PRTS, or antibodies specific for PRTS may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seifhaner et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies,

WO 01/46443

PCT/US00/34811

or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, B.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and

WO 01/46443

PCT/US00/34811

analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PRTS to quantify the levels of PRTS expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendez, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteomic level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seifhamer (1997) *Electrophoresis* 18:533-537), so proteomic toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profiles. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the

WO 01/46443

PCT/US00/34811

individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in *DNA Microarrays: A Practical Approach*, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PRTS may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PRTS on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as

WO 01/46443

PCT/US00/34811

linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PRTS, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRTS and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PRTS, or fragments thereof, and washed. Bound PRTS is then detected by methods well known in the art. Purified PRTS can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PRTS specifically compete with a test compound for binding PRTS. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRTS.

In additional embodiments, the nucleotide sequences which encode PRTS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in

WO 01/46443

PCT/US00/34811

particular U.S. Ser. No. 60/172,055, U.S. Ser. No. 60/177,334, U.S. Ser. No. 60/178,884, and U.S. Ser. No. 60/179,903, are expressly incorporated by reference herein.

EXAMPLES

5 I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. The Incyte cDNAs shown for SEQ ID NO:15 were derived from cDNA libraries constructed from small intestine, ovary, lung, skin, breast, prostate epithelium, and mixed myometrial tissues, umbilical cord blood, and teratocarcinoma cells
10 which contained neuronal precursors. The Incyte cDNAs shown for SEQ ID NO:17 were derived from cDNA libraries constructed from a bronchial epithelium primary cell line, dermal microvascular endothelial cells, pancreas, ileum tissue associated with Crohn's disease, rib bone tissue associated with Patau's syndrome, kidney, thoracic dorsal root ganglion, and penis corpus cavernosum tissue. The Incyte cDNA shown for SEQ ID NO:18 was derived from a cDNA library constructed from brain
15 tumor tissue. The Incyte cDNAs shown for SEQ ID NO:19 were derived from cDNA libraries constructed from adrenal gland, colon, and breast tissue. The Incyte cDNAs shown for SEQ ID NO:20 were derived from cDNA libraries constructed from T-lymphocytes, lung, breast, and penis corpus cavernosum tissues. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIzol
20 (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)⁺ RNA was isolated
25 using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth, CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin, TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA
30 libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIP^T plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the

WO 01/46443

PCT/US00/34811

appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,

5 PBLUESCRIPT plasmid (Stratagene), PSPORI1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), FBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

10 II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL

15 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) *Anal. Biochem.* 216:1-14). Host cell lysis and thermal

20 cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the

25 MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as

30 the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI

35 PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading

WO 01/46443

PCT/US00/34811

frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or GenScan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Coned, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multiple sequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match

WO 01/46443

PCT/US00/34811

between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:15-28. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative proteases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpr1 and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteases, the encoded polypeptides were analyzed by querying against PFAM motifs for proteases. Potential proteases were also identified by homology to Incyte cDNA sequences that had been annotated as proteases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr1 public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

30 "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory

WO 01/46443

PCT/US00/34811

and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of PRTS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:15-28 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:15-28 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for

WO 01/46443

PCT/US00/34811

Genome Research (WIGR), and Génethon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Génethon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (ocyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and this product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For

WO 01/46443

PCT/US00/34811

example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

5 Alternatively, polynucleotide sequences encoding PRTS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system;

10 embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer,

15 cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PRTS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

20 VIII. Extension of PRTS Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using

25 OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension

30 was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme

WO 01/46443

PCT/US00/34811

(Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair U7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides

WO 01/46443

PCT/US00/34811

designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:15-28 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Cl of [γ - 32 P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing. See, e.g., Baldeschwieler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schna (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schna, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array

WO 01/46443

PCT/US00/34811

elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

10 Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse
15 transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified
20 using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

25 Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5
30 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR
35 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and

WO 01/46443

PCT/US00/34811

coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

WO 01/46443

PCT/US00/34811

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835II analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

20 XI. Complementary Polynucleotides

Sequences complementary to the PRTS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRTS. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PRTS. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PRTS-encoding transcript.

XII. Expression of PRTS

30 Expression and purification of PRTS is achieved using bacterial or virus-based expression systems. For expression of PRTS in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*lac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.

35 Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic

WO 01/46443

PCT/US00/34811

resistant bacteria express PRTS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PRTS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PRTS

5 by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA

10 91:3224-3227; Saadig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PRTS is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized

15 glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PRTS at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins

20 (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified PRTS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX, where applicable.

XIII. Functional Assays

PRTS function is assessed by expressing the sequences encoding PRTS at physiologically

25 elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome

30 formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser opties-

WO 01/46443

PCT/US00/34811

based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PRTS on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PRTS and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PRTS and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of PRTS Specific Antibodies

PRTS substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PRTS amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PRTS activity by, for example, binding the peptide or PRTS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

WO 01/46443

PCT/US00/34811

XV. Purification of Naturally Occurring PRTS Using Specific Antibodies

Naturally occurring or recombinant PRTS is substantially purified by immunoaffinity chromatography using antibodies specific for PRTS. An immunoaffinity column is constructed by covalently coupling anti-PRTS antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRTS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PRTS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRTS binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PRTS is collected.

XVI. Identification of Molecules Which Interact with PRTS

PRTS, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PRTS, washed, and any wells with labeled PRTS complex are assayed. Data obtained using different concentrations of PRTS are used to calculate values for the number, affinity, and association of PRTS with the candidate molecules.

Alternatively, molecules interacting with PRTS are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PRTS may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of PRTS Activity

Protease activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) *Proteolytic Enzymes: A Practical Approach*, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are

WO 01/46443

PCT/US00/34811

performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

5 An alternate assay for ubiquitin hydrolase activity measures the hydrolysis of a ubiquitin precursor. The assay is performed at ambient temperature and contains an aliquot of PRTS and the appropriate substrate in a suitable buffer. Chemically synthesized human ubiquitin-valine may be used as substrate. Cleavage of the C-terminal valine residue from the substrate is monitored by capillary electrophoresis (Franklin, K. et al. (1997) *Anal. Biochem.* 247:305-309).

10 In the alternative, an assay for protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for PRTS is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5
15 to RSGFP4. When the fusion protein is incubated with PRTS, the substrate is cleaved, and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in energy transfer which is quantified by comparing the emission spectra before and after the addition of PRTS (Mitra, R.D. et al. (1996) *Gene* 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and PRTS is introduced on an inducible vector so
20 that FRET can be monitored in the presence and absence of PRTS (Sagot, I. et al. (1999) *FEBS Lett.* 447:53-57).

XVIII. Identification of PRTS Substrates

Phage display libraries can be used to identify optimal substrate sequences for PRTS. A random hexamer followed by a linker and a known antibody epitope is cloned as an N-terminal
25 extension of gene III in a filamentous phage library. Gene III codes for a coat protein, and the epitope will be displayed on the surface of each phage particle. The library is incubated with PRTS under proteolytic conditions so that the epitope will be removed if the hexamer codes for a PRTS cleavage site. An antibody that recognizes the epitope is added along with immobilized protein A. Uncleaved phage, which still bear the epitope, are removed by centrifugation. Phage in the supernatant are then
30 amplified and undergo several more rounds of screening. Individual phage clones are then isolated and sequenced. Reaction kinetics for these peptide substrates can be studied using an assay in Example XVII, and an optimal cleavage sequence can be derived (Ke, S.H. et al. (1997) *J. Biol. Chem.* 272:16603-16609).

To screen for in vivo PRTS substrates, this method can be expanded to screen a cDNA

WO 01/46443

PCT/US00/34811

expression library displayed on the surface of phage particles (T7SELECT™10-3 Phage display vector, Novagen, Madison, WI) or yeast cells (pYD1 yeast display vector kit, Invitrogen, Carlsbad, CA). In this case, entire cDNAs are fused between Gene III and the appropriate epitope.

XIX. Identification of PRIS Inhibitors

5 Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PRIS activity is measured for each well and the ability of each compound to inhibit PRIS activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PRIS activity.

10 In the alternative, phage display libraries can be used to screen for peptide PRIS inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PRIS and incubated with a random peptide phage display library or a cyclic peptide library (Koivunen, E. et al. (1999) Nat. Biotechnol. 17:768-774). Unbound phage are washed away and selected phage amplified and rescreened for several more rounds. Candidates are tested for
15 PRIS inhibitory activity using an assay described in Example XVII.

Various modifications and variations of the described methods and systems 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 certain embodiments, it should be

20 understood that the invention as claimed should not be unduly limited to such specific embodiments.

Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

WO 01/46443

PCT/US00/34811

Table 1

Project ID	Peptide SEQ ID NO.	Peptide ID	Peptide SEQ ID NO.	Peptide ID
1714816	1	1714844CD1	15	1714846CB1
1856589	2	1856589CD1	16	1856589CB1
2617672	3	2617672CD1	17	2617672CB1
2769104	4	2769104CD1	18	2769104CB1
4802789	5	4802789CD1	19	4802789CB1
60116897	6	60116897CD1	20	60116897CB1
1866356	7	1866356CD1	21	1866356CB1
1872095	8	1872095CD1	22	1872095CB1
2278688	9	2278688CD1	23	2278688CB1
4043361	10	4043361CD1	24	4043361CB1
531758	11	531758CD1	25	531758CB1
747204	12	747204CD1	26	747204CB1
747204.1	13	747204.1CD1	27	747204.1CB1
747204.1	14	747204.1CD2	28	747204.1CB2

WO 01/46443

PCT/US00/34811

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	1714345	g6941890	0.0	Ubiquitin-specific protease [Mus musculus] [Valero, R. et al. (1999) Genomics 62:395-405]
2	1856589	g1143194	1.2e-45	Proteinase [Homo sapiens] [Yu, J.X. et al. (1994) J. Biol. Chem. 269:18643-18648]
3	2617672	g432927	3.0e-113	Ubiquitin-interstitial nephritis antigen TTN-Ag [Mus musculus]
4	2769104	g175694	3.2e-28	Human complement C1r [Homo sapiens]
5	4802789	g4454565	4.1e-30	Ubiquitin processing protease [Homo sapiens]
6	6011697	g2885747	0.0	Ubl1 protease [Homo sapiens]
7	1866356CD1	g2098623	1.2e-68	Ubl1 protease family P2 [Caenorhabditis elegans]
8	1872095CD1	g2347400	1.7e-23	Ubiquitin-specific protease [Arabidopsis thaliana]
9	2278686CD1	g1184161	0.0	Aminopeptidase [Mus musculus]
10	3043361CD1	g2643781	2.6e-104	Futurate pyroglutamate-peptidase I [Mus musculus]
11	3317958CD1	g480950	5.7e-16	Carboxy-esterase [Homo sapiens]
12	7257324CD1	g2116650	1.1e-78	Alpha-1-antitrypsin [Carcopithecus aethiops] [Colau, B. et al. (1994) DNA 3:327-330; Yoshida, K. et al. (1999) J. Biochem. Mol. Biol. Biophys. 3:59-62]
13	7472038CD1	g293230	4.0e-106	Aspartic protease [Neotoma mexicanus]
14	7472041CD1	g3088553	1.1e-14	Cystatin-related epididymal spermatogenic protein [Homo sapiens] [Connell, G.H. et al. (1993) Proc Natl Acad Sci USA 90:1165-1169]

WO 01/46443

PCT/US00/34811

Table 3 (cont.)

SFO ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Phosphorylation Sites	Formaldehyde Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	2765104CD1	187	S67 T162 S131 S134 W138	M147	CUB domain (extracellular domain found in complement proteins): G10-Y160 Complement C1r/C1s repeat: C36-V163, Q51-Y160, K24-Y150 Signal peptide: M1-A35	BLAST-DBOM MOTIFS
5	4602789CD1	289	T18 S24 S105 T313 S235 S261 S17 S102 S108 S188 S225 T265 S271	N119 N116	Transmembrane domain: G95-L92 Ubiquitin C-term. hydrolase signature: G191-L208	SPScan BLAST BLAST-DBOM MOTIFS
6	60116897CD1	960	S225 S483 T57 T87 S124 T197 S21 T343 S157 S107 S234 S304 S17 S916 S926 T933 S10 S94 T383 S221 T556 S303 S359 S432 S486 S568 S740 S781 T930 T951 Y312 Y622 Y679 Y885	B95 M103 M119 N219 S294 S405 S431 N950 S714 R073	Zn metalloprotease family M1: L69-G458 Zn membrane atanyl dipeptidase: V357-T520, K386-Y398 Neutral Zn-protease: W64-S500, G529-L837, T527-S899, V849-L956 Neutral Zn-protease, Zn binding region: V357-S376, Y367-F377	SPScan BLAST-DBOM MOTIFS
7	1866356CD1	525	S2 S30 M155 T311 S168 S390 T479 S522 S107 S122 S155 S228		Signal peptide: M1-S28 Signature: co the peptidase family R2 P313-R63 E157-S422	SPScan BLAST-DBOM

WO 01/46443

PCT/US00/34811

Table 3 (cont.)

SPC ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	137205CDL	795	S274 S279 S522 Y223 T683 T851 S274 S314 S332 T337 S377 S379 S383 S282 S370 S272 S285 S337 S620 T631	N171 N381 N443 N448 K336 N647 K370 N436 N711 N712 N720 N758	Ubiquitin carboxyl-terminal hydrolase 1 motif: G199-L213 Ubiquitin carboxyl-terminal hydrolase 2 motif: Y293-E510 Ubiquitin carboxyl-terminal hydrolase family UCH-L1: T198-L229 Ubiquitin carboxyl-terminal hydrolases family UCH-2: K589-K701 Protease, Ubiquitin hydrolase, ubiquitin-specific enzyme, deubiquitinating carboxyl-terminal thioesterase, processing, cofactor: P00742, S470-L541 Ubiquitin carboxyl-terminal hydrolases family 2: DMC0659 P40818 782-1103: L203-E386	MOTIFS MOTIFS PRODBR-PEEM H03BR-PEEM BLAST-PRODBM ELANSI-DOXO

WO 01/46443

PCT/US00/34811

Table 3 (cont.)

STQ ID NO:	Inverte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7257324CDL	414	S93 T94 T223 T258 T16 S216 T129 S182 S235 S300 S316 S396 Y118	N221 N233 N267	Scpains protein signatures BL00184: M7-T94, A173-T193, T200-M241, V106-F332, D387-F411 Scpains signatures: G164-K414 Scpains, serins protease inhibitor, signal, precursor, glycoprotein, plasma, proteinase: P30232, A41-F411 Scpains: D60112 P01009 A7-413: D54-N410 Sigma, peptide: M1-G19 Scpains (serine protease inhibitors): M5-F411	ELJMS - BLOCKS PROFLeSear PRODOM ELAST-DOHO HAMER SFSKAN HAMER-PPAM
13	7472038CDL	397	E127 T166 S217 T381 S137 X340 S18 S1190 T154 S252	M156 M166 M169 M178 M179 M202 M245 M258	pepsin (A1) aspartic protease signature PR00792A: M156-M184 M160-M184 Aspartyl protease, hydrolyase precursor, signal, zymogen, glycoprotein, multigene: P65-S307 Bubalytic and viral aspartyl proteases: DND123 Q03198 19-385: R93-A995 Aspartyl protease: V31-V104, V278-L289 Bubalytic and viral aspartyl protease: P65-S307 Bubalytic and viral aspartic proteases BL00141: F91-S106, D184-S195, G233-G244, V278-L287, I370-A323	ELJMS - PRINTS ELAST-PRODOM ELAST-DOHO NOTIPS HAMER-PPAM ELJMS - BLOCKS

WO 01/46443

PCT/US00/34811

Table 3 (cont.)

SEQ ID NO.	Lecythe Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	747204L001	145	T76 S13 S19 S37 T83 S105	R42 N54 N57 Y94 N98 M131 M132	Cysteine proteases, inhibitors: DG0022 P01035 1-110: S30-G134 Cysteine proteases inhibitor: R66-F89 Signal peptide: M1-G23	BLAST-DH0 SLIMPS- ECCCS RMSR PScan DREN-E2M PROLIScan
					Cysteatin domain: G10-S13 Cysteine proteases inhibitors signature: M53-S100	

WO 01/46443

PCT/US00/34811

Table 4

Polynucleotide SEQ ID No.	Inverse Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
15	1714846CB1	4028	1349-1415, 1-199, 1903-3217	6831476AL (S1MNGR01)	1	689
				6773215AL (OVARDR01)	650	1371
				6425728AL (G0G0R07)	928	1635
				1872084ES (SKINBL01)	1575	1995
				396127HL (BRSTN075)	1964	2110
				6433345EL (LUNGFR07)	1929	2526
				4402573AL (S1MNS72)	532	2868
				62953152L (B21P0R01)	2586	3140
				1748465F (UCRNGR02)	3038	3621
				24010119 (HNF0R01)	3652	4033
				5876414L (H1X30R01)	3385	539
				70152358V1	1	589
				70151001V1	359	824
				70157461V1	686	1218
60106256B2	976	1422				
17	2657672CB1	1911	5-619	548554HL (REFIN01)	1	268
				2170381FS (EMDOR03)	150	575
				1437060E1 (PANC078)	476	1031
				70092221V1	774	1352
				14288458L (S1MNS701)	1170	1408
				33900668L (BONRPT01)	1318	1569
				2994150HL (KIDNPT02)	1422	1715
				3501527HL (DST0R04)	1644	1850
				3228672HL (SPACR07)	1760	2026
				70156781V1	143	326
				70156720V1	312	584
				34948395E1 (ADRNUT01)	1	685
18	2769164CB1	854	1-176	70005193D1	660	1266
				26306357G (COLNUT15)	708	1364
				605612H (BRSTN01)	1138	1385
19	4802785CB1	1385	1-23, 343-503	70005193D1	660	1266
				26306357G (COLNUT15)	708	1364
19	4802785CB1	1385	1-23, 343-503	605612H (BRSTN01)	1138	1385

WO 01/46443

PCT/US00/34811

Table 4 (cont.)

