

(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2004-515217**(P2004-515217A)**

(43) 公表日 平成16年5月27日(2004.5.27)

(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 O 4 5
A 6 1 K 38/00	A 6 1 K 45/00	4 B O 2 4
A 6 1 K 45/00	A 6 1 K 49/00	A 4 B O 5 O
A 6 1 K 49/00	A 6 1 P 1/04	4 B O 6 3
A 6 1 P 1/04	A 6 1 P 29/00	4 B O 6 4
審査請求 未請求 予備審査請求 有 (全 306 頁) 最終頁に続く		

(21) 出願番号	特願2002-504617 (P2002-504617)	(71) 出願人	301005050
(86) (22) 出願日	平成13年6月13日 (2001.6.13)		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12月13日 (2002.12.13)		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/US2001/019178		
(87) 国際公開番号	W02001/098468	(74) 代理人	100089266
(87) 国際公開日	平成13年12月27日 (2001.12.27)		弁理士 大島 陽一
(31) 優先権主張番号	60/212, 336	(72) 発明者	ユエ、ヘンリー
(32) 優先日	平成12年6月16日 (2000.6.16)		アメリカ合衆国カリフォルニア州94087・サニーバイル・ルイスアベニュー 826
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/213, 955		
(32) 優先日	平成12年6月22日 (2000.6.22)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/215, 396		
(32) 優先日	平成12年6月29日 (2000.6.29)		
(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 を同定及び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、宿主細胞、抗体、アゴニスト、及びアンタゴニストを提供する。更に、本発明は、P R T S の異常な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: 21 (SEQ ID NO: 1-21) からなる群から選択されたアミノ酸を含むポリペプチドと、

(b) SEQ ID NO: 1-21 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% の同一性を有する天然のアミノ酸配列を含むポリペプチドと、

(c) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、

(d) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択される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: 1-21 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: 22-42 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: 22-42 からなる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と、

(b) SEQ ID NO: 22-42 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% の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチドと、

(c) 前記 (a) のポリヌクレオチドに相補的なポリヌクレオチドと、

(d) 前記 (b) のポリヌクレオチドに相補的なポリヌクレオチドと、

(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単離されたポリヌクレオチド。

【請求項 12】

請求項 1 1 のポリヌクレオチドの少なくとも 6 0 個の連続するヌクレオチドを含む単離されたポリヌクレオチド。

【請求項 1 3】

サンプルにおける請求項 1 1 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に相補的な配列を含む少なくとも 2 0 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に特異的にハイブリダイズする、該ステップと、

10

(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、存在する場合には必要に応じて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。

【請求項 1 4】

前記プローブが少なくとも 6 0 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 1 3 に記載の方法。

【請求項 1 5】

サンプルにおける請求項 1 1 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法を用いて、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、

20

(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、存在する場合には必要に応じて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。

【請求項 1 6】

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 1 7】

前記ポリペプチドが、S E Q I D N O : 1 - 2 1 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 1 6 の組成物。

【請求項 1 8】

30

機能的な P R T S (新規のプロテアーゼ) 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 1 6 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 1 9】

請求項 1 のポリペプチドのアゴニストとして有効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 2 0】

40

請求項 1 9 の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。

【請求項 2 1】

機能的な P R T S 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 2 0 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 2】

請求項 1 のポリペプチドのアンタゴニストとして有効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) 前記サンプルにおける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

50

とするスクリーニング方法。

【請求項 2 3】

請求項 2 2 の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。

【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、そのポリペプチド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、前記試験化合物の非存在下での請求項 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、

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を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有効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、前記化合物の非存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。

【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を、請求項 1 1 のポリヌクレオチドの少なくとも 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、

(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。

【請求項 2 9】

生体サンプルにおける P R T S の発現に関連する病態または疾患の診断テストであって、
(a) 請求項 10 の抗体と前記ポリペプチドとの結合に好適な条件下で、前記生体サンプルを請求項 10 の抗体と結合させて、抗体／ポリペプチド複合体を形成するステップと、
(b) 前記複合体を検出するステップとを含み、
前記複合体の存在が、前記生体サンプルにおける前記ポリヌクレオチドの存在と相関性を有することを特徴とする診断テスト。

【請求項 30】

前記抗体が、

- (a) キメラ抗体、
- (b) 一本鎖抗体、
- (c) F a b 断片、
- (d) F (a b ') ₂ 断片、または
- (e) ヒト化抗体であることを特徴とする請求項 10 に記載の抗体。

10

【請求項 31】

請求項 10 の抗体及び許容される賦形剤を含む組成物。

【請求項 32】

患者において、P R T S の発現に関連する病態または疾患を診断する方法であって、請求項 31 の組成物を前記患者に有効量投与することを含むことを特徴とする方法。

【請求項 33】

前記抗体が標識されていることを特徴とする請求項 31 に記載の組成物。

20

【請求項 34】

患者において、P R T S の発現に関連する病態または疾患を診断する方法であって、請求項 33 の組成物を前記患者に有効量投与することを含むことを特徴とする方法。

【請求項 35】

請求項 10 の抗体の特異性を有するポリクローナル抗体を作製する方法であって、

(a) 抗体反応が起こる条件下で、S E Q I D N O : 1 - 2 1 からなる群から選択されたアミノ酸配列またはその免疫原性断片を有するポリペプチドで動物を免疫化するステップと、

(b) 前記動物から抗体を単離するステップと、

(c) 前記単離された抗体を前記ポリペプチドでスクリーニングして、S E Q I D N O : 1 - 2 1 からなる群から選択されたアミノ酸配列を有するポリペプチドに特異的に結合するポリクローナル抗体を同定するステップとを含むことを特徴とするポリクローナル抗体作製方法。

30

【請求項 36】

請求項 35 の方法により作製された抗体。

【請求項 37】

請求項 36 の抗体及び適当な担体を含む組成物。

【請求項 38】

請求項 10 の抗体の特異性を有するモノクローナル抗体を作製する方法であって、

(a) 抗体反応が起こる条件下で、S E Q I D N O : 1 - 2 1 からなる群から選択されたアミノ酸配列またはその免疫原性断片を有するポリペプチドで動物を免疫化するステップと、

(b) 前記動物から抗体を産生する細胞を単離するステップと、

(c) 前記抗体産生細胞を不死化細胞と融合し、モノクローナル抗体産生ハイブリドーマ細胞を形成するステップと、

(d) 前記ハイブリドーマ細胞を培養するステップと、

(e) S E Q I D N O : 1 - 2 1 からなる群から選択されたアミノ酸配列を有するポリペプチドに特異的に結合する前記培養モノクローナル抗体から単離するステップとを含むことを特徴とするモノクローナル抗体作製方法。

40

【請求項 39】

50

請求項 38 の方法で作製したモノクローナル抗体。

【請求項 40】

請求項 39 の抗体及び適当な担体を含む組成物。

【請求項 41】

F a b 発現ライブラリのスクリーニングにより得られた請求項 10 の抗体。

【請求項 42】

組換え免疫グロブリンライブラリのスクリーニングにより得られた請求項 10 の抗体。

【請求項 43】

サンプルにおける S E Q I D N O : 1 - 2 1 からなる群から選択されたアミノ酸配列を有するポリペプチドを検出する方法であって、

10

(a) 請求項 10 の抗体と前記ポリペプチドとの特異的な結合が許容される条件下で、前記抗体を前記サンプルと共にインキュベートするステップと、

(b) 特異的な結合を検出するステップとを含み、

前記特異的な結合が、前記サンプルに S E Q I D N O : 1 - 2 1 からなる群から選択されたアミノ酸配列を有するポリペプチドが存在することを示唆することを特徴とする方法。

【請求項 44】

S E Q I D N O : 1 - 2 1 からなる群から選択されたアミノ酸配列を有するポリペプチドをサンプルから精製する方法であって、

20

(a) 請求項 10 の抗体と前記ポリペプチドとの特異的な結合が許容される条件下で、前記抗体を前記サンプルと共にインキュベートするステップと、

(b) 前記サンプルから前記抗体を分離して、S E Q I D N O : 1 - 2 1 からなる群から選択されたアミノ酸配列を有する精製されたポリペプチドを得るステップとを含むことを特徴とする方法。

【請求項 45】

S E Q I D N O : 1 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 46】

S E Q I D N O : 2 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 47】

S E Q I D N O : 3 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

30

【請求項 48】

S E Q I D N O : 4 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 49】

S E Q I D N O : 5 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 50】

S E Q I D N O : 6 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 51】

S E Q I D N O : 7 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 52】

S E Q I D N O : 8 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

40

【請求項 53】

S E Q I D N O : 9 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 54】

S E Q I D N O : 10 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 55】

S E Q I D N O : 11 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 56】

S E Q I D N O : 12 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 57】

S E Q I D N O : 13 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

50

【請求項 58】

SEQ ID NO: 14 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 59】

SEQ ID NO: 15 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 60】

SEQ ID NO: 16 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 61】

SEQ ID NO: 17 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 62】

SEQ ID NO: 18 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

10

【請求項 63】

SEQ ID NO: 19 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 64】

SEQ ID NO: 20 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 65】

SEQ ID NO: 21 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 66】

SEQ ID NO: 22 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 67】

20

SEQ ID NO: 23 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 68】

SEQ ID NO: 24 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 69】

SEQ ID NO: 25 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 70】

SEQ ID NO: 26 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

30

【請求項 71】

SEQ ID NO: 27 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 72】

SEQ ID NO: 28 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 73】

SEQ ID NO: 29 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

40

【請求項 74】

SEQ ID NO: 30 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 75】

SEQ ID NO: 31 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 76】

SEQ ID NO: 32 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 77】

50

SEQ ID NO: 33 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 78】

SEQ ID NO: 34 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 79】

SEQ ID NO: 35 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 80】

SEQ ID NO: 36 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。 10

【請求項 81】

SEQ ID NO: 37 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 82】

SEQ ID NO: 38 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 83】

SEQ ID NO: 39 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。 20

【請求項 84】

SEQ ID NO: 40 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 85】

SEQ ID NO: 41 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【請求項 86】

SEQ ID NO: 42 のポリヌクレオチド配列を含む請求項 11 に記載のポリヌクレオチド。

【発明の詳細な説明】 30

【0001】

(技術分野)

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ペプチド結合の加水分解、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、発生または発達障害、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影響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関与する。プロテアーゼは、調節タンパク質の半減期を変えることで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関係する。 40

【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 50

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 ω -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.)。

【0004】

10

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)。

20

【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るプロトロンビンの蛋白分解によって、3つの異なった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。

30

【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)。App1e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)。App1e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

40

50

、グランザイム、カリクレイン、および組織プラスミノーデンアクチベーターおよびウロキナーゼプラスミノーデ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)。プロリルオリゴペプチダーゼ (S9) ファミリーには、特異性が大幅に異なる真核生物及び原核生物の酵素が含まれる。ジペプチジルペプチダーゼ IV (DPP-IV) は CD26 と同一であり、ペプチドホルモンの不活性化及び T 細胞の成長の調節に関係する (Kahne ら (1999) *Int. J. Mol. Med.* 4:3-15、Mentlein, R. (1999) *Regul. Pept.* 85:9-24 を参照)。DPP-IV の阻害が II 型糖尿病の治療に提案され (Holst, J. J. and C. F. Deacon (1998) *Diabetes* 47:1663-1670)、血清の DPP-IV 活性の低下が食欲不振及び大食患者に見られた (van Wes, D. ら (2000) *Eur. Arch. Psych. Clin. Neurosci.* 250:86-92)。

【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応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)。組織プラスミノー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. および P. H. Lange (1989) *Urology* 33:11-16)。

【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ッシングに作用する。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ッ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、Clpプロテアーゼ、およびプロテアソームが含まれる。Clp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、Clp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)。

10

【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ジーン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)。

20

30

40

【0011】

システインプロテアーゼ

50

システインプロテアーゼ (C P) 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 C P の約半数がウイルスのみに存在する。C P 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 C P ファミリーには、パパイン様酵素 (C 1) およびカルパイン (C 2) 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 (K a r r e r, 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 (l o b e) の間に配置された触媒部位を備えた2ローブ分子 (b i l o b e d m o l e c u l e) 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 (R a w l i n g s, N. D. および A. J. B a r r e t t (1994) *Meth. Enzymol.* 244:461-486を参照)。

10

【0012】

或る種の C P は至る所で発現されるが、別の種類の C P は免疫系の細胞によってのみ発現される。注目すべきは、C P 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関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 (K e y s z e r, G. M. (1995) *Arthritis Rheum.* 38:976-984)。

20

【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 E F 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 (C h a n, S. L. および M. P. M a t t s o n (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。A L S、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 (C h a n および M a t t s o n, 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 (M c C r a c k e n, 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 2 A 型に関係する (M i n a m i, N. 他 (1999) *J. Neurol. Sci.* 171:31-37)。

30

40

【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

50

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、小さなスペーサー領域によって分離された1つの大きなサブユニット (p 20) および1つの小さなサブユニット (p 10)、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D 297) が切断され、スペーサーおよびプロドメインが除去され、p 10 / p 20 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; Salvesson, 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増大し、細胞死の減少が癌に関連することが報告された (Salvesson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pol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pH 2-3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個の残基から成る第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。

40

【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

50

び血圧の調節を担うホルモンであるアンギオテンシンのプロセシングの第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)。

【0018】

メタロプロテアーゼ

メタロプロテアーゼは、活性のために通常はマンガン若しくは亜鉛のような金属イオンを必要とする。マンガン金属酵素の例には、アミノペプチダーゼP及びヒトプロリンジペプチダーゼ (PEPD) がある。アミノペプチダーゼPは、様々な炎症反応で活性化されたノナペプチド、即ちブラジキニンを分解しうる。アミノペプチダーゼPは、冠状虚血／再灌流障害に関係する。アミノペプチダーゼP阻害剤をラットに投与にすると、心臓保護効果が見られた (Ersahin, C. ら (1999) J. Cardiovasc. Pharmacol. 34: 604-611)。