Polynucleotide SEQ ID No:	Incyto Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
20	60116897CE1	3323	2502-2610, 1-735 1122-1879	3154611FG (TUXMNT02)	740	834
				6011691RL	1119	1236
				2832568F8 (TUXMNT03)	1519	1657
				2830940F7 (TUXMNT02)	1877	2078
				6510679HL (LUNGTT01)	2135	2180
				2845923F6 (BNSFTT03)	2358	2641
				3200003F6 (PENCNT02)	2752	2862
				283292225 (BNSFTT03)	3233	3323
				3201617F6 (PENCNT02)	3419	3523
				283292225 (BNSFTT03)	3523	3627
21	1866356	2123	1-1590	283292225 (BNSFTT03)	2004	2053
				5726444HL (BNSFTT03)	244	304
				373992576 (BNSFTT03)	1669	1765
				25859086 (BNSFTT01)	643	1073
				6157887HL (BNSFTT03)	1871	2092
				6269726HL (BNSFTT03)	1046	1705
				4570803HL (SHLAD-T03)	555	249
				257179HL (BNSFTT01)	2234	2895
				144288146 (TUXMNT03)	1368	1619
				1388162HL (CARGDT02)	1545	1809
22	1872095	2893	584-1266, 1-567 2839-2893	1344669HL (PENCNT01)	1714	1962
				58401873VA	1898	2451
				44975686 (PENCNT02)	219	734
				44975686 (PENCNT02)	737	1459
				44975686 (PENCNT02)	1459	2199
				44975686 (PENCNT02)	2199	2875
				44975686 (PENCNT02)	2875	3551
				44975686 (PENCNT02)	3551	4227
				44975686 (PENCNT02)	4227	4903
				44975686 (PENCNT02)	4903	5579

WO 01/46443

PCT/US00/34811

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Inverte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
23	2278688	4170	1-245, 3069-3624, 1149-1809	2254713HI (OVAR1P01)	453	710
				097483RI (PIVUN001)	2611	3220
				271744HI (BEAK023)	1433	1667
				442296HI (BEAD010)	1258	1500
				1378162HI (LORNG010)	1110	1309
				3076825HI (BOUK001)	2137	2191
				1556490HI (LORNG033)	755	1064
				255447HI (SCY5G002)	399	4170
				466217HI (BEST012)	1	271
				353760HI (BEAL012)	241	241
				134323HI (BEAL002)	241	241
				134323HI (BEAL002)	241	241
				42856756 (BRAS010)	72	576
24	4043361	757	1-66	1289595HI (BRAS011)	3530	4149
				23739876 (SCLNG001)	1570	2094
				097483RI (PIVUN001)	2913	3639
				2598679HI (UPRNG02)	1817	2134
				2110561HI (BEAT003)	669	950
				3011419HI (MUSNG07)	2369	2622
				1394210HI (UPRNG03)	1002	1396
				4880281HI (UPRNG01)	524	767
				40433616 (LORNG035)	1	593
				6772887I (OVAR001)	436	4216
				6121924HI (BEAL005)	1022	1538
				708224HI (BRAS002)	426	1382
				26	7257324	1497
322863156 (SCLNG02)	416	474				
7257324HI (SCLNG01)	1	474				
27	7472038	1194	1-29, 788-1194	GMV-96436155_002.seq	1	1194
				GMV-96436155_003.seq	1	438
28	7472041	438	1-27	GMV-96436155_004.seq	1	438
				GMV-96436155_005.seq	1	438

WO 01/46443

PCT/US00/34811

Table 5

Polynucleotide SEQ ID NO:	Protein Accession ID	Representative Library
15	1714846CBI	LIN280R07
16	5856595CBI	BROSN0T18
17	2617572CBI	PANCM0T08
18	2769104CBI	COLAN0T02
19	4802785CBI	ADREPU07
20	60116897CBI	TYRAN0T03
21	5866356CBI	HHTZAF01
22	1872091CBI	THYAN0T03
23	2878885CBI	LALAT0T02
24	4043351CBI	LUNSN0T02
25	3938795CBI	KIDAN0T03
26	7231521CBI	SKLBN0T04

WO 01/46443

PCT/US00/34811

Table 6

Library	Vector	Library Description
LUNG0007	P1N1CY	This normalized lung tissue library was constructed from RNA isolated from a lung tissue library. This library was normalized in two rounds using conditions adapted from Soares et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232 and Bonaldi et al. (1996) Genom. Res. 6:791, except that a significantly longer (48 hours/round) denaturing hybridization was used.
PROSTAT8	P1N1CY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrectomy. Pathology indicated adenocarcinoma hyperplasia. This tissue was associated with a grade 3 transitional cell carcinoma. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
PANCRE008	P1N1CY	Library was constructed using RNA isolated from pancreatic tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology for the associated tumor tissue indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract. Previous surgeries included a total splenectomy, cholecystectomy, and abdominal hysterectomy. Family history included cardiovascular disease, type II diabetes, and stomach cancer.
COLON002	P1N1CY	Library was constructed using RNA isolated from diseased ascending colon tissue removed from a 25-year-old Caucasian female during a multiple segmental resection of the large bowel. Pathology indicated moderately to severely active chronic ulcerative colitis, involving the entire colon specimen and sparing 2 cm of the attached ileum. Grossly, the specimen showed continuous involvement from the rectum proximally; marked mucosal atrophy and no skip areas were identified. Microscopically, the specimen showed dense, predominantly mucosal inflammation and crypt abscesses. Patient history included benign large bowel neoplasm. Previous surgeries included a polypectomy.
ADREN007	P1N1CY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 43-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated pheochromocytoma.

WO 01/46443

PCT/US00/34811

Table 6 (cont.)

Library	Vector	Library Description
TYMOT03	PINCY	Library was constructed using RNA isolated from nonactivated T41 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-12 and 2i-transfected CO3 cells.
HWTZEA01	PINDESCRIP1	Library was constructed at Stratagene (SP437231) using RNA isolated from the HWTZ cell line (derived from a human testocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with retinoic acid for 24 hours.
LAPROT02	PINCY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during amniopasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and Type II diabetes.
LUNENOT05	PINCY	Library was constructed using RNA isolated from lung tissue removed from a 62-year-old Caucasian female. Pathology for the associated tumor tissue indicated a grade I spindle cell carcinoma forming a nodule. Patient history included depression, thrombophlebitis, and hyperlipidemia. Family history included cerebrovascular disease, atherosclerotic coronary artery disease, breast cancer, colon cancer, type II diabetes, and malignant skin melanoma.
THYROT03	PINCY	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.
KIDNMO05	REPORT1	Library was constructed using RNA isolated from the kidney tissue of a 2-day-old Hispanic female, who died from cerebral anoxia. Family history included congenital heart disease.
SKINMO04	PINCY	Library was constructed using RNA isolated from breast skin tissue removed from a 70-year-old Caucasian female during a breast biopsy and resection.

WO 01/46443

PCT/US00/34811

Table 7

Program	Description	Reference	Parameter Threshold
ABIFACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABIPARACEL FDP	A Fast Data Finder useful in comparing and elucidating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI Autoassembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastn, blastx, blastp, and blastc.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	ESYs: Probability value= 1.0E-8 or less Fdd Length sequence: Probability value= 1.0E-10 or less
PASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. PASTA comprises at least five functions: fasta, ffast, ffastx, ffastc, and ffastd.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	ESYs: facta E value=1.0E-5 Assembled ESYs: fasta identity= 95% or greater and Match length=200 bases or greater; fastx E value= 1.0E-8 or less Fdd Length sequence: fastx score=100 or greater
BLIMPS	A BLASTs Improved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) <i>Nucleic Acids Res.</i> 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) <i>Methods Enzymol.</i> 256:88-105; and Atwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-5 or less
FINDMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:3501-3531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 16:3209-3222; Durbin, R. et al. (1998) <i>Our World View in a Nucleic Acid</i> , Cambridge Univ. Press, pp. 1-590.	PFAM hit: Probability value= 1.0E-5 or less Signal peptide hit: Score= 0 or greater

WO 01/46443

PCT/US00/34811

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Grisbow, M. et al. (1988) <i>CABIOS</i> 4:61-66; Grisbow, M. et al. (1987) <i>Nucleic Acids Res.</i> 15:146-159; Baroch, A. et al. (1997) <i>Nucleic Acids Res.</i> 25:217-221.	Normalized quality scores, GCC-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Pired	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) <i>Genome Res.</i> 8:175-185; Ewing, B. and P. Green (1998) <i>Genome Res.</i> 8:186-194.	
Phrap	A Plus Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489; Smith, T.F. and M.S. Waterman (1981) <i>J. Mol. Biol.</i> 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater, Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) <i>Genome Res.</i> 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) <i>Protein Engineering</i> 10:1-6; Clavette, J.M. and S. Audic (1997) <i>CABIOS</i> 12:431-439.	Scores=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Pearson, B. and P. Argos (1994) <i>J. Mol. Biol.</i> 237:182-192; Pearson, B. and P. Argos (1996) <i>Protein Sci.</i> 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonhammer, E.L. et al. (1998) <i>Proc. Sixth Int. Conf. on Intelligent Systems for Mol. Biol.</i> , Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Baroch, A. et al. (1997) <i>Nucleic Acids Res.</i> 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

WO 01/46443

PCT/US00/34811

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14,
 - b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14,
 - c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and
 - 10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-14.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:15-28.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - 35 b) recovering the polypeptide so expressed.

WO 01/46443

PCT/US00/34811

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
- 5 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28,
b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28,
c) a polynucleotide sequence complementary to a),
d) a polynucleotide sequence complementary to b), and
10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
- 15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization
20 complex is formed between said probe and said target polynucleotide or fragments thereof, and
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
- 25 15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.
- 35

WO 01/46443

PCT/US00/34811

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.
18. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition of claim 16.
19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
- 10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.
20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.
- 15 21. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 20.
- 20 22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.
- 25 23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.
24. A method for treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 23.
- 30 25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:
- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
35 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

WO 01/46443

PCT/US00/34811

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- 5 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in
10 the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method
15 comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
b) detecting altered expression of the target polynucleotide, and
c) comparing the expression of the target polynucleotide in the presence of varying amounts of
20 the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at
25 least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
c) quantifying the amount of hybridization complex; and
30 d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

WO 01/46443

PCT/US00/34811

```

<110> INCYTE GENOMICS, INC.
      YANG, Junming
      RAUGHN, Mariab R.
      BURFORD, Neil
      AU-YOUNG, Janice
      LU, Myung Aina M.
      REDDY, Roopa
      YUE, Henry
      NGUYEN, Dammiel B.
      TANG, Y. Tom
      YAO, Monique G.
      TAL, Preeti

<130> PROTEASES

<130> PI-0003 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/172,055; 60/177,334; 60/178,884; 60/179,903
<151> 1999-12-23; 2000-01-21; 2000-01-28; 2000-02-02

<160> 28
<170> PERL Program

<210> 1
<211> 1055
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1714846CD1

<400> 1
Met Thr Val Glu Gln Asn Val Leu Gln Gln Ser Ala Ala Gln Lys
 1      5      10      15
His Gln Gln Thr Phe Leu Asn Gln Leu Arg Glu Ile Thr Gly Ile
 20     25     30
Asn Asp Thr Gln Ile Leu Gln Gln Ala Leu Lys Asp Ser Asn Gly
 35     40     45
Asn Leu Glu Leu Ala Val Ala Phe Leu Thr Ala Lys Asn Ala Lys
 50     55     60
Thr Pro Gln Gln Glu Thr Thr Tyr Tyr Gln Thr Ala Leu Pro
 65     70     75
Gly Asn Asp Arg Tyr Ile Ser Val Gly Ser Gln Ala Asp Thr Asn
 80     85     90
Val Ile Asp Leu Thr Gly Asp Asp Lys Asp Asp Leu Gln Arg Ala
 95    100    105
Ile Ala Leu Ser Leu Ala Glu Ser Asn Arg Ala Phe Arg Glu Thr
110    115    120
Gly Ile Thr Asp Glu Glu Gln Ala Ile Ser Arg Val Leu Glu Ala
125    130    135
Ser Ile Ala Glu Asn Lys Ala Cys Leu Lys Arg Thr Pro Thr Glu
140    145    150
Val Trp Arg Asp Ser Arg Asn Pro Tyr Asp Arg Lys Arg Gln Asp
155    160    165
Lys Ala Pro Val Gly Leu Lys Asn Val Gly Asn Thr Cys Trp Phe
170    175    180
Ser Ala Val Ile Gln Ser Leu Phe Asn Leu Leu Glu Phe Arg Arg
185    190    195
Leu Val Leu Asn Tyr Lys Pro Pro Ser Asn Ala Gln Asp Leu Pro
200    205    210
Arg Asn Gln Lys Glu His Arg Asn Leu Pro Phe Met Arg Glu Leu
215    220    225
Arg Tyr Leu Phe Ala Leu Leu Val Gly Trp Lys Arg Lys Tyr Val

```

WO 01/46443

PCT/US00/34811

230	235	240
Asp Pro Ser Arg Ala Val Glu Ile Leu Lys Asp Ala Phe Lys Ser		
245	250	255
Asn Asp Ser Gln Gln Asp Val Ser Glu Phe Thr His Lys Leu		
260	265	270
Leu Asp Trp Leu Glu Asp Ala Phe Gln Met Lys Ala Glu Glu Glu		
275	280	285
Thr Asp Glu Glu Lys Pro Lys Asn Pro Met Val Glu Leu Phe Tyr		
290	295	300
Gly Arg Phe Leu Ala Val Gly Val Leu Glu Gly Lys Lys Phe Glu		
305	310	315
Asn Thr Gln Met Phe Gly Gln Tyr Pro Leu Gln Val Asn Gly Phe		
320	325	330
Lys Asp Leu His Glu Cys Leu Glu Ala Ala Met Ile Glu Gly Met		
335	340	345
Ile Glu Ser Leu His Ser Glu Asn Ser Gly Lys Ser Gly Gln Glu		
350	355	360
His Trp Phe Thr Glu Leu Pro Pro Val Leu Thr Phe Glu Leu Ser		
365	370	375
Arg Phe Glu Phe Asn Gln Ala Leu Gly Arg Pro Glu Lys Ile His		
380	385	390
Asn Lys Leu Glu Phe Pro Gln Val Leu Tyr Leu Asp Arg Tyr Met		
395	400	405
His Arg Asn Arg Glu Ile Thr Arg Ile Lys Arg Glu Glu Ile Lys		
410	415	420
Arg Leu Lys Asp Tyr Leu Thr Val Leu Gln Gln Arg Leu Glu Arg		
425	430	435
Tyr Leu Ser Tyr Gly Ser Gly Pro Lys Arg Phe Pro Leu Val Asp		
440	445	450
Val Leu Gln Tyr Ala Leu Glu Phe Ala Ser Ser Lys Pro Val Cys		
455	460	465
Thr Ser Pro Val Asp Asp Ile Asp Ala Ser Ser Pro Pro Ser Gly		
470	475	480
Ser Ile Pro Ser Gln Thr Leu Pro Ser Thr Thr Glu Gln Gln Gly		
485	490	495
Ala Leu Ser Ser Glu Leu Pro Ser Thr Ser Pro Ser Ser Val Ala		
500	505	510
Ala Ile Ser Ser Arg Ser Val Ile His Lys Pro Phe Thr Gln Ser		
515	520	525
Arg Ile Pro Pro Asp Leu Pro Met His Pro Ala Pro Arg His Ile		
530	535	540
Thr Glu Glu Glu Leu Ser Val Leu Glu Ser Cys Leu His Arg Trp		
545	550	555
Arg Thr Glu Ile Glu Asn Asp Thr Arg Asp Leu Gln Glu Ser Ile		
560	565	570
Ser Arg Ile His Arg Thr Ile Glu Leu Met Tyr Ser Asp Lys Ser		
575	580	585
Met Ile Gln Val Pro Tyr Arg Leu His Ala Val Leu Val His Glu		
590	595	600
Gly Gln Ala Asn Ala Gly His Tyr Trp Ala Tyr Ile Phe Asp His		
605	610	615
Arg Glu Ser Arg Trp Met Lys Tyr Asn Asp Ile Ala Val Thr Lys		
620	625	630
Ser Ser Trp Glu Glu Leu Val Arg Asp Ser Phe Gly Gly Tyr Arg		
635	640	645
Asn Ala Ser Ala Tyr Cys Leu Met Tyr Ile Asn Asp Lys Ala Gln		
650	655	660
Phe Leu Ile Gln Glu Glu Phe Asn Lys Glu Thr Gly Gln Pro Leu		
665	670	675
Val Gly Ile Glu Thr Leu Pro Pro Asp Leu Arg Asp Phe Val Glu		
680	685	690
Glu Asp Asn Gln Arg Phe Glu Lys Glu Leu Glu Glu Trp Asp Ala		
695	700	705
Gln Leu Ala Gln Lys Ala Leu Gln Glu Lys Leu Leu Ala Ser Gln		
710	715	720
Lys Leu Arg Glu Ser Glu Thr Ser Val Thr Thr Ala Gln Ala Ala		
725	730	735

WO 01/46443

PCT/US00/34811

Gly Asp Pro Glu Tyr Leu Glu Gln Pro Ser Arg Ser Asp Phe Ser
740 745 750
Lys His Leu Lys Glu Glu Thr Ile Gln Ile Ile Thr Lys Ala Ser
755 760 765
His Glu His Glu Asp Lys Ser Pro Glu Thr Val Leu Gln Ser Ala
770 775 780
Ile Lys Leu Glu Tyr Ala Arg Leu Val Lys Leu Ala Gln Glu Asp
785 790 795
Thr Pro Pro Glu Thr Asp Tyr Arg Leu His His Val Val Val Tyr
800 805 810
Phe Ile Cln Asn Gln Ala Pro Lys Lys Ile Ile Glu Lys Thr Leu
815 820 825
Leu Glu Gln Phe Gly Asp Arg Asn Leu Ser Phe Asp Glu Arg Cys
830 835 840
His Asn Ile Met Lys Val Ala Gln Ala Lys Leu Glu Met Ile Lys
845 850 855
Pro Glu Glu Val Asn Leu Glu Glu Tyr Glu Glu Trp His Gln Asp
860 865 870
Tyr Arg Lys Phe Arg Glu Thr Thr Met Tyr Leu Ile Ile Gly Leu
875 880 885
Glu Asn Phe Gln Arg Glu Ser Tyr Ile Asp Ser Leu Leu Phe Leu
890 895 900
Ile Cys Ala Tyr Gln Asn Asn Lys Glu Leu Leu Ser Lys Gly Leu
905 910 915
Tyr Arg Gly His Asp Glu Glu Leu Ile Ser His Tyr Arg Arg Glu
920 925 930
Cys Leu Leu Lys Leu Asp Glu Gln Ala Ala Glu Leu Phe Glu Ser
935 940 945
Gly Glu Asp Arg Glu Val Asn Asn Gly Leu Ile Ile Met Asn Glu
950 955 960
Phe Ile Val Pro Phe Leu Pro Leu Leu Leu Val Asp Glu Met Glu
965 970 975
Glu Lys Asp Ile Leu Ala Val Glu Asp Met Arg Asn Arg Trp Cys
980 985 990
Ser Tyr Leu Gly Gln Glu Met Glu Pro His Leu Gln Glu Lys Leu
995 1000 1005
Thr Asp Phe Leu Pro Lys Leu Leu Asp Cys Ser Met Glu Ile Lys
1010 1015 1020
Ser Phe His Glu Pro Pro Lys Leu Pro Ser Tyr Ser Thr His Glu
1025 1030 1035
Leu Cys Glu Arg Phe Ala Arg Ile Met Leu Ser Leu Ser Arg Thr
1040 1045 1050
Pro Ala Asp Gly Arg
1055