【0019】

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met 20
zincinsとして知られ、アミノペプチダーゼB及びN、アンギ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。

【0020】

アンギ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、高いアミノペプチダーゼB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 30
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し得る (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)。

【0021】

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 40
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 50

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)。V I I C型エーラーズ-ダンロス症候群は、プロコラーゲンI N-プロティナーゼ遺伝子の変異によって引き起こされる(Colige, A. ら (1999) Am. J. Hum. Genet. 65:308-317)。

【0022】

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; Tarabolletti, 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)。

【0023】

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。

【0024】

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) Na

ture 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。

【0025】

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ンドメイン(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)。

【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、ペプチド結合の加水分解、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、発生または発達障害、上皮の疾患、神経の疾患、および生殖障害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影響についての評価にも有用である。

における外来性化合物の影響についての評価にも有用である。

【0027】

(発明の要約)

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」、「PRTS-14」、「PRTS-15」、「PRTS-16」、「PRTS-17」、「PRTS-18」、「PRTS-19」、「PRTS-20」、及び「PRTS-21」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: 1-21からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b) SEQ ID NO: 1-21からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c) SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d) SEQ ID NO: 1-21とからなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された単離されたポリペプチドを提供する。別法では、SEQ ID NO: 1-21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: 1-21からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b) SEQ ID NO: 1-21からなる群から選択された

アミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された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: 1-21からなる群から選択されたポリペプチドをコードする。更なる別法では、このポリヌクレオチドは、SEQ ID NO: 22-42からなる群から選択される。

【0029】

更に、本発明は、(a)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b)SEQ ID NO: 1-21からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された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: 1-21からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b)SEQ ID NO: 1-21からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された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: 1-21からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b)SEQ ID NO: 1-21からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された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。

【0032】

更に、本発明は、(a)SEQ ID NO: 22-42からなる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と、(b)SEQ ID NO: 22-42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%の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチドと、(c)前記(a)のポリヌクレオチドに相補的なポリヌクレオチドと、(d)前記(b)のポリヌクレオチドに相補的なポリヌクレオチドと、(e)前記(a)-(d)のRNA等価物とで構成される群から選択された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60個の連続するヌクレオチドを含む。

【0033】

更に本発明は、(a)SEQ ID NO: 22-42からなる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と、(b)SEQ ID NO: 22-42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%の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチドと、(c)前記(a)のポリヌクレオチドに相補的なポリヌクレオチドと、(d)前記(b)のポリヌクレオチドに相補的なポリヌクレオチドと、(e)前記(a)-(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有するサンプルにおける標的ポリヌクレオチドを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含む少なくとも20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、存在する場合には必要に応じて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60個の連続するヌクレオチドを含む。

10

20

30

40

50

【0034】

更に本発明は、(a)SEQ ID NO: 22-42からなる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と、(b)SEQ ID NO: 22-42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%の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチドと、(c)前記(a)のポリヌクレオチドに相補的なポリヌクレオチドと、(d)前記(b)のポリヌクレオチドに相補的なポリヌクレオチドと、(e)前記(a)-(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有するサンプルにおける標的ポリヌクレオチドを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法を用いて、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、存在する場合には必要に応じてその収量を測定するステップとを含む。

【0035】

更に本発明は、(a)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b)SEQ ID NO: 1-21からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された有効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、SEQ ID NO: 1-21からなる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。本発明は更に、患者にこの組成物を投与することを含む、機能的PRTSの過剰な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。

【0036】

更に本発明は、(a)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b)SEQ ID NO: 1-21からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d)SEQ ID NO: 1-21からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択されたポリペプチドのアゴニストとして有効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PRTSの過剰な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。

【0037】

更に、本発明は、(a) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択されたポリペプチドのアンタゴニストとして有効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PRTS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。

10

【0038】

更に本発明は、(a) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された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つの試験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。

20

【0039】

更に本発明は、(a) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を含むポリペプチドと、(b) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸と少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチドと、(c) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d) SEQ ID NO: 1-21 からなる群から選択されたアミノ酸配列を有するポリペプチドの免疫原性断片とで構成される群から選択された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、この試験化合物の非存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。

30

【0040】

更に本発明は、SEQ ID NO: 22-42 からなる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有効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。

40

【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

50

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が、(1) SEQ ID NO: 22-42 からなる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と、(2) SEQ ID NO: 22-42 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%の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチドと、(3) 前記(1)に相補的な配列を有するポリヌクレオチドと、(4) 前記(2)のポリヌクレオチドに相補的なポリヌクレオチドと、(5) 前記(1)-(4)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の少なくとも20個の連続する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: 22-42 からなる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と、(2) SEQ ID NO: 22-42 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%の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチドと、(3) 前記(1)のポリヌクレオチドに相補的なポリヌクレオチドと、(4) 前記(2)のポリヌクレオチドに相補的なポリヌクレオチドと、(5) 前記(1)-(4)のRNA等価物とで構成される群から選択される。別法では、標的ポリヌクレオチドは、その標的ポリヌクレオチドの断片である。

10

20

30

40

50

【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。従来の特許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「PRTS」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PRTSのアミノ酸配列を指す。

【0046】

用語「アゴニスト」は、PRTSの生物学的活性を強める、或いは模倣する分子を指す。このアゴニストは、PRTSに直接相互作用するか、或いはPRTSが関与する生物学的経路の成分と作用して、PRTS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、PRTS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。

【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。

【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。

【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F a b及びF（a b'）₂、及びそれらの断片、F v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。

【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。

10

【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。

20

【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。

30

【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。伸長及び重複の両方によって構築されるコンセンサス配列もある。

40

【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、 β シートや α 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。

【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。

30

【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】

「異なる発現」とは、少なくとも2つの異なったサンプルの比較により決められる遺伝子またはタンパク質の発現の上昇即ちアップレギュレーションや、減少即ちダウンレギュレーション、または発現しないことを指す。このような比較は、例えば、処置したサンプルと未処置のサンプルとの間、または病変サンプルと正常なサンプルとの間で行う。

40

【0064】

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、60、75、100、150、250若しくは500個の連続するヌクレオチド

50

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500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。

【0065】

SEQ ID NO: 22-42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: 22-42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: 22-42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: 22-42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: 22-42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

【0066】

SEQ ID NO: 1-21のある断片は、SEQ ID NO: 22-42のある断片によってコードされる。SEQ ID NO: 1-21のある断片は、SEQ ID NO: 1-21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: 1-21のある断片は、SEQ ID NO: 1-21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: 1-21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

【0067】

「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。

【0068】

「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。

【0069】

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

【0070】

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、Higgins, 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ンメントの場合、デフォルトパラメータは、Ktuple=2、gap penalty=5、window=4、「diagonals saved」=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

【0071】

別法では、National Center for Biotechnology Information (NCBI) Basic Local Alignment Sear

50

ch 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/bl2.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

【0072】

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。

【0073】

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。

【0074】

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。

【0075】

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ンメントの場合、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。PAM250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CLUSTAL Vによって報告される。