<210> 2
<211> 358
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1856589CD1

<400> 2
Met Gly Ala Ala Thr Cys Arg Gly Ser Arg Ile Pro Ser Gly Pro
1 5 10 15
Pro Val Cln Gly Glu Arg Ser Ala Pro Arg Phe Gly Val Thr Ser
20 25 30
Leu Ser Leu Trp Pro Ala Asp Phe Lys Asp Asn Trp Arg Ile Ala
35 40 45
Gly Ser Arg Gln Glu Val Ala Leu Ala Gly Glu Pro Ala Asp Gln
50 55 60
Gln Cln Thr His Leu Arg Arg Leu Pro Tyr Arg Gln Thr Leu Gly
65 70 75
Tyr Lys Glu Asp Thr Thr Asn Pro Val Cys Gly Glu Pro Trp Trp
80 85 90

WO 01/6443

PCT/US00/34811

Ser Glu Asp Leu Glu Leu Thr Arg His Trp Pro Trp Glu Val Ser
95 100 105
Leu Arg Met Glu Asn Glu His Val Cys Gly Gly Ala Leu Ile Asp
110 115 120
Pro Ser Trp Val Val Thr Ala Ala His Cys Ser Gln Gly Thr Lys
125 130 135
Glu Tyr Ser Val Val Leu Gly Thr Ser Lys Leu Gln Pro Met Asn
140 145 150
Phe Ser Arg Ala Leu Trp Val Pro Val Arg Asp Ile Ile Met His
155 160 165
Pro Lys Tyr Trp Gly Arg Ala Phe Ile Met Gly Asp Val Ala Leu
170 175 180
Val His Leu Gln Thr Pro Val Thr Phe Ser Glu Tyr Val Gln Pro
185 190 195
Ile Cys Leu Pro Glu Pro Asn Phe Asn Leu Lys Val Gly Thr Gln
200 205 210
Cys Trp Val Thr Gly Trp Ser Gln Val Lys Gln Arg Phe Ser Gly
215 220 225
Ser Thr Ala Asn Ser Met Leu Thr Pro Glu Leu Gln Glu Ala Glu
230 235 240
Val Phe Ile Met Asp Asn Lys Arg Cys Asp Arg His Tyr Lys Lys
245 250 255
Ser Phe Phe Pro Leu Val Val Pro Leu Val Leu Gly Asp Met Ile
260 265 270
Cys Ala Thr Asn Tyr Gly Glu Asn Leu Cys Tyr Gly Asp Ser Gly
275 280 285
Gly Pro Leu Ala Cys Glu Val Glu Gly Arg Trp Ile Leu Ala Gly
290 295 300
Val Leu Ser Trp Glu Lys Ala Cys Val Lys Ala Gln Asn Pro Gly
305 310 315
Val Tyr Thr Arg Val Thr Lys Tyr Thr Lys Trp Ile Lys Lys Gln
320 325 330
Met Ser Asn Gly Ala Phe Ser Gly Pro Cys Ala Ser Ala Cys Leu
335 340 345
Leu Phe Leu Cys Trp Pro Leu Gln Pro Gln Met Gly Ser
350 355

<210> 3

<211> 467

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<222> Incyte ID No: 2617672CD1

<400> 3

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

WO 01/66443

PCT/US00/34811

Asn Gln Gly Asn Tyr Gly Trp Gln Ala Gly Asn His Ser Ala Phe 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

<210> 4
 <211> 197
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2769104CD1

<400> 4
 Met Pro Gly Pro Arg Val Trp Gly Lys Tyr Leu Trp Arg Ser Pro 15
 1
 His Ser Lys Gly Cys Pro Gly Ala Met Trp Trp Leu Leu Leu Trp 20
 20
 Gly Val Leu Gln Ala Cys Pro Thr Arg Gly Ser Val Leu Leu Ala 35
 35
 Gln Glu Leu Pro Gln Gln Leu Thr Ser Pro Gly Tyr Pro Glu Pro 50
 50
 Tyr Gly Lys Gly Gln Glu Ser Ser Thr Asp Ile Lys Ala Pro Glu 65
 65
 Gly Phe Ala Val Arg Leu Val Phe Gln Asp Phe Asp Leu Glu Pro 80
 80

WO 01/66443

PCT/US00/34811

```

Ser Gln Asp Cys Ala Gly Asp Ser Val Thr Ile Ser Phe Val Gly
95 100 105
Ser Asp Pro Ser Gln Phe Cys Gly Gln Gln Gly Ser Pro Leu Gly
110 115 120
Arg Pro Pro Gly Gln Arg Glu Phe Val Ser Ser Gly Arg Ser Leu
125 130 135
Arg Leu Thr Phe Arg Thr Gln Pro Ser Ser Glu Asn Lys Thr Ala
140 145 150 155
His Leu His Lys Gly Phe Leu Ala Leu Tyr Gln Thr Val Gly Glu
155 160 165
Cys Pro Ser Trp Gly Cys Arg Glu Gly Ala Ser Val Pro Ser His
170 175 180
Asp Pro Gly Ile Phe Lys Pro
185

```

```

<210> 5
<211> 289
<212> FRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyta ID No: 4802789CD1

```

```

<400> 5
Met Arg Val Lys Asp Pro Thr Lys Ala Leu Pro Glu Lys Ala Lys
1 5 10 15
Arg Ser Lys Arg Pro Thr Val Pro His Asp Glu Asp Ser Ser Asp
20 25 30
Asp Ile Ala Val Gly Leu Thr Cys Gln His Val Ser His Ala Ile
35 40 45
Ser Val Asn His Val Lys Arg Ala Ile Ala Glu Asn Leu Trp Ser
50 55 60
Val Cys Ser Glu Cys Leu Lys Glu Arg Arg Phe Tyr Asp Gly Gln
65 70 75
Leu Val Leu Thr Ser Asp Ile Trp Leu Cys Leu Lys Cys Gly Phe
80 85 90
Gln Gly Cys Gly Lys Asn Ser Glu Ser Gln His Ser Leu Lys His
95 100 105
Phe Lys Ser Ser Arg Thr Glu Pro His Cys Ile Ile Ile Asn Leu
110 115 120
Ser Thr Trp Ile Ile Trp Cys Tyr Glu Cys Asp Glu Lys Leu Ser
125 130 135
Thr His Cys Asn Lys Lys Val Leu Ala Gln Ile Val Asp Phe Leu
140 145 150
Gln Lys His Ala Ser Lys Thr Gln Thr Ser Ala Phe Ser Arg Ile
155 160 165
Met Lys Leu Cys Glu Glu Lys Cys Glu Thr Asp Glu Ile Gln Lys
170 175 180
Gly Gly Lys Cys Arg Asn Leu Ser Val Arg Gly Ile Thr Asn Leu
185 190 195
Gly Asn Thr Cys Phe Phe Asn Ala Val Met Gln Asn Leu Ala Gln
200 205 210
Thr Tyr Thr Leu Thr Asp Leu Met Asn Gln Ile Lys Glu Ser Ser
215 220 225
Thr Lys Leu Lys Ile Phe Pro Ser Ser Asp Ser Gln Leu Asp Pro
230 235 240
Leu Val Val Glu Leu Ser Arg Pro Gly Pro Leu Thr Ser Ala Leu
245 250 255
Phe Leu Phe Leu His Ser Met Lys Glu Thr Glu Lys Gly Pro Leu
260 265 270
Ser Pro Lys Val Leu Phe Asn Gln Leu Cys Gln Lys Trp Val His
275 280 285
Leu His Leu Ile

```

<210> 6

WO 01/46443

PCT/US00/34811

<211> 960
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 66116897CD1

```

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

```

WO 01/46443

PCT/US00/34811

Ile Ser Lys Pro Ala Glu Thr Pro Thr Gln Ile Gln Glu Met Phe
 440
 Asp Glu Val Ser Tyr Asn Lys Gly Ala Cys Ile Leu Asn Met Leu
 455
 Lys Asp Phe Leu Gly Glu Glu Lys Phe Gln Lys Gly Ile Ile Gln
 470
 Tyr Leu Lys Lys Phe Ser Tyr Arg Asn Ala Lys Asn Asp Asp Leu
 485
 Trp Ser Ser Leu Ser Asn Ser Cys Leu Glu Ser Asp Phe Thr Ser
 500
 Gly Gly Val Cys His Ser Asp Pro Lys Met Thr Ser Asn Met Leu
 515
 Ala Phe Leu Gly Glu Asn Ala Glu Val Lys Glu Met Met Thr Thr
 530
 Trp Thr Leu Gln Lys Gly Ile Pro Leu Leu Val Val Lys Gln Asp
 545
 Gly Cys Ser Leu Arg Leu Gln Gln Glu Arg Phe Leu Gln Gly Val
 560
 Phe Gln Glu Asp Pro Glu Trp Arg Ala Leu Gln Glu Arg Tyr Leu
 575
 Trp His Ile Pro Leu Thr Tyr Ser Thr Ser Ser Ser Asn Val Ile
 590
 His Arg His Ile Leu Lys Ser Lys Thr Asp Thr Leu Asp Leu Pro
 605
 Glu Lys Thr Ser Trp Val Lys Phe Asn Val Asp Ser Asn Gly Tyr
 620
 Tyr Ile Val His Tyr Glu Gly His Gly Trp Asp Gln Leu Ile Thr
 635
 Gln Leu Asn Gln Asn His Thr Leu Leu Arg Pro Lys Asp Arg Val
 650
 Gly Leu Ile His Asp Val Phe Gln Leu Val Gly Ala Gly Arg Leu
 665
 Thr Leu Asp Lys Ala Leu Asp Met Thr Tyr Tyr Leu Gln His Glu
 680
 Thr Ser Ser Pro Ala Leu Leu Glu Gly Leu Ser Tyr Leu Glu Ser
 695
 Phe Tyr His Met Met Asp Arg Arg Asn Ile Ser Asp Ile Ser Gln
 710
 Asn Leu Lys Arg Tyr Leu Leu Gln Tyr Phe Lys Pro Val Ile Asp
 725
 Arg Gln Ser Trp Ser Asp Lys Gly Ser Val Trp Asp Arg Met Leu
 740
 Arg Ser Ala Leu Leu Lys Leu Ala Cys Asp Leu Asn His Ala Pro
 755
 Cys Ile Gln Lys Ala Ala Glu Leu Phe Ser Gln Trp Met Glu Ser
 770
 Ser Gly Lys Leu Asn Ile Pro Thr Asp Val Leu Lys Ile Val Tyr
 785
 Ser Val Gly Ala Gln Thr Thr Ala Gly Trp Asn Tyr Leu Leu Glu
 800
 Gln Tyr Glu Leu Ser Met Ser Ser Ala Glu Gln Asn Lys Ile Leu
 815
 Tyr Ala Leu Ser Thr Ser Lys His Gln Glu Lys Leu Leu Lys Leu
 830
 Ile Glu Leu Gly Met Glu Gly Lys Val Ile Lys Thr Gln Asn Leu
 845
 Ala Ala Leu Leu His Ala Ile Ala Arg Arg Pro Lys Gly Gln Gln
 860
 Leu Ala Trp Asp Phe Val Arg Glu Asn Trp Thr His Leu Leu Lys
 875
 Lys Phe Asp Leu Gly Ser Tyr Asp Ile Arg Met Ile Ile Ser Gly
 890
 Thr Thr Ala His Phe Ser Ser Lys Asp Lys Leu Gln Glu Val Lys
 905
 Leu Phe Phe Glu Ser Leu Glu Ala Gln Gly Ser His Leu Asp Ile
 920
 Phe Gln Thr Val Leu Glu Thr Ile Thr Lys Asn Ile Lys Trp Leu
 935

WO 01/46443

PCT/US00/34811

Glu	Lys	Asn	Leu	935	Pro	Thr	Leu	Arg	Thr	940	Trp	Leu	Met	Val	Asn	Thr	945
				950						955							960

<210> 7
 <211> 525
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1066356CD1

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

WO 01/46443

PCT/US00/34811

Asp Gly Gln Glu Asp Gly Ser Val Ile Glu Arg Gln Val Gly Thr
 395 400 405
 Phe Glu Arg His Thr Lys Gly Ile Gly Arg Lys Val Met Glu Arg
 410 415 420
 Gln Gly Trp Ala Glu Gly Gln Gly Leu Gly Cys Arg Cys Ser Gly
 425 430 435
 Val Pro Glu Ala Leu Asp Ser Asp Gly Gln His Pro Arg Cys Lys
 440 445 450
 Arg Gly Leu Gly Tyr His Gly Glu Lys Leu Gln Pro Phe Gly Gln
 455 460 465
 Leu Lys Arg Pro Arg Arg Asn Gly Leu Gly Leu Ile Ser Thr Ile
 470 475 480
 Tyr Asp Glu Pro Leu Pro Gln Asp Gln Thr Glu Ser Leu Leu Arg
 485 490 495
 Arg Gln Pro Pro Thr Ser Met Lys Phe Arg Thr Asp Met Ala Phe
 500 505 510
 Val Arg Gly Ser Ser Cys Ala Ser Asp Ser Pro Ser Leu Pro Asp
 515 520 525

<210> 8
 <211> 795
 <212> FRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1872095CD1

<400> 8
 Met Ile Thr Val Leu Ile Arg Ser Leu Thr Thr Asp Pro Asn Val
 1 5 10 15
 Lys Asp Ala Ser Met Thr Gln Ala Leu Cys Arg Met Ile Asp Trp
 20 25 30
 Leu Ser Trp Pro Leu Ala Gln His Val Asp Thr Trp Val Ile Ala
 35 40 45
 Leu Leu Lys Gly Leu Ala Ala Val Gln Lys Phe Thr Ile Leu Ile
 50 55 60
 Asp Val Thr Leu Leu Lys Ile Glu Leu Val Phe Asn Arg Leu Trp
 65 70 75
 Phe Pro Leu Val Arg Pro Gly Ala Leu Ala Val Leu Ser His Met
 80 85 90
 Leu Leu Ser Phe Gln His Ser Pro Glu Ala Phe His Leu Ile Val
 95 100 105
 Pro His Val Val Asn Leu Val His Ser Phe Lys Asn Asp Gly Leu
 110 115 120
 Pro Ser Ser Thr Ala Phe Leu Val Gln Leu Thr Glu Leu Ile His
 125 130 135
 Cys Met Met Tyr His Tyr Ser Gly Phe Pro Asp Leu Tyr Glu Pro
 140 145 150
 Ile Leu Glu Ala Ile Lys Asp Phe Pro Lys Pro Ser Glu Glu Lys
 155 160 165
 Ile Lys Leu Ile Leu Asn Gln Ser Ala Tyr Thr Ser Gln Ser Asn
 170 175 180
 Ser Leu Ala Ser Cys Leu Ser Arg Leu Ser Gly Lys Ser Glu Thr
 185 190 195
 Gly Lys Thr Gly Leu Ile Asn Leu Gly Asn Thr Cys Tyr Met Asn
 200 205 210
 Ser Val Ile Gln Ala Leu Phe Met Ala Thr Asp Phe Arg Arg Gln
 215 220 225
 Val Leu Ser Leu Asn Leu Asn Gly Cys Asn Ser Leu Met Lys Lys
 230 235 240
 Leu Gln His Leu Phe Ala Phe Leu Ala His Thr Gln Arg Glu Ala
 245 250 255
 Tyr Ala Pro Arg Ile Phe Phe Glu Ala Ser Arg Pro Pro Trp Phe
 260 265 270
 Thr Pro Arg Ser Gln Gln Asp Cys Ser Glu Tyr Leu Arg Phe Leu

WO 01/46443

PCT/US00/34811

275 280 285
 Leu Asp Arg Leu His Glu Glu Glu Lys Ile Leu Iys Val Gln Ala
 290 295 300
 Ser His Lys Pro Ser Glu Ile Leu Glu Cys Ser Glu Thr Ser Leu
 305 310 315
 Gln Glu Val Ala Ser Lys Ala Ala Val Leu Thr Glu Thr Pro Arg
 320 325 330
 Thr Ser Asp Gly Glu Lys Thr Leu Ile Glu Iys Met Phe Gly Gly
 335 340 345
 Iys Leu Arg Thr His Ile Arg Cys Leu Asn Cys Arg Ser Thr Ser
 350 355 360
 Gln Lys Val Glu Ala Phe Thr Asp Leu Ser Leu Ala Phe Cys Pro
 365 370 375
 Ser Ser Ser Leu Glu Asn Met Ser Val Gln Asp Pro Ala Ser Ser
 380 385 390
 Pro Ser Ile Gln Asp Gly Gly Leu Met Gln Ala Ser Val Pro Gly
 395 400 405
 Pro Ser Glu Glu Pro Val Val Tyr Asn Pro Thr Thr Ala Ala Phe
 410 415 420
 Ile Cys Asp Ser Leu Val Asn Glu Lys Thr Ile Gly Ser Pro Pro
 425 430 435
 Asn Glu Phe Tyr Cys Ser Glu Asn Thr Ser Val Pro Asn Glu Ser
 440 445 450
 Asn Lys Ile Leu Val Asn Lys Asp Val Pro Gln Lys Pro Gly Gly
 455 460 465
 Glu Thr Thr Pro Ser Val Thr Asp Leu Leu Asn Tyr Phe Leu Ala
 470 475 480
 Pro Glu Ile Leu Thr Gly Asp Asn Gln Tyr Tyr Cys Glu Asn Cys
 485 490 495
 Ala Ser Leu Gln Asn Ala Glu Lys Thr Met Gln Ile Thr Glu Glu
 500 505 510
 Pro Glu Tyr Leu Ile Leu Thr Leu Leu Arg Phe Ser Tyr Asp Gln
 515 520 525
 Lys Tyr His Val Arg Arg Lys Ile Leu Asp Asn Val Ser Leu Pro
 530 535 540
 Leu Val Leu Glu Leu Pro Val Lys Arg Ile Thr Ser Phe Ser Ser
 545 550 555
 Leu Ser Glu Ser Trp Ser Val Asp Val Asp Phe Thr Asp Leu Ser
 560 565 570
 Glu Asn Leu Ala Lys Lys Leu Lys Pro Ser Gly Thr Asp Glu Ala
 575 580 585
 Ser Cys Thr Lys Leu Val Pro Tyr Leu Leu Ser Ser Val Val Val
 590 595 600
 His Ser Gly Ile Ser Ser Glu Ser Gly His Tyr Tyr Ser Tyr Ala
 605 610 615
 Arg Asn Ile Thr Ser Thr Asp Ser Ser Tyr Gln Met Tyr His Gln
 620 625 630
 Ser Glu Ala Leu Ala Leu Ala Ser Ser Gln Ser His Leu Leu Gly
 635 640 645
 Arg Asp Ser Pro Ser Ala Val Phe Glu Gln Asp Leu Glu Asn Lys
 650 655 660
 Glu Met Ser Lys Glu Trp Phe Leu Phe Asn Asp Ser Arg Val Thr
 665 670 675
 Phe Thr Ser Phe Gln Ser Val Gln Lys Ile Thr Ser Arg Phe Pro
 680 685 690
 Lys Asp Thr Ala Tyr Val Leu Leu Tyr Iys Lys Gln His Ser Thr
 695 700 705
 Asn Gly Leu Ser Gly Asn Asn Pro Thr Ser Gly Leu Trp Ile Asn
 710 715 720
 Gly Asp Pro Pro Leu Gln Lys Glu Leu Met Asp Ala Ile Thr Lys
 725 730 735
 Asp Asn Lys Leu Tyr Leu Gln Glu Gln Glu Leu Asn Ala Arg Ala
 740 745 750
 Arg Ala Leu Gln Ala Ala Ser Ala Ser Cys Ser Phe Arg Pro Asn
 755 760 765
 Gly Phe Asp Asp Asn Asp Pro Pro Gly Ser Cys Gly Pro Thr Gly
 770 775 780