【0076】

別法では、NCBI BLAST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「BLAST 2 Sequences」ツールVersion 2.0.12 (Apr-21-2000)でblastp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。

【0077】

Matrix: BLOSUM62
Open Gap: 11 及び Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 3
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15、20または30、40、50、70、150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。

【0078】

「ヒト人工染色体 (HAC)」は、約6 kb (キロベース) ~ 10 MbのサイズのDNA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。

【0079】

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。

【0080】

「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 (stringency) 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68℃で、約6×SSC、約1% (w/v) のSDS、並びに約100 µg/mlのせん断して変性したサケ精子DNAが含まれる。

【0081】

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pHにおける特定の配列の熱融点 (Tm) より約5~20℃低く選択される。このTmは、(所定のイオン強度とpHの下) 標的の配列の50%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。Tm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、Sambrook, J. 他による, 1989, Molecular Cloning: A L

10

20

30

40

50

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。

【0082】

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。

10

【0083】

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、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。

20

【0084】

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。

【0085】

「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。

【0086】

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PRTS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。

30

【0087】

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。

【0088】

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。

40

【0089】

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。

【0090】

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。

【0091】

50

「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。

【0092】

「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。 10

【0093】

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。

【0094】

「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。 20

【0095】

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。 30

【0096】

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。 40

【0097】

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 50

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「ミス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、上記した方法に限定されるものではない。

【0098】

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。

【0099】

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。

【0100】

「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。

【0101】

「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な部分である。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。

【0102】

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、糖鎖のバックボーンがデオキシリボースではなくリボースからなる。

【0103】

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。

【0104】

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。

10

【0105】

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、最も好ましいしくは90%以上除去されたものを指す。

20

【0106】

「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。

【0107】

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、チップ、スライド、ウェハ、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

【0108】

「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。

【0109】

「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、及び微粒子照射が含まれるが、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

【0110】

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

50

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。

【0111】

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。

【0112】

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、あるポリペプチド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。

【0113】

(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、細胞増殖異常、発生または発達障害、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。

【0114】

表 1 は、本発明の完全長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 (Polynucleotide 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 Polynucleotide ID) の両方によって示されている。

【0115】

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て本明細書の一部とする。

【0116】

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。

【0117】

表2および表3は共に、本発明の各ポリペプチドの特性を一覧にしたものであって、これらの特徴が、請求するポリペプチドがプロテアーゼであることを立証する。SEQ ID NO: 1が、M1残基からG225残基においてヒトユビキチン特異的プロセシングプロテアーゼ (GenBank ID g9971757) と48%の同一性を有することがBasic Local Alignment Search Tool (BLAST) によって示された (表2を参照)。BLASTの確率スコアは $1.00e-49$ である。これは、探しているポリペプチド配列アラインメントが偶然の一致により得られる確率を示す。また、SEQ ID NO: 1がユビキチンカルボキシル末端ヒドロラーゼ触媒部位ドメインを含むことが、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおける統計的に有意な一致の検索により確定された (表3を参照)。スコアが53.4ビットであり、探しているポリペプチド配列アラインメントが偶然の一致により得られる確率を示すE値が $4.9e-12$ である。このモチーフの存在が、BLIMPS (確率スコアが $2.6e-4$) 及びMOTIFS分析により確証された。このことから、SEQ ID NO: 1がユビキチンカルボキシル末端ヒドロラーゼであることが更に証明された。別の例では、SEQ ID NO: 2が、ヒトprostasin、即ちセリンプロテアーゼ (GenBank ID g1143194) のアミノ酸残基の15-235において45%の同一性を有することがBasic Local Alignment Search Tool (BLAST) によって示された (表2を参照)。探しているポリペプチド配列アラインメントが偶然の一致により得られる確率を示すBLAST確率スコアは $1.3e-46$ である。また、SEQ ID NO: 2がトリプシンファミリーセリンプロテアーゼ活性部位ドメインを含むことが、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおける統計的に有意な一致の検索により確定された。この一致は、確率スコアが $2.7e-58$ である。BLIMPS、MOTIFS、及びPROFILES CAN分析により、このドメインの存在が確認された (表3を参照)。また、BLIMPS分析によりkringleドメインも明らかにされ、SEQ ID NO: 2がトリプシンファミリーのセリンプロテアーゼであることが更に証明された。別の例では、SEQ ID NO: 7が様々なペプチドを加水分解するジペプチダーゼであり (Kozak, E. and S. Tate (1982) J. Biol. Chem. 257: 6322-6327)、ペネム及びカルバペネムなどの抗生物質の β ラクタム環の加水分解に関係する (C

ampbellら(1984) J. Biol. Chem. 259: 14586-14590)。SEQ ID NO: 7が、ヒトジペプチダーゼ前駆体(GenBank ID g219600)に対して、全長411のアミノ酸の377のアミノ酸において48%の配列同一性を有することが、Basic Local Alignment Search Tool (BLAST)によって示された。探しているポリペプチド配列アラインメントが偶然の一致により得られる確率を示すBLAST確率スコアは 1.1×10^{-88} である。更に、本発明のプロテアーゼが腎ジペプチダーゼドメインを有することが、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおける統計的に有意な一致の検索により確定された。腎ジペプチダーゼPFAMヒットのHMMスコアは412.7である。更にBLIMPS、MOTIFS、BLAST-DOMO、及びBLAST-PRODOM分析のデータにより、SEQ ID NO: 7が腎ジペプチダーゼであることが確証された。BLIMPS-BLOCKSヒットにより、1040-1537の範囲の局所領域にスコアが付けられた。BLAST-DOMOヒット確率スコアは 5.2×10^{-85} である。BLAST-PRODOMヒット確率スコアは 4.7×10^{-73} である。別の例では、SEQ ID NO: 8が、ヒト膜貫通トリプターゼ(GenBank ID g6103629)と86%の同一性を有することがBasic Local Alignment Search Tool (BLAST)によって示された(表2を参照)。探しているポリペプチド配列アラインメントが偶然の一致により得られる確率を示すBLAST確率スコアは 3.9×10^{-166} である。SEQ ID NO: 8は、トリプシンファミリープロテアーゼ活性部位ドメインを含み、確率スコアが 5.3×10^{-89} であることが、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおける統計的に有意な一致の検索により確定された。BLIMPS、MOTIFS、及びPROFILES CAN分析により、このモチーフの存在が確認された。また、BLIMPS分析により、SEQ ID NO: 8がkringleドメイン及びI型フィブロンネクチンドメインを含むことも示された。HMMERを用いる分析により、膜貫通ドメインの存在が明らかにされた(表3を参照)。これらの分析を総合すると、SEQ ID NO: 8がトリプシンファミリーのプロテアーゼの膜貫通メンバーであることが分かる。別の例では、SEQ ID NO: 17が、或る部分においてヒト膜貫通型セリンプロテアーゼ1(MT-SP1, GenBank ID g6002714)と同一性が44%であることが、Basic Local Alignment Search Tool (BLAST)によって示された(表2を参照)。探しているポリペプチド配列アラインメントが偶然の一致により得られる確率を示すBLAST確率スコアは 5.1×10^{-94} である。SEQ ID NO: 17がトリプシンファミリーのセリンプロテアーゼ活性部位ドメインを含むことが、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおける統計的に有意な一致の検索により確定された(表3を参照)。HMMを基にした分析により、更にSEQ ID NO: 17のN末端に近い膜貫通ドメインが明らかにされた。低密度リボタンパク質受容体、及びMT-SP1(PDOC00929)を含むその他のタンパク質に見られる或るドメインもこの方法で同定された。トリプシン活性部位モチーフの存在がPROFILES CAN、BLIMPS、及びMOTIFS分析により確認された。BLIMPS分析により、kringleドメイン及びI型フィブロンネクチンドメインの存在が明らかにされた。これらのデータを総合すると、更にSEQ ID NO: 17がトリプシンファミリーのセリンプロテアーゼの膜貫通メンバーであることが確証された。同様の方法で、SEQ ID NO: 3-6、SEQ ID NO: 9-16、及びSEQ ID NO: 18-21を分析し、アノテーションを付けた。SEQ ID NO: 1-21の分析に用いたアルゴリズム及びパラメータを表7に記載する。

【0118】

表4に示されているように、本発明の完全長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 (Inc yte 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: 22-42 を同定するため、或いは SEQ ID NO: 22-42 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完全長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、それぞれの完全長配列に対する列 5 の cDNA 配列及び/または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。

【0119】

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、7246467T8 はインサイト cDNA 配列の識別番号であり、PROSTM Y01 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、71041539V1) に由来する。または、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、g5745066) の識別番号の場合もある。または、列 5 の識別番号は、ゲノム DNA の Genscan 分析によって推定されるコード領域の場合もある。例えば、GNN. g7208751_000002_002. edit は、Genscan 推定コード配列の識別番号であり、g7208751 は Genscan に用いた配列の GenBank 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、列 5 の識別番号は、「エキソンステッチ (exon-stitching)」アルゴリズムによって得られた cDNA および Genscan 推定エキソンの両方からなる集合体の場合もある。例えば、FL1389845_00001 はステッチ配列であり、1389845 はアルゴリズムが適用された配列のクラスターの識別番号であり、00001 はアルゴリズムによって生成された推定の番号である (実施例 5 を参照)。または、列 5 の識別番号は、「エキソンストレッチ (exon-stretching)」アルゴリズムによって得られた cDNA および Genscan 推定エキソンの両方からなる集合体の場合もある。例えば、FL2256251_g7708357_000002_g6103629 は、ストレッチ配列の識別番号であって、2256251 はインサイトプロジェクト識別番号であり、g7708357 はエキソンストレッチアルゴリズムが適用されたヒトゲノム配列の GenBank 識別番号であり、g6103629 は最も近い GenBank タンパク質相同体の GenBank 識別番号である (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。

【0120】

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完全長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。

【0121】

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

10

20

30

40

50

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。

【0122】

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: 22-42 からなる群から選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: 22-42 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。

【0123】

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: 22-42 からなる群から選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: 22-42 からなる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。

【0124】

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。

【0125】

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有する 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。

【0126】

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。

【0127】

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: 22-42 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。

【0128】

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)。

【0129】

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 Applic 1: 111-119 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。

【0130】

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) ライブラリが完全長の 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。

【0131】

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 つの異なった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。

10

【0132】

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。

20

【0133】

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、スプライスバリエーションの作製等が可能である。

30

【0134】

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. Biotechnol. 17: 259-264; Cramer, 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。

40

【0135】

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

50

、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。

10

【0136】

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）。

20

【0137】

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で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.* 20: 1-18-162. を参照）。

30

【0138】

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）。

40

【0139】

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

50

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。

【0140】

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。

【0141】

PRTSの発現に酵母の発現系の使用が可能である。 α 因子やアルコールオキシダーゼやPGH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、酵母菌サッカロミセス・セレビジエまたは 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) Met

h o d s E n z y m o l . 1 5 3 : 5 1 6 - 5 4 4、及び S c o r e r . C . A . ら
(1 9 9 4) B i o / T e c h n o l o g y 1 2 1 - 1 8 1 - 1 8 4 . を参照)。

【0142】

植物系も P R T S の発現に使用可能である。P R T S をコードする配列の転写は、例えば、C a M V 由来の 3 5 S 及び 1 9 S 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 T M V (例えば、C o r u z z i , G . ら (1 9 8 4) E M B O J . 3
: 1 6 7 1 - 1 6 8 0 ; B r o g l i e , R . ら (1 9 8 4) S c i e n c e 2
2 4 : 8 3 8 - 8 4 3 ; および W i n t e r , J . ら (1 9 9 1) R e s u l
t s P r o b l . C e l l D i f f e r . 1 7 : 8 5 - 1 0 5 を参照) 由来のオメ
ガリーダー配列と組み合わせて促進される。これらの作製物は、直接の D N A 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。(10
例えば、The McGraw Hill Yearbook of Science and
Technology (1 9 9 2) M c G r a w H i l l N Y , p p . 1 9 1 - 1 9
6 を参照)。

【0143】

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 P R T S 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 E 1 または E 3 領域への挿入により、感染した宿主細胞に P R T S を発現
する生ウイルスを得ることが可能である (L o g a n , J . 及び S h e n k , T . (1 20
9 8 4) P r o c . N a t l . A c a d . S c i . 8 1 : 3 6 5 5 - 3 6 5 9 を参照
) 。さらに、ラウス肉腫ウイルス (R S V) 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、S V 4 0 または E B V を基にしたベクターを用いることが可能で
ある。

【0144】

ヒト人工染色体 (H A C) 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 D N A の断片を供給可能である。治療のために約 6 k b ~ 1 0 M b の H A C s 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、H a r r i n g t o n . J . J . 他 (1 9 9 7) N a t G e n 30
e t . 1 5 : 3 4 5 - 3 5 5 . を参照)。

【0145】

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 P R T
S 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、P R T S 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 40
増殖可能である。

【0146】

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 t k⁻ または a p
r⁻ 細胞において使用される。(例えば、W i g l e r , M . 他 (1 9 7 7) C e l
l 1 1 : 2 2 3 - 2 3 2 ; 及び L o w y , I . 他 (1 9 8 0) C e l l 2 2 : 8 1
7 - 8 2 3 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 d h f r はメトトレキセートに対する耐性を与え、n
e o はアミノグリコシッドネオマイシン及び G - 4 1 8 に対する耐性を与え、a l s 或い 50

はpatはクロルスルフロン (chlorosulfuron)、ホスフィノトリシ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; Colbrese-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)。アニトシアニン、緑色蛍光タンパク質 (GFP; Clontech)、 β 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)。

【0147】

マーカー遺伝子の発現の存在／非存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。誘導または選択に応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。

【0148】

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。

【0149】

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 (RIA)、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イムノアッセイ (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)。

【0150】

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。

10

【0151】

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。

20

【0152】

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「プレプロ」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類の宿主細胞 (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。