WO 01/46443

PCT/US00/34811

Gly Gly Gly Gly Gly Gly Phe Asn Thr Val Gly Arg Ile Val Phe
785 790 795

<210> 9
<211> 919
<212> PRQ
<213> Homo sapiens

<220>
<221> misc_feature
<223> Inocyte ID No: 2279688CD1

<400> 9
Met Trp Leu Ala Ala Ala Ala Pro Ser Leu Ala Arg Arg Leu Leu
1 5 10 15
Phe Leu Gly Pro Pro Pro Pro Leu Leu Leu Leu Val Phe Ser
20 25 30
Arg Ser Ser Arg Arg Arg Leu His Ser Leu Gly Leu Ala Ala Met
35 40 45
Pro Glu Lys Arg Pro Phe Glu Arg Leu Pro Ala Asp Val Ser Pro
50 55 60
Ile Asn Cys Ser Leu Cys Leu Lys Pro Asp Leu Leu Asp Phe Thr
65 70 75
Phe Glu Gly Lys Leu Glu Ala Ala Ala Gln Val Arg Gln Ala Thr
80 85 90
Asn Gln Ile Val Met Asn Cys Ala Asp Ile Asp Ile Ile Thr Ala
95 100 105
Ser Tyr Ala Pro Glu Gly Asp Glu Glu Ile His Ala Thr Gly Phe
110 115 120
Asn Tyr Gln Asn Glu Asp Glu Lys Val Thr Leu Ser Phe Pro Ser
125 130 135
Thr Leu Gln Thr Gly Thr Gly Thr Leu Lys Ile Asp Phe Val Gly
140 145 150
Glu Leu Asn Asp Lys Met Lys Gly Phe Tyr Arg Ser Lys Tyr Thr
155 160 165
Thr Pro Ser Gly Glu Val Arg Tyr Ala Ala Val Thr Gln Phe Glu
170 175 180
Ala Thr Asp Ala Arg Arg Ala Phe Pro Cys Trp Asp Glu Pro Ala
185 190 195
Ile Lys Ala Thr Phe Asp Ile Ser Leu Val Val Pro Lys Asp Arg
200 205 210
Val Ala Leu Ser Asn Met Asn Val Ile Asp Arg Lys Pro Tyr Pro
215 220 225
Asp Asp Glu Asn Leu Val Glu Val Lys Phe Ala Arg Thr Pro Val
230 235 240
Met Ser Thr Tyr Leu Val Ala Phe Val Val Gly Glu Tyr Asp Phe
245 250 255
Val Glu Thr Arg Ser Lys Asp Gly Val Cys Val Arg Val Tyr Thr
260 265 270
Pro Val Gly Lys Ala Glu Gln Gly Lys Phe Ala Leu Glu Val Ala
275 280 285
Ala Lys Thr Leu Pro Phe Tyr Lys Asp Tyr Phe Asn Val Pro Tyr
290 295 300
Pro Leu Pro Lys Ile Asp Leu Ile Ala Ile Ala Asp Phe Ala Ala
305 310 315
Gly Ala Met Glu Asn Trp Gly Leu Val Thr Tyr Arg Glu Thr Ala
320 325 330
Leu Leu Ile Asp Pro Lys Asn Ser Cys Ser Ser Ser Arg Gln Trp
335 340 345
Val Ala Leu Val Val Gly His Glu Leu Ala His Gln Trp Phe Gly
350 355 360
Asn Leu Val Thr Met Glu Trp Trp Thr His Leu Trp Leu Asn Glu
365 370 375
Gly Phe Ala Ser Trp Ile Glu Tyr Leu Cys Val Asp His Cys Phe
380 385 390
Pro Glu Tyr Asp Ile Trp Thr Gln Phe Val Ser Ala Asp Tyr Thr

WO 01/46443

PCT/US00/34811

Arg	Ala	Gln	Glu	Leu	Asp	Ala	Leu	Asp	Asn	Ser	His	Pro	Ile	Glu	405
				410					415					420	
Val	Ser	Val	Gly	His	Pro	Ser	Glu	Val	Asp	Glu	Ile	Phe	Asp	Ala	435
				435					440					445	
Ile	Ser	Tyr	Ser	Lys	Gly	Ala	Ser	Val	Ile	Arg	Met	Leu	His	Asp	450
				440					445					450	
Tyr	Ile	Gly	Asp	Lys	Asp	Phe	Lys	Lys	Gly	Met	Asn	Met	Tyr	Leu	465
				455					460					465	
Thr	Lys	Phe	Gln	Gln	Lys	Asn	Ala	Ala	Thr	Glu	Asp	Leu	Trp	Glu	480
				470					475					480	
Ser	Leu	Glu	Asn	Ala	Ser	Gly	Lys	Pro	Ile	Ala	Ala	Val	Met	Asn	495
				485					490					495	
Thr	Trp	Thr	Lys	Gln	Met	Gly	Phe	Pro	Leu	Ile	Tyr	Val	Glu	His	510
				500					505					510	
Glu	Gln	Val	Glu	Asp	Asp	Arg	Leu	Leu	Arg	Leu	Ser	Gln	Lys	Lys	525
				515					520					525	
Phe	Cys	Ala	Gly	Gly	Ser	Tyr	Val	Gly	Glu	Asp	Cys	Pro	Gln	Trp	540
				530					535					540	
Met	Val	Pro	Ile	Thr	Ile	Ser	Thr	Ser	Glu	Asp	Pro	Asn	Gln	Ala	555
				545					550					555	
Lys	Leu	Lys	Ile	Leu	Met	Asp	Lys	Pro	Glu	Met	Asn	Val	Val	Leu	570
				560					565					570	
Lys	Asn	Val	Lys	Pro	Asp	Gln	Trp	Val	Lys	Leu	Asn	Leu	Gly	Thr	585
				575					580					585	
Val	Gly	Phe	Tyr	Arg	Thr	Gln	Tyr	Ser	Ser	Ala	Met	Leu	Glu	Ser	600
				590					595					600	
Leu	Leu	Pro	Gly	Ile	Arg	Asp	Leu	Ser	Leu	Pro	Pro	Val	Asp	Arg	615
				605					610					615	
Leu	Gly	Leu	Gln	Asn	Asp	Leu	Phe	Ser	Leu	Ala	Arg	Ala	Gly	Ile	630
				620					625					630	
Ile	Ser	Thr	Val	Glu	Val	Leu	Lys	Val	Met	Glu	Ala	Phe	Val	Asn	645
				635					640					645	
Glu	Pro	Asn	Tyr	Thr	Val	Trp	Ser	Asp	Leu	Ser	Cys	Asn	Leu	Gly	660
				650					655					660	
Ile	Leu	Ser	Thr	Leu	Leu	Ser	His	Thr	Asp	Phe	Tyr	Glu	Glu	Ile	675
				665					670					675	
Gln	Glu	Phe	Val	Lys	Asp	Val	Phe	Ser	Pro	Ile	Gly	Glu	Arg	Leu	690
				680					685					690	
Gly	Trp	Asp	Pro	Lys	Pro	Gly	Glu	Gly	His	Leu	Asp	Ala	Leu	Leu	705
				695					700					705	
Arg	Gly	Leu	Val	Leu	Gly	Lys	Leu	Gly	Lys	Ala	Gly	His	Lys	Ala	720
				710					715					720	
Thr	Leu	Glu	Glu	Ala	Arg	Arg	Arg	Phe	Lys	Asp	His	Val	Glu	Gly	735
				725					730					735	
Lys	Gln	Ile	Leu	Ser	Ala	Asp	Leu	Arg	Ser	Pro	Val	Tyr	Leu	Thr	750
				740					745					750	
Val	Leu	Lys	His	Gly	Asp	Gly	Thr	Thr	Leu	Asp	Ile	Met	Leu	Lys	765
				755					760					765	
Leu	His	Lys	Gln	Ala	Asp	Met	Gln	Glu	Glu	Lys	Asn	Arg	Ile	Glu	780
				770					775					780	
Arg	Val	Leu	Gly	Ala	Thr	Leu	Leu	Pro	Asp	Leu	Ile	Gln	Lys	Val	795
				785					790					795	
Leu	Thr	Phe	Ala	Leu	Ser	Glu	Glu	Val	Arg	Pro	Gln	Asp	Thr	Val	810
				800					805					810	
Ser	Val	Ile	Gly	Gly	Val	Ala	Gly	Gly	Ser	Lys	His	Gly	Arg	Lys	825
				815					820					825	
Ala	Ala	Trp	Lys	Phe	Ile	Lys	Asp	Asn	Trp	Glu	Glu	Leu	Tyr	Asn	840
				830					835					840	
Arg	Tyr	Gln	Gly	Gly	Phe	Leu	Ile	Ser	Arg	Leu	Ile	Lys	Leu	Ser	855
				845					850					855	
Val	Glu	Gly	Phe	Ala	Val	Asp	Lys	Met	Ala	Gly	Glu	Val	Lys	Ala	870
				860					865					870	
Phe	Phe	Glu	Ser	His	Pro	Ala	Pro	Ser	Ala	Glu	Arg	Thr	Ile	Gln	885
				875					880					885	
Gln	Cys	Cys	Glu	Asn	Ile	Leu	Leu	Asn	Ala	Ala	Trp	Leu	Lys	Arg	900
				890					895					900	

WO 01/46443

PCT/US00/34811

Asp Ala Glu Ser Ile His Gln Tyr Leu Leu Gln Arg Lys Ala Ser 915
 905
 Pro Pro Thr Val

<210> 10
 <211> 209
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4043361CD1

<400> 10
 Met Glu Gln Pro Arg Lys Ala Val Val Val Thr Gly Phe Gly Pro
 1 5 10 15
 Phe Gly Glu His Thr Val Asn Ala Ser Trp Ile Ala Val Gln Glu
 20 25 30
 Leu Glu Lys Leu Gly Leu Gly Asp Ser Val Asp Leu His Val Tyr
 35 40 45
 Glu Ile Pro Val Glu Tyr Gln Thr Val Gln Arg Leu Ile Pro Ala
 50 55 60
 Leu Trp Glu Lys His Ser Pro Gln Leu Val Val His Val Gly Val
 65 70 75
 Ser Gly Met Ala Thr Thr Val Thr Leu Glu Lys Cys Gly His Asn
 80 85 90
 Lys Gly Tyr Lys Gly Leu Asp Asn Cys Arg Phe Cys Pro Gly Ser
 95 100 105
 Gln Cys Cys Val Glu Asp Gly Pro Glu Ser Ile Asp Ser Ile Ile
 110 115 120
 Asp Met Asp Ala Val Cys Lys Arg Val Thr Thr Leu Gly Leu Asp
 125 130 135
 Val Ser Val Thr Ile Ser Gln Asp Ala Gly Arg Lys Lys Pro Phe
 140 145 150
 Pro Ala Lys Gly Asp Cys Val Phe Cys Arg Arg Arg Ala Arg
 155 160 165
 Ser Leu Gln Ala Gln Cys Gly Phe Ser Leu Thr Pro Ala Leu Glu
 170 175 180
 Leu Leu Pro Val Pro Phe Leu Lys Leu Leu Cys Pro Gly Pro Pro
 185 190 195
 Arg Arg Arg Arg Ile Cys Arg Ile Leu Pro Gly Ala Gly Leu
 200 205 210

<210> 11
 <211> 77
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3937958CD1

<400> 11
 Met Gly Lys Glu Lys Ala Leu Ser Leu Gln Met Met Lys Tyr Trp
 1 5 10 15
 Ala Asn Phe Ala Arg Thr Gly Asn Pro Asn Asp Gly Asn Leu Pro
 20 25 30
 Cys Trp Pro Arg Tyr Asn Lys Asp Glu Lys Tyr Leu Gln Leu Asp
 35 40 45
 Phe Thr Thr Arg Val Gly Met Lys Leu Lys Glu Lys Lys Met Ala
 50 55 60
 Phe Trp Met Ser Leu Tyr Gln Ser Gln Arg Pro Glu Lys Gln Arg
 65 70 75
 Gln Phe

WO 01/46443

PCT/US00/34811

```

<210> 12
<211> 414
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7257324Cu1

<400> 12
Met Asn Pro Thr Leu Gly Leu Ala Ile Phe Leu Ala Val Leu Leu
1      5      10      15
Thr Val Lys Gly Leu Leu Lys Pro Ser Phe Ser Pro Arg Asn Tyr
20     25     30
Lys Ala Leu Ser Glu Val Gln Gly Trp Lys Gln Arg Met Ala Ala
35     40     45
Lys Glu Leu Ala Arg Gln Asn Met Asp Leu Gly Phe Lys Leu Leu
50     55     60
Lys Lys Leu Ala Phe Tyr Asn Pro Gly Arg Asn Ile Phe Leu Ser
65     70     75
Pro Leu Ser Ile Ser Thr Ala Phe Ser Met Leu Cys Leu Gly Ala
80     85
Gln Asp Ser Thr Leu Asp Glu Ile Lys Gln Gly Phe Asn Phe Arg
95     100    105
Lys Met Pro Glu Lys Asp Leu His Glu Gly Phe His Tyr Ile Ile
110    115    120
His Glu Leu Thr Gln Lys Thr Gln Asp Leu Lys Leu Ser Ile Gly
125    130    135
Asn Thr Leu Phe Ile Asp Gln Arg Leu Gln Pro Gln Arg Lys Phe
140    145    150
Leu Glu Asp Ala Lys Asn Phe Tyr Ser Ala Glu Thr Ile Leu Thr
155    160    165
Asn Phe Gln Asn Leu Glu Met Ala Gln Lys Gln Ile Asn Asp Phe
170    175    180
Ile Ser Gln Lys Thr His Gly Lys Ile Asn Asn Leu Ile Glu Asn
185    190    195
Ile Asp Pro Gly Thr Val Met Leu Leu Ala Asn Tyr Ile Phe Phe
200    205    210
Arg Ala Arg Trp Lys His Glu Phe Asp Pro Asn Val Thr Lys Glu
215    220    225
Glu Asp Phe Phe Leu Glu Lys Asn Ser Ser Val Lys Val Pro Met
230    235    240
Met Phe Arg Ser Gly Ile Tyr Gln Val Gly Tyr Asp Asp Lys Leu
245    250    255
Ser Cys Thr Ile Leu Glu Ile Pro Tyr Gln Lys Asn Ile Thr Ala
260    265    270
Ile Phe Ile Leu Pro Asp Glu Gly Lys Leu Lys His Leu Glu Lys
275    280    285
Gly Leu Gln Val Asp Thr Phe Ser Arg Trp Lys Thr Leu Leu Ser
290    295    300
Arg Arg Val Val Asp Val Ser Val Pro Arg Leu His Met Thr Gly
305    310    315
Thr Phe Asp Leu Lys Lys Thr Leu Ser Tyr Ile Gly Val Ser Lys
320    325    330
Ile Phe Glu Glu His Gly Asp Leu Thr Lys Ile Ala Pro His Arg
335    340    345
Ser Leu Lys Val Gly Glu Ala Val His Lys Ala Glu Leu Lys Met
350    355    360
Asp Glu Arg Gly Thr Glu Gly Ala Ala Gly Thr Gly Ala Gln Thr
365    370    375
Leu Pro Met Glu Thr Pro Leu Val Val Lys Ile Asp Lys Pro Tyr
380    385    390
Leu Leu Leu Ile Tyr Ser Glu Lys Ile Pro Ser Val Leu Phe Leu
395    400    405
Gly Lys Ile Val Asn Pro Ile Gly Lys
410