30

【0153】

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、PRTS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 (GST)、マルトース結合タンパク質 (MBP)、チオレドキシン (Trx)、カルモジュリン結合ペプチド (CBP)、6-His、FLAG、c-mc、赤血球凝集素 (HA) が含まれるが、これらに限定されるものではない。GST及びMBP、Trx、CBP、6-His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 (phenylarsine oxide)、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。FLAG、c-mc、及び赤血球凝集素 (HA) 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、PRTS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、PRTS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、Ausubel, (1995、前出 ch 10) 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

40

50

パク質の発現及び精製を促進できる。

【0154】

本発明の別の実施例では、TNTウサギ網状赤血球可溶化液またはコムギ胚芽抽出系（Promegea）を用いて *in vitro* で放射能標識したPRTSの合成が可能である。これらの系は、T7またはT3、SP6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、³⁵S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。

【0155】

本発明のPRTSまたはその断片を用いて、PRTS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、PRTS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。

【0156】

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PRTS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（Coligan, J. E. 他（1991）Current Protocols in Immunology 1（2）の5章等を参照）。同様に、化合物は、PRTS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PRTS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。PRTSを発現する細胞またはPRTSを含有する細胞膜断片を試験化合物と接触させて、PRTS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。

【0157】

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PRTSと結合させるステップと、PRTS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。

【0158】

本発明のPRTSまたはその断片を用いて、PRTS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、PRTSが少なくとも1つの試験化合物と結合する、PRTS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PRTSの活性が試験化合物非存在下でのPRTSの活性と比較する。試験化合物の存在下でのPRTSの活性の変化は、PRTS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PRTSの活性に適した条件下でPRTSを含む *in vitro* 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、PRTS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。

【0159】

別の実施例では、胚性幹細胞（ES細胞）における相同組換えを用いて動物モデル系内で、PRTS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「ノックアウト」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, 175, 383号及び第5, 767, 337号等を参照）。例えば129/SvJ細胞株等のマウスES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ES細胞は、ネオマイシンホスホトランスフェラーゼ遺伝子（neo: Capecchi, M. R. (1989) Science 244:1288-1292）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、Cre-loxP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ノックアウトする（Marth, J. D. (1996) Clin. Invest. 97:1999-2002; Wagner, K. U. 他 (1997) Nucleic Acids Res. 25:4323-4330）。形質転換したES細胞を同定し、例えばC57BL/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。

【0160】

PRTSをコードするポリヌクレオチドを*in vitro*でヒト胚盤胞由来のES細胞において操作することが可能である。ヒトES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（Thomson, J. A. 他 (1998) Science 282:1145-1147）。

【0161】

PRTS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「ノックイン」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。ノックイン技術を用いて、PRTSをコードするポリヌクレオチドの或る領域を動物ES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PRTSを乳汁内に分泌するなどPRTS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（Janne, J. 他 (1998) Biotechnol. Annu. Rev. 4:55-74）。

【0162】

（治療）

PRTSは、ペプチド結合を加水分解するのに有用である。PRTS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、PRTSの発現は血液組織、神経組織、生殖組織、内分泌組織、尿性器組織、病変組織、奇形癌組織、及び腫瘍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、細胞増殖異常、発生または発達障害、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。

【0163】

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、 α -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、子癰前症、子癰、妊娠性急性肝脂肪、妊娠性肝臓胆汁うっ滞と、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）、肺胞たんばく症、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（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（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、脾臓、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、癲癇、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-トゥ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（Syndenham'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、先端線維性軟肉、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、ビダール苔癬、asteatotic eczema、うっ血性皮膚炎、うっ血性潰瘍、脂漏性皮膚炎、乾癬、扁平苔癬、バラ色秕糠疹、膿疱疹、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 corn eal 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、癲癇、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（cerebellar retinal 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、

10

20

30

40

50

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。

【0164】

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。

【0165】

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。

10

【0166】

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。

【0167】

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、細胞増殖異常、発生または発達障害、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ッティング或いは運搬機構として間接的に用いられ得る。

20

【0168】

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。

【0169】

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て投与することもできる。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。

30

【0170】

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。詳しくは、精製されたPRTS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PRTSと特異的に結合するものを同定が可能である。PRTS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、Fabフラグメント、及びFab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。

40

【0171】

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、PRTS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

50

に用いられるアジュバントの中では、BCG (*bacilli Calmette-Guérin*) 及び *Corynebacterium parvum* が特に好ましい。

【0172】

PRTS 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。PRTS アミノ酸の短いストレッチは、KLH 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。

【0173】

PRTS 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 EBV-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、Kohler, G. ら. (1975) *Nature* 256:495-497; Kozbor, D. ら. (1985) *J. Immunol. Methods* 81-8-42; Cote, R. J. ら. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; Cole, S. P. ら. (1984) *Mol. Cell Biol.* 62:109-120 を参照）。

【0174】

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、Morrison, S. L. 他. (1984) *Proc. Natl. Acad. Sci.* 81-4851-4855; Neuberger, M. S. 他. (1984) *Nature* 312:604-608; Takeda, S. ら. (1985) *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、PRTS 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、Burton D. R. (1991) *Proc. Natl. Acad. Sci.* 88:11120-3 を参照）。

【0175】

抗体は、リンパ球集団の中の *in vivo* 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、Orlandi, R. 他. (1989) *Proc. Natl. Acad. Sci.* 86:3833-3837; Winter, G. 他. (1991) *Nature* 349:293-299 を参照）。

【0176】

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、限定するものではないが、抗体分子のペプシン消化によって生成される $F(ab')_2$ 断片、及び $F(ab')_2$ 断片のジスルフィド架橋を減少させて生成される Fab 断片が含まれる。別法では、所望の特異性を有するモノクローナル Fab 断片の迅速且つ容易な同定が可能となるように、Fab 発現ライブラリを作製することもできる（例えば、Huse, W. D. ら. (1989) *Science* 254:1275-1281 を参照）。

【0177】

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。確立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アッセイ、または免疫放射線アッセイの様々なプロトコルが、当分野では周知である。通常このようなイムノアッセイでは、PRTS とその特異的抗体との間で

10

20

30

40

50

形成された複合体の測定が含まれる。二つの非干渉性 P R T S 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 (P o u n d、前出)。

【0178】

ラジオイムノアッセイ技術と共に S c a t c h a r d 分析などの様々な方法を用いて、P R T S に対する抗体の親和性を評価する。親和性を結合定数 K_a で表すが、この K_a は、平衡状態の下で P R T S 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 P R T S 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 K_a は、P R T S 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 P R T S エピトープに単一特異的なモノクローナル抗体医薬の K_a は、親和性の真の測定値を表す。 K_a 値が $10^9 \sim 10^{12} \text{ L/mol}$ の高親和性抗体医薬は、P R T S 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。 K_a 値が $10^6 \sim 10^7 \text{ L/mol}$ の低親和性抗体医薬は、P R T S 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 (i m m u n o p u r i f i c a t i o n) 及び類似の処理に用いるのが好ましい。(C a t t y, 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)。

10

【0179】

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 $1 \sim 2 \text{ mg/ml}$ の特異的な抗体、好ましくは $5 \sim 10 \text{ mg/ml}$ 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、P R T S 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、C a t t y, 前出, 及び C o l i g a n 他、前出を参照)。