```

WO 01/46443

PCT/US00/34811

<210> 13
<211> 397
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472038001

```

<400> 13
Met Pro Arg Ala Ile Ser Pro Leu Met Arg Phe Gln His Pro Val
1      5      10      15
Ser Cys Lys Leu Gln Leu Tyr Arg Val Pro Leu Arg Arg Phe Pro
20     25     30
Ser Ala Arg His Arg Phe Glu Lys Leu Gly Ile Arg Met Asp Arg
35     40     45
Leu Arg Leu Lys Tyr Ala Glu Glu Val Ser His Phe Arg Gly Glu
50     55     60
Trp Asn Ser Ala Val Lys Ser Thr Pro Leu Ser Asn Tyr Leu Asp
65     70     75
Ala Gln Tyr Phe Gly Pro Ile Thr Ile Gly Thr Pro Pro Gln Thr
80     85
Phe Lys Val Ile Phe Asp Thr Gly Ser Ser Asn Leu Trp Val Pro
95     100    105
Ser Ala Thr Cys Ala Ser Thr Met Val Ala Cys Arg Val His Asn
110    115    120
Arg Tyr Phe Ala Lys Arg Ser Thr Ser His Gln Val Arg Gly Asp
125    130    135
His Phe Ala Ile His Tyr Gly Ser Gly Ser Leu Ser Gly Phe Leu
140    145    150
Ser Thr Asp Thr Val Arg Val Ala Gly Leu Glu Ile Arg Asp Gln
155    160    165
Thr Phe Ala Glu Ala Thr Glu Met Pro Gly Pro Ile Phe Leu Ala
170    175    180
Ala Lys Phe Asp Gly Ile Phe Gly Leu Ala Tyr Arg Ser Ile Ser
185    190    195
Met Gln Arg Ile Lys Pro Pro Phe Tyr Ala Met Met Glu Gln Gly
200    205    210
Leu Leu Thr Lys Pro Ile Phe Ser Val Tyr Leu Ser Arg Asn Gly
215    220    225
Glu Lys Asp Gly Gly Ala Ile Phe Phe Gly Gly Ser Asn Pro His
230    235    240
Tyr Tyr Thr Gly Asn Phe Thr Tyr Val Gln Val Ser His Arg Ala
245    250    255
Tyr Trp Gln Val Lys Met Asp Ser Ala Val Ile Arg Asn Leu Glu
260    265    270
Leu Cys Gln Gln Gly Cys Glu Val Ile Ile Asp Thr Gly Thr Ser
275    280    285
Phe Leu Ala Leu Pro Tyr Asp Gln Ala Ile Leu Ile Asn Glu Ser
290    295    300
Ile Gly Gly Thr Pro Ser Ser Phe Gly Gln Phe Leu Val Pro Cys
305    310    315
Asp Ser Val Pro Asp Leu Pro Lys Ile Thr Phe Thr Leu Gly Gly
320    325    330
Arg Arg Phe Phe Leu Glu Ser His Glu Tyr Val Phe Arg Asp Ile
335    340    345
Tyr Gln Asp Arg Arg Ile Cys Ser Ser Ala Phe Ile Ala Val Asp
350    355    360
Leu Pro Ser Pro Ser Gly Pro Leu Trp Ile Leu Gly Asp Val Phe
365    370    375
Leu Gly Lys Tyr Tyr Thr Glu Phe Asp Met Glu Arg His Arg Ile
380    385    390
Gly Phe Ala Asp Ala Arg Ser
395

```

<210> 14
<211> 145

WO 01/46443

PCT/US00/34811

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472041CD1

```

<400> 14
Met Gly Ile Gly Cys Trp Arg Asn Pro Leu Leu Leu Ile Ala
1 5 10 15
Leu Val Leu Ser Ala Lys Leu Gly His Phe Gln Arg Trp Glu Gly
20 25 30
Phe Gln Gln Lys Leu Met Ser Lys Lys Asn Met Asn Ser Thr Leu
35 40 45
Asn Phe Phe Ile Gln Ser Tyr Asn Asn Ala Ser Asn Asp Thr Tyr
50 55 60
Leu Tyr Arg Val Gln Arg Leu Ile Arg Ser Gln Met Gln Leu Thr
65 70 75
Thr Gly Val Glu Tyr Ile Val Thr Val Lys Ile Gly Trp Thr Lys
80 85 90
Cys Lys Arg Asn Asp Thr Ser Asn Ser Ser Cys Pro Leu Gln Ser
95 100 105
Lys Lys Leu Arg Lys Ser Leu Ile Cys Glu Ser Leu Ile Tyr Thr
110 115 120
Met Pro Trp Ile Asn Tyr Phe Gln Leu Trp Asn Asn Ser Cys Leu
125 130 135
Glu Ala Glu His Val Gly Arg Asn Leu Arg
140 145

```

<210> 15
<211> 4828
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: V/14846C31

```

<400> 15
gccatccggc ggggcgcgct cctccggctc cctctctccc ttccccaag cagcccgcg 60
accgcaacc aagaaacatc cgaaccgctg acccccccct actggctcgc gctctcccg 120
ggcccgccgt cctccgccc ctaatggccc gggcccgccg ggaagagagg cgtcagggcg 180
ggcccgctgg agacgtgagg cggccgcccgt ggcctccaca gtggcgcttt cgcgcgctgg 240
cggcggtgcc cgcgcacgcc ggcgcacatc gccttcggcg ctggctggcg gggcgctgt 300
cctccagggc cgtccggccc gctccctgga gctcggcgga ggcggcagc caggggcggc 360
ggagcggcga ggaacgggac gccaccgccc ccgcgcgccc cgcgcgccc gggcccatga 420
cgtggagaca gaacgctgtg cagcagagcc cggcgcacaa gcaccagag acgcttttga 480
atccactgag aqaanttacc gggatcaatg acaccagat attacagcaa gcttgaagg 540
atagtaatgg aaacttgaa ttacagctgg ctttccctac tgcgaagaat gctaaagcc 600
ctcagcaaga ggaacaacct taclaccaaa cagcacttcc tggcaatgat agatacatc 660
gtgtgggaag ccaagcagat acaaatgtga ttgatctcac tggagatgat aagatgatc 720
ttcagagagc aattgccttg agtttggccc aatcaaacag ggcattcagg gagactgaa 780
taactgatga ggaacaagcc attagcagag ttcttgaagc cagcatagca gagaataaag 840
cstgtctgaa gaggacacct acagaagttt ggaaggattc tggaaacct taigtatgaa 900
aaagcaaga aaagctccc gttgggctaa gaaatgttgg caatacttgt ggttttgg 960
ctgttatcca gtcattattt aatccttgg aattagaag attagtctg aattacaagc 1020
ccccatcaaa tgcctaaagt ttccccgaa accaaaaggc acatcggaat ttgcttttta 1080
tgctgagct gaggtatcta ttgcacttc ttgttggtac caaaaggaag taigttyatc 1140
catcaagagc agttgaattt cttaaagatg ctttcaaatc aantgactca cagcagcaag 1200
atgtgagtga gtttacaac aaattattag attggttaga agatgccttc caaatgaaag 1260
ctgaagagga gacgatgaa gagaagccaa agaaccocat ggtagagttg ttctatggca 1320
gatcctggc tggggagta cttyaggtta aaaaaattga aaccactgaa atgttttggc 1380
agtaacctc taaggtaat gggttcaag atctgpatga gtccctgaa cctgcaaga 1440
tgaagagga attgagctc ttacattcag agnattcagg aaatccggc caagagcatt 1500
ggtttactga attaccacct ggttaacat ttgaattgtc aagattgaa ttbaatcagg 1560
cattgggaag accagaaaa attcaaca caattgaatt tcccgaagtt ttatatttgg 1620
acagatacat gccagaaaac aggaataaa caagaattaa gaggyaagag atcaagagac 1680

```

WO 01/46443

PCT/US00/34811

```

tgaagaatta cctceogyta ttacaacaaa gctagesag AlalltAagc latggttccg 1746
gtcccaaacg altccccttg gtgatgttcc ttcagtagcc attggaattt gcctcaagta 1800
aacctgtttg cacttctect gtbacagata ttgacgtag tcccacact agtggttcca 1860
taacctcaaa gacattacca agcaaacag acaaacaggg agcctatct tccagactgc 1920
caagcaalc accctcaaca gttgctgcca ttccatcgag atcagtaata caaaaacct 1980
ttactcaatc ccggtacact ccagatltgc ccatgcatcc ggcaccaagg ccaataaagg 2040
aggaagaaat ttctgtgtgt gaaagttgtt taactogctg gaggaagaa atggaagatg 2100
accocagaga ttgcaggaaa agcattacca gaatccatcg aacaatgaa ttaatgtaat 2160
ctgacaatc latgatacaa gttccctatc gatfacatgc cgttttagtt cccgaaggcc 2220
aagctaagtc tgggcactac tgggcataca tttttgatca tctgtaaagc agatggtatg 2280
agtcaaatga tattgtctgtg acaaaaatcat catgggaaga gctagtgaag gactcttttg 2340
gtggllatag aaatgccagt gcatactggt Laatgtacat aatgataag gcacagtcc 2400
taatacaaga gtagtllaat aaagaacctg ggcagccctc tgttggtata gaaacattac 2460
caccggattc gagagatttt gllgaggaag acaaccaagc atttgaaaa gaactagaag 2520
aatgpaagc acaactgtcc cagaaggttt tgcagaaaaa gcttttagcg tcccagaat 2580
tgaagagtc agagacttct gtgacacag ccccaagcagc aggaagcccc gaatactag 2640
agcagccatc aagaagtgat ttctcaagc acttgaaga agaacattt caaataatta 2700
caagggcacc acatgagcat gaagataaaa gctctgaaac agttttcgag tccgcaatta 2760
agttgaaata tgcagggttg gttaaagttg ccaagaaga caccacca gaaaccgat 2820
atcgtttaca tcatgtagt gtctacttta tccagaaca ggcaccaag aaaaattatg 2880
agaaacatt actagaacaa rrtggagata gaaatttgag tttgatgaa aggtgtaca 2940
acaacatgaa agttgctcaa gccaaactgg aaalgtaaa acctgaagaa gtaaacctgg 3000
aggaatagaa gtagtggcat cagatltata ggaactcag gaaacact atgtatcca 3060
taatbggct agaaatttt caaagaaaa gttatataga ttccctgtct tctcactct 3120
gtgclatca gataacaaa gactcttgt ctaaaagcct ataccagaa catgatgag 3180
aatgatac acattataga agagaatgtt tgcataaatt aatgagcaa gccagcaac 3240
tcttgaatc tggagaggtat cgagaagtaa ccaatggttt gattatcatg aatgattta 3300
tgtcccaatt ttgcccatta tctctggtg atgaaatgaa agaaaagat atactagct 3360
tagaagat atgaaatgaa tgggttcccl accttggta agaaatgaa ccaaccctcc 3420
aagaaagct gacagatttt ttgcaaaac tgccttgatg tctctgag attaaaagt 3480
tccatggcc accgaagttt ccttcaatll cccagatg acicctgtag agatttccc 3540
gaactatgt gtccctcagt cgaactcctg ctgatggag ataaactgca cacttccct 3600
gaacacactc lataaactct ttttagttct taaccttgc cttcctgtca accggtttgc 3660
tgttctgctc tatagttttt aactttttt taltttaata actgcaaaag acaaaatgac 3720
tatacagact ttagtccagc tgcagacaa atagctgaaa atgcaatg gcctcagac 3780
tttaaccgga actgatgat aatcacaat ctatgtgat tbatatgyc aaaaatagc 3840
tttlycacc ttccgttgc agtataact tgccttlatc ttttctttc caaacgctt 3900
caattcagtc tggatctctc catgactaca gcaatttaag tgttcagcac tgtgtacga 3960
acaatatt ttgtagcttg taatgaaat aaagataaa gttttatta tggctacct 4020
aaaaaaa

```

```

<210> 16
<211> 1422
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> Incyte ID No: 1856589CB1

```

```

<400> 16
ggccocggca ggcaggggtg gtcgocaggg agcogtcaac tgcctctcc ctcogcetc 60
ccctcagggc cagggggcca ggaacccgga gcagcggat gggagccgc acctytagg 120
gctccaggat cccagcggc ccccagctcc agggggaacg cagtgcgccc cgtctcggty 180
ttacttccct cagcgtggg ccaggggaat tcaaggaaa ctggagatt ccggctcca 240
gacagaaat ggcctggca ggtgagcctg cayaccagca acagacacat ctgctggagc 300
tccctatcg ccagcaactg ggttataaag agyacacaa caatccagtt tctgtgagc 360
cctgggtgct gtaggatttg gaaatgacc gccattggcc ctgggaggtg agcctccgga 420
tggaaaatga gcaactgtgt ggaagggccc tcaattgccc cagctgggtg gtagctcgg 480
cccaatgag ccaaggcacc aaagagtaat cagtgggtct tggcactcc aagctcagc 540
ccatgaact ccgaggggccc cteTgggtcc ctgtgagggc catcatatg caacccaagt 600
actggagccg agcctcaatc atgggtgagc ttgcccctgt ccaactcaaa acacctgtca 660
cctccagta gtaactgaa ccaactgtcc tccgggagc caatttcaac ctgaggtgt 720
gagcagatg ttgggtgact ggttgagccc aggttaagca ggcgttttca ggtccacag 780
caactccat gotgacocca gtagctcagg agcctgaggt gttatcatg gacaacaaga 840
gagtgaccg gcaatacaag aagtcctctt tcccctagt tgcctccct gctcctgggg 900
acatgatctg tgcacacaa latggggaaa actgtgata tggggattct ggagggccat 960

```

WO 01/46443

PCT/US00/34811

```

tggctgtgga agttgagggc agatgggatt tggctggggt gttgtcctgg gaaagggcct 1020
gggtggaagg acagaaatccu gglgtgtaca cccggctcnc caaalacacc aaatggatca 1080
agaagcaaat gagcaatgga gcotttctag gtccctgtgc ctctgcttcc ctctgttcc 1140
lgtgtggccc gctgcaagccc agatgggctt cctgacctcc ctactcttcc ctctctctgc 1200
ctgtcctctg ctgaaatgggg ccagatgggt tgaccaggtt cagtgttcca tctcaaaaa 1260
gaatcagggc agggaaaggt aaccctctgg agaatgggtc tggcttttgg ctccctgtga 1320
gggaagatgt ggtgtatgac taggcttgg gtgagcggg aagggaaag- gtgcccaga 1380
agattcttgg aatctgggac cagggagagca gggat'aaac at 1422

```

<210> 17
 <211> 1911
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2617672CE1

```

<400> 17
cccacccctc cgcggggcgt cgcagagcca ggagggggag gcccgggggc cagcctgggc 60
cccagcccac accttcaaca gggcccaggc gccaccalgt ggcgaatgcc actgggggcta 120
ctgtlgttgc tgcagcthgc tggccacttg gctctgggltg cccagcaggg tctlgyggcgc 180
cggagctkag caccgggctt gcaacctggg ggcctctggg ccggcggagg ccgtactctg 240
caggagccag acctgtgctg ccggggcgtl gccacagact gtacctgccc ctactctggc 300
gccatctggt actgtgacct ctctgtcaac cgcacggctt ccgactgtct cctgtacttc 360
tgggacctct gactcgggct gccacccctt tttcccaga tccaaggatg tatgatgga 420
ggtcgtatct atccagctct gggcaactac tgggcaact gtaaccgttg cactgcccag 480
gagaacagge agtggcagtg tgaccaagaa ccatgctctg tggatocaga catgatcaaa 540
gccatcaacc agggcaacta tggctggcag gctgggaacc acagcgcctt ctggggcaatg 600
accctggatg agggcaatcg ctaccctctg ggcaccatcc gccactcttc ctggctcatg 660
aacatgatg hantttatad agtgtgaaac ccaggggggg tgcctccacc agcctctgag 720
gacctgaga agtguccaa cctgatctat gacactcttg accaaggcaa ctgtgcaagc 780
tcctggacct tctccagag agctgtygca tccyatctgt tctcaatcca tctctctgga 840
cacatgagcc ctgtctctgc gcccccagac ctgtctctt gtgacacca ccagcaagcag 900
ggctgcccgg gtggcgtctt cpatgtgtcc tgggtgtlcc tgcgtctccg aggggtggtg 960
tctgacctct gctacccttt ctggggcctt gaaagagacc aggtctggcc tgcctcccc 1020
tgtatgatgc acagccagac catgggtctg ggaagctgcc aggcaccctg ccaatgcccc 1080
aacagctatg ttatatacaa tgactctctc caagctact ctgtctacc ctctggctcc 1140
aacgcaaggg agatcatgaa gtagctgatg gagaatggcc ctgtccagc ctctctggag 1200
gtgcaagggc actctctctt ataacaagga ggcactctac gccacaagcc agtgagcctt 1260
gggagccag agagataccg ccggcatggg cccactctag tcaagatcac agtatgggga 1320
yagggagacc tgcagatgg aggacactc aatactgga ctgcccaca ctctggggc 1380
ccagcctggg gcagaggggg ccactctccc atctgtggg cgtctaatga gtgcccctc 1440
agagacttgc tgcggggcgt ctggggcccgc gtaggcatgg aggacatggg tcatcactga 1500
ggctgctggc accacggggg gtcaggcctg ggtaccaggc taagggccgg cggaaagggc 1560
cccaggggg cgtggaaccc agctctggcc gcaagagccc gggggccagf cggggccag 1620
ggcctactc ccggggccgg tccgctgac ccagcctccc gctgggagc cggggccag 1680
cggactctgc gggccccag acctccagtt ggggagcggg caggcccttg cctgggaaga 1740
gccagctgc agatcccagg cctctgggcc cccactctaa gactacaaa gccaggaacc 1800
ctcaagctc cagccccact acccccacc cccctcttat tctattctt cagatattta 1860
ttttctttt cactgtttta aaaaataaac aaagtattga taaaaaaaa a 1911

```

<210> 18
 <211> 854
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2769104C31

```

<400> 18
caccttttgt tccctatcct gggccagttc tctcggaggt cccagatgc cagttccaga 60
tgccctggacc cagagtgctg gggaaatctc tctggagaag cctctactcc aaagcctgto 120
cagggcctaat gtagtgctg ctctctctgg gactctctca ggcctgccc acccgggct 180
ccgtctctctt ggcaccaag ctccccagc agctgacatc cccgggttac ccagagccct 240
atggccaagg ccaagagagc agcacggaca tcaaggtctc agagggctt gctgtgagcc 300

```

WO 01/46443

PCT/US00/34811

```

tgcctctcca ggaacttcgac ctggagccgt cccaggactg tgcaggggac tctgtccaca 360
tctcatttgt cyygtgggat ccaagccagt tetgtggtca gcaaggclcc cctctgggca 420
ggccccctgg toagaggag ttgtatcct caggagagag ttgctgctg acctccgca 480
cacagccttc ctctggagac aagactgccc acctccaca gggcttctct gcccctacc 540
aaaccctggg tgaagtgtccc tctgggggt gcaggagag agcctctgt cccggccatg 600
accctgggat ctccaagcct taagtggaag cttagtggac agctyagpct gggppactca 660
ggcaactcgg gctggtatccc agcctggccc ctctggaac gcaaacctat taagagacag 720
cgtgactcga gcccccagcg gtrgtttcca tgcagaita caggcccagt gtttgcagat 780
catctcattc ttaagagat gcccaaaatc cagatttita agttaaaita taaatttica 840
aaanaaaaaa aaaa 854

```

```

<210> 19
<211> 1386
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4802789CE1