20

【0180】

本発明の別の実施例では、P R T S 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、治療目的で利用することができる。ある実施態様では、P R T S 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 (D N A 及び R N A、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、P R T S 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まざまな位置から設計可能である。

30

【0181】

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、S l a t e r, J. E. 他 (1998) J. A l l e r g y C l i n. I m m u n o l. 102 (3): 469-475; and S c a n l o n, 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、M i l l e r, A. D. (1990) B l o o d 76: 271; A u s u b e l, 前出; U c k e r t, W. and W. W a l t h e r (1994) P h a r m a c o l. T h e r. 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ップ、及び当分野で周知のその他の系が含まれる (R o s s i, J. J. (1995) B r. M e d. B u l l. 51 (1): 217-225; B o a d o, R. J. 他 (1998) J. P h a r m. S c i. 87 (11):

40

50

1308-1315; and Morris, M. C. 他 (1997) *Nucleic Acids Res.* 25 (14): 2730-2736. を参照)。

【0182】

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; Poeschbla, 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。

【0183】

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トランスポソン (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。

【0184】

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、PCDNA 3.1、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β-

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-R E X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PVGRR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) 10、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。

【0185】

市販のリボ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(Neu mann, 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 20ェクションプロトコルを変更する必要がある。

【0186】

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. 30 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 40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 50

, U. 他 (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290)。

【0187】

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
 或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 10
 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 (1
 997) *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。

【0188】

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或いは
 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 20
 。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 (199
 9) *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5,804,41
 3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 30
 d92についての記載がある。また上記特許には、ICP4、ICP27及びICP2
 2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.* 73:5
 19-532 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。

【0189】

別法では、α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 40
 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α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)。α
 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α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 50

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α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、α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α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。αウイルスの感染性cDNAクローンの操作、αウイルスcDNA及びRNAのトランスフェクション、並びにαウイルスの感染方法は当分野で周知である 10

【0190】

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。三重鎖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、Ge, J. E. ら. (1994) *In: Huber, B. E. 及び B. I. Carr, Molecular and Immunologic 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。 20

【0191】

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。

【0192】

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、標的分子をスキャン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。 30

【0193】

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。 40

【0194】

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 50

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、チオール、及び同様の修飾形態だけでなく、イノシン、キューオシン (*queosine*)、及びワイブトシン (*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。

【0195】

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。

10

【0196】

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 (Bruce, T. W. 他 (1997) 米国特許第5,686,242号、Bruce, T. W. 他 (2000) 米国特許第6,022,691号)。

20

30

40

【0197】

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

50

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）。

【0198】

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。

【0199】

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。

10

【0200】

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、骨髓内、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。

【0201】

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、針注射を用いなくて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。

20

【0202】

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。

30

【0203】

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ポソーム製剤は、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）。

【0204】

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。

40

【0205】

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

50

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、 LD_{50}/ED_{50} 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、 ED_{50} 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。

【0206】

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。 10

【0207】

通常、薬用量は投与の経路によって異なるが、約0.1~100,000 μ g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。 20

【0208】

(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。 30

【0209】

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、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(photo metric)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。 40

【0210】

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値の変化を監視する。

【0211】

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 50

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、プローブがPRTSをコードする自然界の配列のみを同定するかどうか、或いはアレルや関連配列コードする自然界の配列のみを同定するかどうかによって決まるであろう。

【0212】

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: 22-42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。

10

【0213】

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。

20

【0214】

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、α-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、

30

40

50

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、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 (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、骨髄腫、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、脾臓、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、癲癇、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—トゥ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 舞踏病 (Syndenham'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、带状疱疹、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、先端線維性軟疣、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、うっ血性皮膚炎、うっ血性潰瘍、脂漏性皮膚炎、乾癬、扁平苔癬、バラ色糝糖疹、膿疱疹、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、ジーマンス水疱魚鱗癬 (ichthyosis bullosa of Siemens)、剥脱性魚鱗癬、手掌足底角化症、掌蹠角皮症 (keratosis palmo plantaris 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、癲癇、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、筋萎縮性側策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、サザン法やノーザン法、ドットブロット法、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、及び変異PRTS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。

【0215】

ある実施態様では、PRTS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。PRTS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、制御サンプルと較べて著しく変わっている場合は、サンプル内のPRTS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。

【0216】

PRTS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。

【0217】

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ベースで繰り返して、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。

【0218】

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。

【0219】

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 *i n v i t r o* 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。

【0220】

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可能である。f S S 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、D N A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S N P (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D N A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、D N A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D N A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S N Pを検出し、特徴付ける。

【0221】

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 (1993) J. I m m u n o l. M e t h o d s, 159:235-44; D u p l a a, C. ら (1993) A n a l. B i o c h e m. 229-236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。

【0222】

更に別の実施例では、本明細書で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。

10

【0223】

別の実施例では、PRTS、PRTSの断片、PRTS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。

20

【0224】

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 (Seilliamer 他、米国特許第5,840,484号の "Comparative Gene Transcript Analysis" 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。

30

【0225】

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。

【0226】

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

40

50

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。

10

【0227】

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。

【0228】

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ケンシングする。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。

20

30

40

【0229】

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ELEMENT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ELEMENT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、Mendoeze, 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

50

アレイエレメントにおける蛍光結合の量を検出し得る。

【0230】

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

10

【0231】

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ケンシングし、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。

【0232】

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。

20

【0233】

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 (1997) 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。

30

【0234】

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. (1

40

50

993) 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)の遺伝とが関連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)。

【0235】

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関連し得る(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 10
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つけることができる。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。

【0236】

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っていなくても、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。 20

【0237】

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。 30

【0238】

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、Geyser, 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 40
ぎに、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。

【0239】

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。

【0240】

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。

【0241】

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。

【0242】

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60/212,336号、同第60/213,995号、同第60/215,396号、同第60/216,821号、及び同第60/218,946号に言及することをもって本明細書の一部とする。 10

【0243】

(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。ある組織をグアニジニウムイソチオシアネート溶液においてホモジナイズして溶解する一方、別の組織を、フェノールにおいて、或いは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であるTRIZOL (Life Technologies) などの変性剤の好適な混合液においてホモジナイズして溶解した。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。RNAは、イソプロパノール或いは酢酸ナトリウムとエタノールのどちらか、或いは別の一般的な方法でこの溶解物から沈殿させた。 20

【0244】

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。 30

【0245】

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させた。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 S1000、SEPHAROSE CL2B、SEPHAROSE CL4Bカラムクロマトグラフィー (Amersham Pharmacia Biotech)、または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)、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)、またはそれら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 40 50

プラスミドを、Stratagene社のXL1-Blue, XL1-Blue MRF、SOLR、またはLife Technologies社のDH5αまたはDH10B、ELECTROMAX DH10Bを含むコンピテント大腸菌細胞に導入し形質転換した。