```

```

<400> 19
gacgctgagg cccggccccg cgggtaata acagatgcyg gtgaaagac caactaaagc 60
ttactctggg aaagccaaaa gaaataaag gctactgta cctctgagt aagpctcttc 120
agatgatatt gctgtagggt taacttgcca acatgtaagt cctgctatca gctgtaatca 180
tgtaaagaga gcaatagctg agaactctgt gtcagltgc tccagatggt taagaagaa 240
aagattctat gatgggcagc tagtacttac tctgtatatt tggttgtgcc tcaagttygt 300
cttccaggga tgtgttaaaa actcgaagag ccaacatca tgaagcact taaagatlc 360
cagaaacagag cccattgta tataatlaa tctgagcaca tggattatc ggtttatga 420
atgtgatgaa aaattatcaa cgcattgtaa taagaaggtt ctggctoaga tagttgatit 480
tctccgaaaa catgctctca aaacacaaac aagtgcaatc tctagaatca tgaactttg 540
tgaagaaaa tggaaacag atgaaatca caggagagc aactgcagaa ctatctctgt 600
agaggaatt acaaatctag gaaactctg ctcttttaac gcaagctatg agaacttgc 660
acagacttat actcttactg atctgatgaa tgcgatacaa gcaagtagta caaactcca 720
gatttttctc tctccagact ctacgtgga cccattggt gtggaactll caagcctgg 780
accactgacc taagccttgt tctgtttct tcaagcagc agggagactg aaaaaggacc 840
acttctctc aaagtctctt ttaactagct ttgtragaag tgggtgcatc taacttaac 900
ataaatatt atgagtaca aaatactaat gtatlcata ttaaacatga atagtgltt 960
ttactgtaac ttgtctctca ttgcccagac catgaaagaa actaaacttt gttcaagctc 1020
tktgtttat gaaatata tctcagctct cagaagagc agtctctct catataaag 1080
gaaatggaga cacagaatc aaatggctca cctagtctga gtgaaagct gagaatcaaa 1140
tggagactct tctcactctg gatgctatg ttgtaatacc ataaagttag aaaaaccatg 1200
agttgtaaaa tctagaagat accgtaagat aacatctaac ctagcttctc taltttaaaa 1260
gatgagctgt gaggcaata gatttaagt gaatttctca aggtattaca gtatgtttaa 1320
aaacccaatc cttatgtgcc tggaaataaa caataaagg atctgacttg aaaaaaaa 1380
aaaaaa 1386

```

```

<210> 20
<211> 1323
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 60116897CE1

```

```

<400> 20
caaatctgca gcaagcatgat ttaagattaa attoatgtat tgaaaatatt gttcagacc 60
catgtgacat aaotggagcc agtgcagtgc catgaagaac tacgagatta gcttggatat 120
taactgtctc tctagagaaat agatttcatg tccattctt ctgcaatggt taattccac 180
agaaaaaaca tgtttaacat tcaagagaga ttttactgct taacagcct ctgccccaa 240
atagtcaatt gttctcagtt ctcaagtcca tctagtctac actcaactga ggaactctgg 300
gcttccagag tagccactaa tggggaacga tttcttggc aggagctaa 360
gtgtcaatc ctctcaatta tgaactctt gtcaccca atctcaactc tctgacttt 420
gttgcactg agaagattga agctctgctc agcaatgcta cccagtttat catctgccc 480
agcaagatc ttgaaatcac gaaatgcaac cttaagttag aggaagatto aagataactg 540
aaacccagaa aagaactgaa agtttggagt taccctgctc atgacaaat tgcactgctg 600
gttccagaga aaottagccc tcaactgaaa tactatgtgg ctatggactc ccaagccag 660

```

WO 01/46443

PCT/US00/34811

```

ttagggtgatg gctttgaagg gllllataaa agcacataca gaactcttgg tggtagaaaca 720
agaattclly cagtaacaga ttttggagca acccaggcac gcatggcttt cccctgattt 780
gatgaacogt tgttcaaaag caactlltca atcaagatac gaagagagag cagggcatatt 840
gcactatccc aontgocaaa ggttaagaca attgaacttg aaggaggtct tttggaugat 900
caactttgaaa ctactgtataa aatgagtaca taacttghta cctacaiaat tttggtakto 960
cactctctga gtggctctac tteatcaggg gtcaaggtgt ccatctatgc atcccagac 1020
aaaaggaatc aaacacacta tgotttgaaq gcatcaactg aagtaactga ttttatlga 1080
aagtaactllg atatcacta tcaactctcc aaactggact taattgctat tctgactrr 1140
gaccctggag ccatgaaaaa ttggggcctc attacatata gggagacgic actgctttt 1200
gaccocaaaga cctctctctg tccgataaa ctgtggttca ccagagtcat agcccatgaa 1260
ctggcgacc accgtgtttg caacctggtc acnatggnaat ggtggaatga tatttggctt 1320
aaggagggtt ttgcaaaaa caiggaactt atcgtcttla atgclacata tccagagctg 1380
caatttgatg actatllttt gaatgtgtgt ttggaagtaa ttacaaaaga ttcaattgaat 1440
tcactcccgc ctatctccaa accagcggaa accccgactc aaatacagga aatgtttgat 1500
gaggtttccl ataacaaagg agcttgtatt ttgatattg tcaaggtttt tctgggtgag 1560
gggaaattcc aqaaggaat aattcagtac ttaaaagagt tcaagctatag aaatgtlaag 1620
aatgatgact tgtggagcag totgtcaaat agttgtttag aaagtattt tacaatcgtt 1680
ggagltttgc attcggatcc caagatgaca agtaacatgc tggcctttct ggggaaaaat 1740
gcagaggtca aayagatgat gactacatgg actctccaga aaggaatccc cctgctgglg 1800
gttaaacaaq accggtgttc actccgactg caacaggagc gcttctctca ggggttttct 1860
caggaagacc ctgaaatggag ggcctcagc gagaggtacc tgtggcatab cccattgaac 1920
tactccacga gctctctcaa tglgalecac agacacatct caaatcama gacagatact 1980
ctggatctac ctgaaaagac cagttggtgt aatlttaagy ttgatctaaa tggttactac 2040
atggtctact atgaggttca tggatgggac caactaata cacagctgaa tccagaaccac 2100
acactcttca gaccatayga cagatagagt ctgatctatg atgtcttca gctagtggg 2160
gcagggagac tgaacctaga caaagctctt gacatgactt actactcca acatgaaaca 2220
agcagccccc cactctctga aggtctgagt tacttggaa ctgtttacca catgatggac 2280
agaaggaata tttcagatct ctctgaaaac ctcaagcgtt acctcttca gtatlttaag 2340
ccagtcatct ccaagcaaaq ctggagtgc aagggtctag tctgggacag gatgctccc 2400
tcggctctct tgaagctggc ctggaactg aaactgtct ctgcatcca gaagctgct 2460
gaaactctct ccaagtggtt ggaatccgt ggaataatga atatacaaa agatlttta 2520
angattgtgt atctgttgg tctcagaca accagagat ggaattacct tttagagcaa 2580
latgaactgt caatgtcaag tctgtaacan aaacaaatct tgtatgctt gtcnccgagc 2640
aagcatcayg aaaaattact gaagttaatl gaactaggaa tggaaagaaa ggtlatcaag 2700
acacagaact tggcagctct coltcatgcy atlgcaagc atgcaaaagg gcagcaacta 2760
gcatgggatt ttglaagaga aatbtgaacc catctctga aaaaattga ctgggctca 2820
latgacataa gnatgacat ctctggcaca acagctcaat tttcttcaa ggaataqtlg 2880
caaggttga aactatctt tgaatctctt gaggctcaag gatcaactct ggalatttt 2940
caaaccttct tggaaagat aaacaaaat ataaatggt tggagaagaa tcttccgct 3000
ctgaggactt ggtcaatggt taactactaa atggtcaata gaaaagtag gctggcgcgc 3060
gtggctcacc cctgtaatcc caqcaacttg ggagctctag aaggcgggat caaggttca 3120
ggagatggag accatctctg ctaaacaggt gagaccocyt ctccgctaaa aatacaaaaa 3180
atagccggg catggttggca ggtgcctgla gtcccagcta ctccgagag tgcagcagga 3240
aaaaggcata aaccggggag gtggagcttg cagtgcagc agattgccc actgattcc 3300
agcctgggtg actgagcgg act 3323

```

```

<210> 21
<211> 2123
<212> DNA
<213> Homo sapiens
<330>
<331> misc_feature
<332> Incyte ID No: 1866356C31

```

```

<400> 21
tgacaatcca agatgggggt gcccgccgag gccgagagg agccgacagt ttacttggta 60
gtgaggggta tcccctcctg gttgcctctg gcccaattac ggagctatatt tagccagtct 120
cggagaagagc gccggcgttg ctctctctgt ttccactacc ggcctcggcc tgagggggcc 180
cctccgaggg ccgctcctaa ctctgcccctt attcctaccg acccagccgc tgagggccag 240
ctctctctct agactctggc caecgatgto cggctctctc ccaactgaga ctctactcca 300
atccagacc ccaactctctg ctgctctctc tctgttaggg ggttggctca agctcagag 360
ctttctgca tgtactcagg ccgcgggtgg ctggatctc accggaactg gctaccgggt 420
cgcctctca tccgagact tggctactct accgagccat caggctctgg ctctctccc 480
tcaagaacc ggaaggaact gcaagttgg aaggcagaya atgaagcctt aacctggct 540
gacctgaagc aactccggga gctgaacca ccaatgtctg tgcocagag gaatgtggg 600
actcccctgc ggtctttttt ggaattgatc cgggctctcc gctaccccc tgggatcac 660

```

WO 01/46443

PCT/US00/34811

```

aoccaagctgc agctccaggt ccccaagaca ggttccctcc ggcgctaccg caalgtgctt 720
tttgatgatg aggactcaga gctctgtggg cagggaugagc ttgtgtatac agcagagggt 780
gaagaataac cccaaggaac ctactctggca gatataccag ccagccctct tggagagcct 840
gaggagaagag tggggcaggga agaggaagaa gagtctcact cagatgagga csgatgaccg 900
ggtgaggaat tggaaagcga tgaagcgccty catgaggacy tgaaccggca gpagcggacc 960
actgagcagc tctttgggga ggaagatggg ctcaagtggg agaagggggy ctctggctg 1020
gtcttttata ctgagtccca gtctctggca gaggagaag ggaatttga tgaacaagca 1080
gocgagactc gggatgtgga catgagtggt taclatgaca gagatggtg agcaagagt 1140
goccgagact ctgtccaaat gcgtctggaa cagagaactcc gagatggaca ggaagatggc 1200
ctgtgatcg aagccaggtt gggcaacttt gaggccaca ccagggcat tggcggaaq 1260
gtgatggagc ggcaggccty ggcctgagggc caggccctgg gctgcagty ctcaagggty 1320
cctgaggccc tggatagtga tggccaacac ccagatgca agcgtggat ggggtaccat 1380
ggagagagc tacagcatt tgggaacty aagagcccc gtagaatgg ctggggctc 1440
atctccaca tctatgalya gctctcaacc caagaccaga gggagtcact gctccyccg 1500
cagccacaca ccagcaaaa gttctggaca gacatggct ttgaggggy tccagttgt 1560
gottcagaca gccctcatt gctctgactga ccgggttggg ggttctctt catagtaca 1620
tgatgaaaac cctctgccc ttgcctcactt accactgaaq cagaaggag tctgggaca 1680
gagctctctg tggctgttc aggtgtttt gttccgacc tgcctgcty ccggttcta 1740
acctcagggg catttttaca aaaagcccc tccctcccc tcccttgya tattagggty 1800
aacgacccct tgtctttgt ctctaacct aatctctgg cttgccctt cctcttgya 1860
gaactttgaa aagctgggtt gagttagct atcagcacag cttctctgg ggaacttga 1920
gggtcccca cgaagccag aagggggaa agggactgg ccgagggag gatctgtgt 1980
gcttggaga gctggccty gttggcctt ccacgcctc tactctctt ggttggact 2040
gcccggggag agtccgggg agtggcttg ggtctcgac gtcacggag aataaagag 2100
tttactctg gaaaaaaa aaa 2123

```

```

<210> 22
<211> 2893
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> IncyLe ID No: 1872095CB1

```

```

<400> 22
atgcatcctt tgaaccttct gttagcattgg caagccttgt gcagcatatt cctcttcaga 60
tgaatacagt tctcatcagg agccttacta cggatccaaa tgtaaaagat gcaagatga 120
ccccagcctt ttgcwaaag attgactggc tctcctggcc attgctctag catgtggata 180
catgggtaat tgcactctg aaagactgg cagctgtcca gaggttact atttctgat 240
atgttacttt gctgaaaata gnaactggtt ttaattgact ttggtttctt ctgttgagac 300
ctggtactct tgcagtctt tctcaalgc tgcctagctt tcaagattct ccagagggct 360
tccatttgal tgttctctat gttgttaatt tggttctatc ttcaaaaat gatggtctgc 420
cttcaagtac agccttctta gtacaattaa cagaattgat acactgtatg atgtaccat 480
atcttggalt tccagatctc latgaaccta ttctggagc aataaaggat tttcctaagc 540
ccatgttaga gaagattaaq ttaattctca atcaagtgyc ctggactctt aaatccatc 600
ctttggcttc ttgctgtctt agactttctg gaaactcga aactgggaaa actggtctb 660
ttaacctagg aaatcacatg tatatgaaca gtttataca agcctgttt atggcccaag 720
atctcagggg acaagtatca tcttbaatc taaatgggtg caattcatta atgaaaana 780
tacagatctt tttgctctt ctggcccata cacagagggg agcatargca cctcggatc 840
tctttgagcc ttccagacct ccatggttta ctcccagac acagcaagac tgttctgat 900
acctcagatt tctcttgac aggcctcatg aagaagaaa gatcttgaat gttcaagcct 960
cacacaagcc ttctgaatt ctggatgca gtgaacttc tttaacaggaa gttagctagta 1020
aagcagcagt actgaagag accctctgta caagtgaagc tgaagaact ctatagaaa 1080
aaatgtttg aggaaaacta cgaactcaca tacyttgttt gaactggag agtaacctac 1140
aaaaagtggg agcctttaca gatctttgoc ttgctcttgc tctctctct tctttgaaa 1200
caatgtctgt ccagatcca gcatctcac ccagtataca agatgggtgt ctaatgcaq 1260
cctctgtacc cgtctctca gaagaaccag tagtttataa tccaacaaca gctgccttca 1320
ctgtgactc actgtgaat gaaaaacca tagggcagtc tctaatgag tttactggt 1380
ctgaaaaaac ttctgtcct aacgaatca acaagattct tgttaataaa gatgtacct 1440
agaaaccag aggtgaaacc acacctcaq taactgactt actaaattat tctttggctc 1500
cagagatct taactggtat aaccaatct atgttgaaaa ctgpcctct ctgcaaatg 1560
ctgagaaaac tatgcaatc accgaggaac ctgaaactt taicttact ctctgagat 1620
ttcatalgaa tgggaagtat catgtgagaa ggaanaattt ahaaatgta tcaactgca 1680
tggttttgga gttgocagtt aaaagaatta ctctctctc ttcattgtca gaaagtgtg 1740
ctgtagatgt gacttcaact gatcttagt agaaccttgc taaaaatta aagccttca 1800
ggactgatga agctctctg acaaaattgg tgcctatct attaagttcc gttgtggttc 1860

```

WO 01/46443

PCT/US00/34811

```

acotctgtat atctctlgaa agtgggcatt actattctta tggcagaat atcaacaagta 1920
cagactcttc atatcagatg taaccaccagt ctgaggtctct ggcatttagca tctccccaga 1980
gtcatttact agggagagat agtcccagtg cagtttttga acaggatttg gaaataaagg 2040
aaatgtcaaa agaatgglll ltattttaatg acagttagagt gcaatttact tcaattcagt 2100
caglccagaa aattacgggc agytttccaa aggacacagc tlatgtgctt ttgtataaaa 2160
aacngatag taetaatagt ttaagtgtta abaccocac cagtggactc tggataaaig 2220
gagocccacc tctacagaaa gaactatagg algtataaac aaaaagcaat aaactatc 2280
tacaggaaea agagttagat gclcyagccc gggcootoca agotgcactc gcllcaatll 2340
catttcggcc caatggalll gatgaacaag ccccccaggy aagctgtgga ccaattggtg 2400
gaggggttg agaggalll aalacagtgg gcaagactgt atttgaacc tggagaggtc 2460
caaaatgcaac tggtcacgaa acgttotaata ctatgactgt taaaatgta gactataaca 2520
aaatctate ttttattttt cattagaccc ttataactca agagaacaca ctcagtgtt 2580
gtttttattt tcttgacaca lttattaaca aaatgcatac tggaaaaaaa aatctacctc 2640
tcaaatctcc attttgtttt abggttagac abgcttgaac aaaaatgttc agaaagaaa 2700
atgtaccagg tcccraatta agotgggtta aatttggtag aagatttaa atgtctatc 2760
ttagctllia ctgaacaaaa aatgttaattt abttagcatt ctttataaaa gaattgagtc 2820
tagaggtaaa aaaaaatct tgtttttaa aaatctctta cgtcttgtgt aattaccctc 2880
attattaat tca 2893

```

```

<210> 23
<211> 4170
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<223> Incyte ID No: 2278688CE1

```

```

<400> 23
gctcccngg tgcctctctt ccggcgtctg ccgcgcctcg ytgatgtgg ctgcagctg 60
ccgcccctc cctcgtctct tccctggccc tccgctctct cccctctctc 120
ttctctctt cagccctctc tctcgcgcc gccctccacg cctgggctc gcccgatgt 180
cggagagag gccctctgag cgcctgcctg ccgatgtct ccccatcaac tgcagccttt 240
gccccaagcc cgaactgctg gacttccctc tccagggcaa gctggagcc gccgcccagg 300
tgggcaggc gaactaatcag attgtgatga attgtctga laltgatatt attacagctt 360
cattctcacc agaagagat gaagaatcac atgtcacagg atttaactat cagaatgaa 420
atgaaaagt caccctgtct tcccttagta ctctgcaaac aggtacggga accttaaga 480
tagttttgt tggagcgtg aatgcaaaaa tgaaggttt ctatagaagt aatatacta 540
cccctcttg agggctggc tatgtctgt taacacagtt tgggctact gatgccgaa 600
ggcctttcc tgcctggal gagcctgcta tcaagcacc tttgatctc tcaattggtt 660
tctcaaga oagatagct ltaaccaaca tgaatgtaat tgaccgaaa ccatacctg 720
atgataaaa tttagtgaa gtyaagttg ccgcacacc tgttatgtc acatatctgg 780
tggcattgt tgtgggtgaa tatgactttg tagaacaag gtcacaagat ggtgtgtgt 840
tccgttttta caactcgtt ggcacaagcag aacaagaaa alttgcctta gaggctgtc 900
claaaccct gactttctc aaggaactct tcaatgttcc ctatctctc cctaaaaatg 960
atctcaatg cattgagac ttgtcagctg gtgcaatgga gaactggggc ctgttactt 1020
ataggcagc tgcattgctt atgtaccaa aaaaatccctg tctctctcc cgcagctgg 1080
tgcctctgt tctgggcent gaaactgccc atcaatgctt tggaaatctt gtaactatg 1140
aatggtgac tcaactctg ttaaatgaa gttttgcac ctgyattgaa tctctgtgt 1200
tagaccactg ctctccagag tatgatattt ggactcagtt tglctctgt gattacacc 1260
gtcccacaga gottgacgcc ttagataaca gccatcttat tgaagtcaat gtggccatc 1320
cactgaggt tcatgagata lttgatgcta tatcatatag caaaggtgca tctgcatcc 1380
gaaigtgca tgaactacat ggggataag actttagaa aggaalgaa atgtatctan 1440
caagttcca caaaaagat gctgcaacag agatctctg gaaaagtta gaaaagctc 1500
gtgtaaac tatagact gtgatgata cctggaccaa caaatggga tttccctca 1560
tttatgtga acctgaacag ctagaagatg acagattatt gaggctgtc caaagaagt 1620
tctgtctgg tgggtcatat gttgtgaa gtttgcacca gtggatggtc cctatcaaa 1680
tctctactg tgaagaccoc aaccaggcca aactaaaaat tetaatggac aagccagaga 1740
tgaatlggt ttgaaaaat gtaaacccag ccaaatgggt gaagttaaac ttgagaacag 1800
ttgggttta tgggcccag tacagctctg ccaatgctga aagtttata ccagcatcc 1860
gggaccttc tctgcccct gtgatctgac ttgatatac gaatgacctc tctctctgt 1920
ctgagctgg atctacact actgttagag ttctaaagt caltggagct tttgtgaa 1980
agcccacta taactgtatg agagactcga cctgtaacct gggatctctc tcaactctc 2040
tgtcccac acactctat gaggaatcc aggagtctt gaaagtgtc ttttcaacta 2100
tagggagag actgggctgg gaocccaaac ctgggaagg tcatctgat gcaactcgt 2160
ggggcttgt tctggaaaa ctaggaaaag caggacataa ggcaacttta gaagaagcc 2220
gtcgtcgtt laaggaccac gtggaaggaa aacagattct ctccgtgat ctgagagtc 2280

```

WO 01/46443

PCT/US00/34811

```

ctgtctatct gactgttttg aagcatggtg atggcaactac tttagatatt atgttaaac 2340
ttcataaaca agcagatag caagaagaga aaaacgaaat cgaagagatc ctltgycota 2400
ctcttttggc tgacctgatt caaaaaglcc tcaaglttgc acttlcagaa gaggtagctc 2450
cacaggacac tgtatcggtg atltggtygg tagctggagg cagcaagcat ggtaggaaag 2520
ctgtctggaa atctataaag gacaactggg aagaacttta taaccyatac cagggaggtc 2580
tctcaatato cagactaata aagctatacag ttgagggatt tgcagttagt aaaaaggtg 2640
gagaggttaa ggtttcttcc gagagtcacc cagctcccct agotgagcgt accatccagc 2700
aggtttgtga aaatactctg ctgaaatgcty cctggctaaa ccgagatgct gegagcater 2760
accagtaoct cottoacggy auggcctcac caccacagat gtaaatctct aggttcycgc 2820
atggcggttt ctgctcgctc gctgcagga taaggtggag ctaccgaaac gctgattcat 2880
atgccaaqaa lttggagtct tctttcaaac cagtgggggt tggacaatga atgagttaa 2940
ctggttctgt ctacactcc agaaltaaat tclaltgaaa aaggaatac agcaaltcag 3000
caaaaaataa atonaaata aanaatgaaa tctgtagtga ataaataga gcatacagaa 3060
actgtgaac ttctggaagc ttctgtagty gtbaaaagta ttcaactcct taotgttaat 3120
gacaggtgtt ctgtttttat accctacca aaggaacta gaggcttctt ggtgaagac 3180
atttttttga agtgggttct gaaaggagcc tataaagcca aggttggtgt ccatlctgg 3240
gaatggttaa acacaaaagg ctgatagcty gtatcacata gttggagtea gtgcataat 3300
ccaaagctct ttttttttt ttggcagcgg gactgatcag gaagataat tcclgcataa 3360
ctcaactcga accaaggttt gtagttagt tttctcctt gccttcctct ctgtctgacc 3420
gacctcttgg caaaaaaaa aaacaaaag caaaaaaaa aaactaccc tgttctggtt 3480
ttttctctcc ctltagttcc accccaacc cccaltccct ggtgtccttc tttagagatg 3540
agaaataata aggaactcct ttctcagcc acaktaaaat agagaacty atatacila 3600
tttttttttt tttaagatg actbataaga accctgaaat ttatataggt gagaataag 3660
aaataaaaag atctccagcc aggccttctt gaaggagtta ttctctaaa aatgtctta 3720
gttgtctgna aagccagctc ttgaactctt tcaaacagat atcaacacty gcttctcccg 3780
gttcaattca tggctgagag agtctagttg taactyctgc agggcttaa acattagtyg 3840
taactgtctt aaaaaacnaa gactgttaag ctgtgtgtgc cactgtttgc ttaacagta 3900
tatctacta ataaagcctca cttatttaat ccaatgagtt taaacttaa atctaltcc 3960
cttctctctt cctacacttt tttttcttt ttcttaaaa aaataktttg tgtttatac 4020
agaaattcat atttgggtg gctlaacggt atttcaagag gtaicagat tgtgagacty 4080
cttctctgna acatttttct gotattgttt taaaaaata attaaaaac agttgycgtt 4140
aataaaaatg tcaatgtgaa aaaaaaaa 4170

```

<210> 24
 <211> /67
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4042361CB1

```

<400> 24
ccagaggctt gcaagcggac agctgtctgg ccagctgcga cagaagcagg tccgagggac 60
agcccgatcc ggcctggag cagcttagga aggcgtggt agtgacgga ttlygscctt 120
ttggggnaaa caecctgaa cgcagctgga ttgcagttca gggctagaa aagctagggc 180
ttggcagcag cgtggacctg catgtgtacg agatctcggg ttggtacca acgctccaga 240
gactcactcc cgcctgtgg gagaagcaca gtcacagcct ggttggtcat gtcgggtgt 300
agggcattgc gaccacagtc acactggaga aatgtgaca caacaagggc tacaaggggc 360
tggacaactg cgtcttttgc cccggtctcc agtgcctggt ggagagcggg cctgaaagca 420
ttgaactcat caicagactg gatgtgtgt gcaagcagat caccacgttg ggcctgatg 480
tgtcgtgtac catctctcag gatgcgggca ggaaaaaacc ctltcctgcc aaaggtgact 540
gtgtttctg ccgccaaggg agggcgggt cctccagggc tcaagtgagg ctctccctga 600
ccccggcctt agaaactttg ccagtgcctt ttctgaaact cctgtgtccc gggccccc 660
ggcggagag patatgcgg atctctctg gggctgggct ctaggagacc ccaaatbtga 720
caccacagaa agcaataaaa acacttgaaa taccgaaaa aaaaaaa 767

```

<210> 25
 <211> 1538
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3937952CB1

<400> 25

WO 01/46443

PCT/US00/34811

```

ggtagatggg agccatggg lggatgagaa gccatggcag agycttttcc tgcataccctc 60
ctcagtttcc ctattccag atgcccggct cccctgtctac ctgtatgact ttgagacca 120
cgctcgtgga ataategta aaccocgca tgatggggaa gaccatgggg atgagatga 180
cttctctt ttggggccctc tggccacag tgcacaaggc cccctgata ccccaactgg 240
gtgtccagtc tcccactct ggatgagjac ccacccctcc atttggctgg cacagggagc 300
tcccagttc ctactctgtt atgtctctcc aattgaaagt ctctgtctcc ggagcagta 360
gagcagcag gagtaggtg gtaggtcagt gtcccctgt ctgtccgaaa tcccactcc 420
catctgccc ccaggtcttt ccattggtta ggagaaagca cttagcctcc agatgatga 480
atactggcc aactttgcc gccaccgaaa ccccaatgat gggaaatgco cctgctggcc 540
agctcaaac aagatgaaa agtaccctga gctggatttt accacaagag gggcagta 600
gtccaaaggag aagaagatgg ctttttggat gactctgtac cagtctcaa gactgagaa 660
gcagaggcaa tctaaaggtt ggtatgtcag gaagagaca aagaggggtt tgcaccacc 720
atccagccc lggggagact agccatggac atacctgggg acaagagttc tcccaccacc 780
agtttagaac tgcaggagct cctctctgcc tccaggccaa agctagagct ctltgcalgt 840
gtgtgggac tgaactggcc ttccagcctt gacatcccat gatgcccctc taactcactg 900
ttgacatcca gttaggccag gccctgtcaa caaccaactg tctcaagctc tccagcctca 960
ggcaaacctc ttttttccc ttctccaat cctcccacc tccaatgtct ccttgtgact 1020
ccttctatg gtaggtctgac ccagactgcc actgcccctg tcaatgacc cagctggca 1080
ttaccatcc atccctgcca acctgtgtcc tctctgttca cattggcctg gagcctagg 1140
gcaggttgt acatggagca aacttttggg aglttgggat cttctctccc acccaactt 1200
atctcccaca gggccactcc aaagtctata cacaggggtg gtctcttcaa taaaagatg 1260
ttgttagaac ctgactttct ccaactctaa aatgggtctg tgaag-gaat gatgtctaa 1320
ttgagacctt gaaagaaagg aagtatgtct gctgttctt tagtggctg tgcctggatg 1380
ctaacctcag tcaaggggtg ctactgcmaa gtgtcctctg atcaaaaaca cttgctttg 1440
gccctctatg gtctcaagt caccctcag gacagccaca ccccagctc acttgtccat 1500
cagtttaggt cttagtgcca catctagat cctctgctc 1538

```

```

<210> 26
<211> 1497
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<223> Inocyte ID No: 7257324CB1

```

```

<400> 26
ggccttactc ttccaagagg ccatggaagt ataataata aagcaagaaa ggcagatgca 60
tttgggtggc tccgtggact tctgaatgta ctgtgagtat gagaccttcc cttccaaaag 120
atccgtgtct tctgtctat tccacacgaa gcttgcctca gatcagggga ggaatgtaga 180
ctgtccacag gtctactact caacaggata ttctccaagg aaaaatgacc ccaactctag 240
cctggccatt tttctggoty ttctctcac ggtgaaaggt cttctaaagg ctagctcttc 300
accaagaaat tataaagctt tgagcaggtt ccaaggaatgg aagcaaaagga tggagagcaa 360
ggagcttga aggcagaaca tggacttagg ctlttaagctg ctcaagaagc tggcctttta 420
caaccctggc aggaacatct tctatctccc cttyagactc tctaacgctt tctccatgct 480
gtgctgggt gccacagaca gcaacctgga ctgagatcag caggggtcca acttccagaa 540
gatgccaaga aagactcttc atgagggctt ccaatacctc atccacgagc tgaccacgaa 600
gaccacggac ctcaaaactga gcatctggaa cacgctgttc attgaccaga gctctgaccc 660
scagcctag tttttggag atgccaagaa cttttacagt gccgaaacca tcccttaccaa 720
ctttcagaat ttggaatgg ctcaagaaca gatcaatgac ttatcagtc aaaaaacca 780
tgggaaatatt aacaacctga tggaaatatt agaccocggc actgtgatgc ttcttgcaa 840
ttatatttct tttcgagcca ggtggaaca tgagtttgat ccaaatgtaa ctaagagga 900
agatltctt cttgagaaaa acagttcagt caaggttccc atgatgttcc ctagtggcat 960
atcccaaggt ggtatgacg ataagctctc ttgcccactc ctgaaatac cctaccagaa 1020
aaatctcaca gccacttcc tcttctctga tgaagggcag ctgaaagcact tggagaggg 1080
atctcagctg gacactttct ccagatggaa acaattactg tcaagcaggg cctgtagact 1140
gtctgtctcc agactccaca tgacgggac cttcgactg aagaagactc tctctcaact 1200
agdygtctcc aaaaatcttg aggaacatgy tgatctcacc aagatcpcct ctcatctcag 1260
ctgaaagtgt ggcagggctg tgcacagcc gtagctgaag atggatgaga ggggtacgga 1320
agggccctt ggcaccggag cacagactct gccatggag acaccactcg tctcaagat 1380
agacaaacc tatctgtctc tgalltaag gagaaaata ccttccctg tcttctctgg 1440
aaagatgttt aacctatg gaaaataaag gagaattctt gcttgcaca aaaaaaa 1497

```

```

<210> 27
<211> 1194
<212> DNA
<213> Homo sapiens

```

WO 01/46443

PCT/US00/34811

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472038CB1

<400> 27
atgcccggg ccattagtc cctgatgag ttcaaacatc cgttcagtg caagctgag 60
ctgtaccgc tcccctcgc cgcctcccc tcggccctc atgcttcga gaagttggy 120
atccgatgg accgctcgc tttaaatgac ggcgaggag tcagccatt ccgtggcag 180
tggaaactcg cggtagagc cacaccctb agcaattacc tagacccca gtactttggc 240
cccatacca ttgtagccc gcgcagaca tcaaggtya tattcgatac ggttccctc 300
aatctctggg tgcctccgc caagtgtgc tccacaatgg tggcctgtc tgtgcacaat 360
cgtactttg ccagcggtc gaccagtcac caggtgagg gagaccatt tgcatacac 420
tatggcagc gcagctgtc cggcttcct tccaccgaca ccgttcggg ggcctggcta 480
gagattcgg atcagcctt ccagcggcc ccgcaatgc cggctccat ctctctgga 540
gcasaalcc accgcatctt tggattggcc tatgcagca tctctatca ggcataaag 600
ccaccatct atgcatgat ggcgcaagga ctctcaaga aaccatatt cagtqlltac 660
cttagcagaa atggcagaaa ggaigtgga gccatctct ttggcggac caatccgat 720
tactacacc gaaactttac ttatgtccg gtgagcacc gtgcctatt gcagtgaaa 780
atggattcc cagllatcc gaatctcag ctatgtcag agggatgta zgtgattac 840
gacacgggca cctcttctc ggcattgcc tacgaccag ctatacltat caatyaatcc 900
atgggggaa ctccctctc ctttggacg ttctatgtc cgtgcgacg cgtaccagc 960
ctgcccama tcaactttac cttggtagg agtagattt tctcggagt caagagfat 1020
gtcttctgg atactacca ggatcgaag aletgctct cggccttat tgcctgmac 1080
ctgccatgc ccagtgacc gctctgatt ctgggggat tgtttttgg caaatactat 1140
actgagttc acatggagc gcatcgcatt gatttcgac atgccaggag ttga 1194

```

```

<210> 28
<211> 438
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472041CB1

<400> 28
atggggatcg galgctggag aaacccccg ctgctctga ttgcctggt cctgtcagcc 60
agctgggto acttccaaag gtgggaggc ttccagcaga agctcatgag caagaagac 120
atgaaltoa cactcaactt ctctatcan tctacacaa atgcagcaa cgcactctac 180
ttatactag tccagcgtc aattegaagt cagatgcag tagcpcagg agtgagtat 240
atagtoactg tgaagatgg ctggacaaa tgcagagga atgacaagc caattctctc 300
tgcccctgc aaagcaagaa gctgagaag agtttaatt gccagctct gatatacac 360
atgcccctga taaactatt ccagctctg acaattctc gctcggaggc cagcactgtg 420
ggcagaacc tcagatga 438

```

【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
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number
WO 01/046443 A3

- (51) International Patent Classification: C12N 15/57, 94B, A01K 6/20K, C07K 16/30, C12Q 1/68, A6, K 38/48, G0, N 3/570
- (21) International Application Number: PCT/US00/34811
- (22) International Filing Date: 19 December 2000 (19.12.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
634114535 23 December 1999 (23.12.1999) US
62077334 21 January 2000 (21.01.2000) US
60178384 28 January 2000 (28.01.2000) US
60179502 2 February 2000 (02.02.2000) US
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US], 3160 Foster Drive, Palo Alto, CA 94304 (US).
- (72) Inventors and
(75) Inventors/Applicants (for US only): YANG, Junming [CN/US], 7125 Bark Lane, San Jose, CA 95129 (US); BAUGHN, Mariah, R. [US/US], 14544 Soudinge Road, San Leandro, CA 94577 (US); BURFORD, Neil [GB/US], 105 Wildwood Circle, Durham, CT 06422 (US); AC-YOUNG, Janice [US/US], 273 Golden Eagle Lane, Brisbane, CA 94003 (US); LÜ, Dyung, Aina, M. [US/US], 234 Coy Drive, San Jose, CA 95123 (US); REDDY, Roopa [IN/US], 1233 W. McKinley Avenue, #3, Sunnyvale, CA 94086 (US); YUE, Henry [US/US], 826 Lees Avenue, Sunnyvale, CA 94087 (US); NGUYEN, Daniel, B. [US/US], 1403 Ridgewood Drive, San Jose, CA 95118 (US); TANG, Y., Tom [CN/US], 4230 Ranwick Court, San Jose, CA 95128 (US); YAO, Monique, G. [US/US], 111 Frederick Court, Mountain View, CA 94043 (US); LALL, Praveen [IN/US], 2382 Lass Drive, Santa Clara, CA 95054 (US).
- (74) Agents: HAMLET COX, Diana et al.; Incyte Genomics, Inc., 3160 Foster Drive, Palo Alto, CA 94304 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EC, 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, MY, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GI, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW); Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM); European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LI, MC, NL, PT, SE, TR); OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report
- (88) Date of publication of the international search report: 11 July 2002
- 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 01/046443 A3

(54) Title: PROTEASES

(57) Abstract: The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/34811

A. CLASSIFICATION OF SUBJECT MATTER		
IPC 7 C12N15/57 C12N9/48 A01K67/00 C07K16/40 C12Q1/68 A61K38/48 G01N33/50		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system to which classification symbols)		
IPC 7 C12N A01K C07K C12Q A61K G01N		
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)		
EPO-Internal, EMBL		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VALERO REBECA ET AL: "USP25, a novel gene encoding a deubiquitinating enzyme, is located in the gene-poor region 21q11.2" GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 62, no. 3, 15 December 1999 (1999-12-15), pages 395-405, XP002153277 ISSN: 0888-7543 the whole document --- -/-	1-7, 11-15
<input checked="" type="checkbox"/>	Further documents are listed in the continuation of box C.	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Patent family members are listed in annex.	
<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another claim or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"T" document published prior to the international filing date but after the priority date claimed</p> <p>"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principles or theory underlying the invention</p> <p>"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</p> <p>"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 combinations being obvious to a person skilled in the art</p> <p>"S" document member of the same patent family</p>		
Date of the actual completion of the international search		Date of making of the international search report
4 September 2001		09 01 2002
Name and mailing address of the ISA European Patent Office, P.O. Box 16 Patentstra. 2 NL 2000 HV Rijswijk Tel: +31 (0) 79 200 2000 Fax: +31 (0) 79 200 2016		Authorized officer ESPEN, J

1

P. 2/2 PCT/US 00/34811 (Revised) (Word) (July 1999)

INTERNATIONAL SEARCH REPORT

Inter. Int'l Application No.
PCT/US 00/34811

C. (Continued) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	GROET J ET AL: "NARROWING OF THE REGIONS OF ALLELIC LOSS IN 21Q11-21 IN SQUAMOUS NON-SMALL CELL LUNG CARCINOMA AND CLONING OF A NOVEL UBIQUITIN-SPECIFIC PROTEASE GENE FROM THE DELETED SEGMENT" GENES, CHROMOSOMES & CANCER, XX, XX, vol. 27, no. 2, February 2000 (2000-02), pages 153-161, XP000943091 the whole document	1-7, 11-15
X	WO 99 46289 A (HUMAN GENOME SCIENCES INC ;NI JIAN (US); ROSEN CRAIG A (US); FERRI) 16 September 1999 (1999-09-16) SEQ ID NO 130 claims 1-23	1-7, 9-18, 25, 26
E	WO 00 78934 A (GROET JUERGEN ;NIZETIC DEAN (GB); UNIV LONDON PHARMACY (GB)) 28 December 2000 (2000-12-28) SEQ ID NOS 1,6 claims 1-35	1-18
E	WO 00 79267 A (GROET JUERGEN ;NIZETIC DEAN (GB); UNIV LONDON PHARMACY (GB)) 28 December 2000 (2000-12-28) SEQ ID NO 1,6 claims 1-11	1-7, 9, 11-18
E	WO 01 21654 A (RIGEL PHARMACEUTICALS INC ;LUO YING (US); MANCEBO HELENA (US)) 29 March 2001 (2001-03-29) SEQ ID NO 1 claims 1-22; figure 2	1-7, 9-14, 25-27
X	DATABASE EMBL [Online] Database entry HS1271038; Accession no AA479313, 23 June 1997 (1997-06-23) HILLIER L ET AL: "Homo sapiens cDNA clone" XP002176711 sequence	11, 12
X	DATABASE EMBL [Online] Database entry HS1272487, Accession no AA482201, 24 June 1997 (1997-06-24) NATIONAL CANCER INSTITUTE, CANCER GENOME ANATOMY PROJECT (CGAP): "Homo sapiens cDNA clone similar to ubiquitin carboxyl-terminal hydrolase 2" XP002176712 sequence	11, 12

	-/--	

1

NOTE: Where a document is cited as being relevant, it is so cited because it is considered to be relevant to the claims of the invention.

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 00/34811

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT	
Category	Relevant to claim No.
X	11,12
DATABASE EMBL [Online] Database entry HSZ12939, Accession no AA307805, 18 April 1997 (1997-04-18) ADAMS MD ET AL: "EST178661 Colon carcinoma (HCC) cell line Homo sapiens cDNA" XP002176713 sequence -----	

1

Form PCT/ISA/210 (continuation of second sheet) July 1993

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 00/34811
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 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.: 20,21,23,24 because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that a 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 6.4(a).	
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows:		
see additional sheet		
1.	<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 checked="" 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.: In part 1-28	
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 00/34811

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 20,21,23,24

Claim 20 refers to an agonist of a polypeptide of claim 1 and claim 23 refers to an antagonist of a polypeptide of claim 1 without giving a true technical characterization. Moreover, no such specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is in fact, a mere recitation of the result to be achieved. The above comments also applies for claims 21 and 24.

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 Application No. PCT/US 00/34811

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International Application No.
PC1/US 00/34811

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9946289 A	16-09-1999	AU 3006799 A	27-09-1999
		EP 1062236 A1	27-12-2000
		WO 9946289 A1	16-09-1999
		EP 1044210 A1	18-10-2000
		WO 9931116 A1	24-06-1999
		US 2001024813 A1	27-09-2001
		-----	-----
WO 0078934 A	28-12-2000	AU 5550800 A	09-01-2001
		AU 5552400 A	09-01-2001
		WO 0078934 A2	28-12-2000
		WO 0079267 A2	28-12-2000
-----	-----	-----	-----
WO 0079267 A	28-12-2000	AU 5550800 A	09-01-2001
		AU 5552400 A	09-01-2001
		WO 0078934 A2	28-12-2000
		WO 0079267 A2	28-12-2000
-----	-----	-----	-----
WO 0121654 A	29-03-2001	AU 7832800 A	24-04-2001
		WO 0121654 A2	29-03-2001
-----	-----	-----	-----

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
A 6 1 P 1/16	A 6 1 P 1/18	4 B 0 6 5
A 6 1 P 1/18	A 6 1 P 5/02	4 C 0 8 4
A 6 1 P 5/02	A 6 1 P 5/14	4 H 0 4 5
A 6 1 P 5/14	A 6 1 P 7/06	
A 6 1 P 7/06	A 6 1 P 9/00	
A 6 1 P 9/00	A 6 1 P 9/10	
A 6 1 P 9/10	A 6 1 P 9/10	1 0 1
A 6 1 P 9/12	A 6 1 P 9/10	1 0 3
A 6 1 P 11/00	A 6 1 P 9/12	
A 6 1 P 11/06	A 6 1 P 11/00	
A 6 1 P 15/00	A 6 1 P 11/06	
A 6 1 P 17/00	A 6 1 P 15/00	
A 6 1 P 17/06	A 6 1 P 17/00	
A 6 1 P 19/02	A 6 1 P 17/06	
A 6 1 P 19/10	A 6 1 P 19/02	
A 6 1 P 25/00	A 6 1 P 19/10	
A 6 1 P 25/14	A 6 1 P 25/00	
A 6 1 P 25/16	A 6 1 P 25/14	
A 6 1 P 25/18	A 6 1 P 25/16	
A 6 1 P 25/28	A 6 1 P 25/18	
A 6 1 P 29/00	A 6 1 P 25/28	
A 6 1 P 31/00	A 6 1 P 29/00	
A 6 1 P 31/04	A 6 1 P 29/00	1 0 1
A 6 1 P 31/12	A 6 1 P 31/00	
A 6 1 P 33/00	A 6 1 P 31/04	
A 6 1 P 35/00	A 6 1 P 31/12	
A 6 1 P 35/02	A 6 1 P 33/00	
A 6 1 P 37/00	A 6 1 P 35/00	
A 6 1 P 37/06	A 6 1 P 35/02	
A 6 1 P 37/08	A 6 1 P 37/00	
A 6 1 P 43/00	A 6 1 P 37/06	
C 0 7 K 16/40	A 6 1 P 37/08	
C 1 2 N 1/15	A 6 1 P 43/00	1 0 5
C 1 2 N 1/19	A 6 1 P 43/00	1 1 1
C 1 2 N 1/21	C 0 7 K 16/40	
C 1 2 N 5/10	C 1 2 N 1/15	
C 1 2 N 9/64	C 1 2 N 1/19	
C 1 2 P 21/02	C 1 2 N 1/21	
C 1 2 Q 1/68	C 1 2 N 9/64	Z
G 0 1 N 33/15	C 1 2 P 21/02	C
G 0 1 N 33/50	C 1 2 Q 1/68	A
G 0 1 N 33/53	G 0 1 N 33/15	Z
G 0 1 N 33/566	G 0 1 N 33/50	Z
	G 0 1 N 33/53	M
	G 0 1 N 33/566	
	C 1 2 N 5/00	A
	A 6 1 K 37/547	

(31)優先権主張番号 60/179,903

(32)優先日 平成12年2月2日(2000.2.2)

(33)優先権主張国 米国(US)

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CR,CU,CZ,DE,DK,DM,DZ,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,RO,RU,SD,SE,S G,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 バーフォード、ニール

アメリカ合衆国コネチカット州06422・ダラム・ワイルドウッドサークル 105

(72)発明者 オウ・ヤング、ジャニス

アメリカ合衆国カリフォルニア州94005・プリズベン・ゴールデンイーグルレーン 233

(72)発明者 リュ、デュング・アイナ・エム

アメリカ合衆国カリフォルニア州95123・サンノゼ・コイドライブ 233

(72)発明者 レディ、ルーパ

アメリカ合衆国カリフォルニア州94086・サニーベイル・#3・ウェストマッキンレーアベニュー 1233

(72)発明者 ユエ、ヘンリー

アメリカ合衆国カリフォルニア州94087・サニーベイル・ルイスアベニュー 826

(72)発明者 ニュエン、ダニエル・ピー

アメリカ合衆国カリフォルニア州95118・サンノゼ・リッジウッドドライブ 1403

(72)発明者 タング、ワイ・トム

アメリカ合衆国カリフォルニア州95118・サンノゼ・ランウィックコート 4230

(72)発明者 ヤオ、モニック・ジー

アメリカ合衆国カリフォルニア州94043・マウンテンビュー・フレデリックコート 111

(72)発明者 ラル、ブリーティ

アメリカ合衆国カリフォルニア州95054・サンタクララ・ラスドライブ 2382

Fターム(参考) 2G045 AA34 AA35 DA13 DA36 FA16 FB02 FB07 FB12 GC15

4B024 AA01 AA11 BA14 BA19 BA61 CA04 CA12 DA01 DA02 DA05

DA06 DA11 EA04 HA08 HA12

4B050 CC01 CC03 CC07 DD11 LL01 LL03

4B063 QA01 QA18 QQ08 QQ36 QQ44 QQ53 QR08 QR14 QR32 QR42

QR55 QR72 QR77 QS34 QX02

4B064 AG26 CA02 CA05 CA10 CA19 CA20 CC24 DA03 DA13

4B065 AA01X AA57X AA87X AA93Y AB01 CA23 CA24 CA33 CA44 CA46

4C084 AA02 AA06 AA07 BA01 BA08 BA22 BA23 CA53 CA56 DC03

NA14 ZA012 ZA022 ZA162 ZA182 ZA262 ZA362 ZA402 ZA452 ZA552

ZA592 ZA612 ZA662 ZA682 ZA752 ZA762 ZA812 ZA892 ZA902 ZA962

ZA972 ZB052 ZB082 ZB112 ZB132 ZB152 ZB212 ZB272 ZB332 ZB352

ZB372 ZC042 ZC062

4H045 AA10 AA11 AA20 AA30 BA09 BA54 CA40 DA56 DA75 EA20

EA50 FA72 FA73 FA74

专利名称(译)	蛋白酶		
公开(公告)号	JP2004500812A	公开(公告)日	2004-01-15
申请号	JP2001546939	申请日	2000-12-19
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	ヤングジュンミング ボーグンマライアール バーフォードニール オウヤングジャニス リュデュングアイナエム レディルーパ ユエヘンリー ニュエンダニエルビー タングワイトム ヤオモニークジー ラルプリーティ		
发明人	ヤング、ジュンミング ボーグン、マライアール バーフォード、ニール オウヤング、ジャニス リュ、デュング・アイナ・エム レディ、ルーパ ユエ、ヘンリー ニュエン、ダニエル・ビー タング、ワイトム ヤオ、モニーク・ジー ラル、プリーティ		
IPC分类号	G01N33/50 A61K38/00 A61K38/48 A61K45/00 A61P1/00 A61P1/04 A61P1/16 A61P1/18 A61P5/02 A61P5/14 A61P7/06 A61P9/00 A61P9/10 A61P9/12 A61P11/00 A61P11/06 A61P15/00 A61P17/00 A61P17/06 A61P19/02 A61P19/10 A61P25/00 A61P25/14 A61P25/16 A61P25/18 A61P25/28 A61P29 /00 A61P31/00 A61P31/04 A61P31/12 A61P33/00 A61P35/00 A61P35/02 A61P37/00 A61P37/06 A61P37/08 A61P43/00 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/48 C12N9/64 C12N15/09 C12N15/57 C12P21/02 C12Q1/68 G01N33/15 G01N33/53 G01N33/566		
CPC分类号	A01K2217/075 A61K38/00 A61P1/00 A61P1/04 A61P1/16 A61P1/18 A61P5/02 A61P5/14 A61P7/06 A61P9/00 A61P9/10 A61P9/12 A61P11/00 A61P11/06 A61P15/00 A61P17/00 A61P17/06 A61P19/02 A61P19/10 A61P25/00 A61P25/14 A61P25/16 A61P25/18 A61P25/28 A61P29/00 A61P31/00 A61P31 /04 A61P31/12 A61P33/00 A61P35/00 A61P35/02 A61P37/00 A61P37/06 A61P37/08 A61P43/00 C12N9/48 C12N9/6421		
FI分类号	C12N15/00.ZNA.A A61K45/00 A61P1/00 A61P1/04 A61P1/16 A61P1/18 A61P5/02 A61P5/14 A61P7 /06 A61P9/00 A61P9/10 A61P9/10.101 A61P9/10.103 A61P9/12 A61P11/00 A61P11/06 A61P15/00 A61P17/00 A61P17/06 A61P19/02 A61P19/10 A61P25/00 A61P25/14 A61P25/16 A61P25/18 A61P25 /28 A61P29/00 A61P29/00.101 A61P31/00 A61P31/04 A61P31/12 A61P33/00 A61P35/00 A61P35/02 A61P37/00 A61P37/06 A61P37/08 A61P43/00.105 A61P43/00.111 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N9/64.Z C12P21/02.C C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.M G01N33 /566 C12N5/00.A A61K37/547		

F-TERM分类号

2G045/AA34 2G045/AA35 2G045/DA13 2G045/DA36 2G045/FA16 2G045/FB02 2G045/FB07 2G045/FB12 2G045/GC15 4B024/AA01 4B024/AA11 4B024/BA14 4B024/BA19 4B024/BA61 4B024/CA04 4B024/CA12 4B024/DA01 4B024/DA02 4B024/DA05 4B024/DA06 4B024/DA11 4B024/EA04 4B024/HA08 4B024/HA12 4B050/CC01 4B050/CC03 4B050/CC07 4B050/DD11 4B050/LL01 4B050/LL03 4B063/QA01 4B063/QA18 4B063/QQ08 4B063/QQ36 4B063/QQ44 4B063/QQ53 4B063/QR08 4B063/QR14 4B063/QR32 4B063/QR42 4B063/QR55 4B063/QR72 4B063/QR77 4B063/QS34 4B063/QX02 4B064/AG26 4B064/CA02 4B064/CA05 4B064/CA10 4B064/CA19 4B064/CA20 4B064/CC24 4B064/DA03 4B064/DA13 4B065/AA01X 4B065/AA57X 4B065/AA87X 4B065/AA93Y 4B065/AB01 4B065/CA23 4B065/CA24 4B065/CA33 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA06 4C084/AA07 4C084/BA01 4C084/BA08 4C084/BA22 4C084/BA23 4C084/CA53 4C084/CA56 4C084/DC03 4C084/NA14 4C084/ZA012 4C084/ZA022 4C084/ZA162 4C084/ZA182 4C084/ZA262 4C084/ZA362 4C084/ZA402 4C084/ZA452 4C084/ZA552 4C084/ZA592 4C084/ZA612 4C084/ZA662 4C084/ZA682 4C084/ZA752 4C084/ZA762 4C084/ZA812 4C084/ZA892 4C084/ZA902 4C084/ZA962 4C084/ZA972 4C084/ZB052 4C084/ZB082 4C084/ZB112 4C084/ZB132 4C084/ZB152 4C084/ZB212 4C084/ZB272 4C084/ZB332 4C084/ZB352 4C084/ZB372 4C084/ZC042 4C084/ZC062 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA09 4H045/BA54 4H045/CA40 4H045/DA56 4H045/DA75 4H045/EA20 4H045/EA50 4H045/FA72 4H045/FA73 4H045/FA74

優先権

60/172055 1999-12-23 US
60/177334 2000-01-21 US
60/178884 2000-01-28 US
60/179903 2000-02-02 US

外部リンク

Espacenet

摘要(译)

本发明提供了鉴定和编码PRTS的人蛋白酶 (PRTS) 和多核苷酸。 本发明还提供表达载体, 宿主细胞, 抗体, 激动剂和拮抗剂。 本发明还提供用于诊断, 治疗或预防与PRTS异常表达有关的疾病的方法。

特許2004-500812
(P2004-500812A)

(43) 公表日 平成16年1月15日(2004.1.15)

(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	2 G 0 4 5
A 6 1 K 38/48	A 6 1 K 45/00	4 B 0 2 4
A 6 1 K 45/00	A 6 1 P 1/00	4 B 0 5 0
A 6 1 P 1/00	A 6 1 P 1/04	4 B 0 6 3
A 6 1 P 1/04	A 6 1 P 1/16	4 B 0 6 4
	審査請求 未請求	予備審査請求 有 (全 229 頁) 最終頁に続く
(21) 出願番号	特願2001-546939 (P2001-546939)	(71) 出願人 301005050
(86) (22) 出願日	平成12年12月19日 (2000.12.19)	インサイト・ゲノミクス・インコーポレイテッド
(85) 翻訳文提出日	平成14年6月24日 (2002.6.24)	アメリカ合衆国カリフォルニア州94304・パロアルト・ボートドライブ 3160
(86) 国際出願番号	PCT/US2000/034811	(74) 代理人 100089266
(87) 国際公開番号	W02001/046443	弁理士 大島 剛一
(87) 国際公開日	平成13年6月28日 (2001.6.28)	ヤング、ジュンミン
(31) 優先権主張番号	60/172,055	(72) 発明者
(32) 優先日	平成11年12月23日 (1999.12.23)	アメリカ合衆国カリフォルニア州95129・サンノゼ・パークレーン 7125
(33) 優先権主張国	米国 (US)	(72) 発明者
(31) 優先権主張番号	60/177,334	ボーゲン、マライア・アール
(32) 優先日	平成12年1月21日 (2000.1.21)	アメリカ合衆国カリフォルニア州94577・サンレアンドロ・サンティアゴロード 14244
(33) 優先権主張国	米国 (US)	最終頁に続く
(31) 優先権主張番号	60/178,884	
(32) 優先日	平成12年1月28日 (2000.1.28)	
(33) 優先権主張国	米国 (US)	

(54) 【発明の名称】 プロテアーゼ