【0246】

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は、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。プラスミドの精製には、MagicまたはWIZARD Minipreps DNA精製システム (Promega)、AGTC Miniprep精製キット (Edge Biosystems, Gaithersburg MD)、QIAGEN社のQIAWELL 8 Plus、QIAWELL 8 Plus Plasmid、またはQIAWELL 8 Ultra Plasmid 精製システム、またはREAL Prep 96プラスミドキットの内の少なくとも1つを用いた。沈殿させた後、0.1 ml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。

【0247】

別法では、ハイスルーブットの直接結合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ステップは、1つの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 (Lab systems Oy, Helsinki, Finland) を用いて蛍光定量的に測定した。

【0248】

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クエンシングした。cDNAのシークエ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클ー或いはPTC-200 thermal cycler (MJ Research) などのハイスルーブット装置を用いて行った。cDNAのシークエンシング反応は、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シークエンシングキットに含まれる試薬を用いて行った。cDNAシークエ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シークエンシングシステム (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シークエンシングシステム (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。

【0249】

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、オープンリーディングフレームを決定するべくcDNA群をスクリーニングした。完全長の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。別法では、本発明のポリヌクレオチドは、完全長翻訳ポリペプチドの任意のメチオニン残基から始まり得る。次に、完全長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LASERGENE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マルチシークエンス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タを用いて作成した。

【0250】

表7は、インサイトcDNAおよび完全長配列の組み立て、および組み立てた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）。

【0251】

完全長の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: 22-42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。

【0252】

4 ゲノムDNAからのコード配列の同定および編集

推定上のプロテアーゼは、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クソンを連結して、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、

F A S T A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。G e n s c a nによって一回で解析できる配列の最大長さは30 kbに設定されている。これらのG e n s c a n推定c D N A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P F A M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c D N A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G e n s c a n推定配列を、B L A S T解析を用いてg e n e p tおよびg b p r i公共データベースの配列と比較した。必要に応じて、G e n s c a n推定c D N A配列を、g e n e p tにおいてB L A S Tで最もヒットした配列と比較して、G e n s c a n推定配列における余分なエクソンや省いてしまったエクソンなどのエラーを修正し、編集した。B L A S T解析を用いてG e n s c a n推定c D N A配列を含むインサイトc D N Aまたは公共のc D N A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c D N AがG e n s c a n推定c D N A配列を含む場合、この情報を用いてG e n s c a n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G e n s c a n推定コード配列とインサイトc D N Aおよび／または公共のc D N A配列を組み立てて作製した。別法では、完全長ポリヌクレオチド配列は、その全てが編集した或いは未編集のG e n s c a n推定コード配列から作製した。

【0253】

5 c D N A配列データを用いるゲノム配列データの組み立て ステッチ配列 (S t i c h e d S e q u e n c e)

部分c D N A配列を、実施例 4に記載したG e n s c a n遺伝子同定プログラムによって推定されたエクソンで伸長した。実施例 3に記載されたように組み立てられた部分c D N AをゲノムDNAにマッピングし、関連するc D N Aと1或いは複数のゲノム配列から推定されるG e n s c a nエクソンとを含む複数のクラスターに分けた。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分析し、c D N Aおよびゲノム情報を統合し、完全長の配列を作製するために、その後に確認、編集、または伸長して潜在的なスプライスバリエーションを生成した。クラスターの2つ以上の配列に配列区間の全長が存在する配列区間を同定し、このように同定した区間を推移により等しいと考えた。例えば、或る区間がc D N Aおよび2つのゲノム配列に存在する場合、これら3つ全ての区間を等しいと考えた。この方法によって、関連しないが連続するゲノム配列をc D N A配列によって繋ぎ1つにした。このようにして同定された区間を、親配列 (p a r e n t s e q u e n c e) 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した。1種類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 (c D N Aとc D N A、またはゲノム配列とゲノム配列) は、親配列の種類が異なる連結 (c D N Aとゲノム配列) よりも優先された。得られたステッチ配列を翻訳し、B L A S T解析でg e n p e p tおよびg b p r i公共データベースにおける配列と比較した。G e n s c a nによって推定された不適当なエクソンを、g e n e p tにおいてB L A S Tで最もヒットした配列と比較して修正した。必要に応じてこのような配列を更なるc N D A配列若しくはゲノムDNA検査により伸長した。

【0254】

ストレッチ配列 (S t r e t c h e d S e q u e n c e)

部分DNA配列をB L A S T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c D N Aを、B L A S Tプログラムを用いてG e n B a n k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、G e n B a n k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c D N A或いはG e n S c a nエク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 (H S P) を用いてキメラタンパク質を作製し、G e n B a n k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G e n B a n kの相同タンパク質に関連して、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検索するために

、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。得られたストレッチ配列を完全な遺伝子を含んでいるかについて検査した。

【0255】

6 PRIS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: 22-42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: 22-42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。

10

【0256】

遺伝子地図の位置は、ヒト染色体の範囲または区間によって表される。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組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」 (<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。

20

【0257】

このようにして、SEQ ID NO: 25が、第5染色体の69.60 cM~76.50 cMの範囲内にマッピングされた。SEQ ID NO: 28が、第16染色体の81.80 cM~84.40 cMの範囲内にマッピングされた。

30

【0258】

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)。

【0259】

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同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、

40

【0260】

【数1】

(BLAST スコア×配列一致率)

5 × (長さ(配列1), 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。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%の同一性で重畳が70%であるか、或いは88%の同一性で重畳が100%であるかの何れかの場合である。積スコア50は、100%の同一性で重畳が50%であるか、或いは79%の同一性で重畳が100%であるかの何れかの場合である。

10

【0261】

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。

20

30

【0262】

8 PRTS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。

40

【0263】

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。

【0264】

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

50

)、E L O N G A S E 酵素 (L i f e T e c h n o l o g i e s)、P f u D N A ポリ
 メラーゼ (S t r a t a g e n e) を含む。プライマーの組、P C I A と P C I B に対
 して以下のパラメータで増幅を行った。

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

10

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。

【0265】

各ウェルのDNA濃度は、1X TE及び0.5µlの希釈していないPCR産物に溶解
 した100µlのP I C O G R E E N 定量試薬 (0.25% (v / v) P I C O G R E
 E N ; M o l e c u l a r P r o b e s , E u g e n e O R) を不透明な蛍光光度計
 プレート (C o m i n g C o s t a r , A c t o n M A) の各ウェルに分配してDN
 A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F l u o r o s k a n I I
 (L a b s y s t e m s O y , H e l s i n k i , F i n l a n d) 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5～10µl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。

20

【0266】

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、C v i J I
 コレラウィルスエンドヌクレアーゼ (M o l e c u l a r B i o l o g y R e s e a r
 c h , M a d i s o n W I) で消化し、p U C 18ベクター (A m e r s h a m P h
 a r m a c i a B i o t e c h) に再連結する前に音波処理またはせん断を行った。シ
 ョ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A g a r A C E (P r o m e
 g a) で消化した。T4リガーゼ (N e w E n g l a n d B i o l a b s , B e v e
 r l y M A) を用いて伸長したクローンをp U C 18ベクター (A m e r s h a m P
 h a r m a c i a B i o t e c h) に再連結し、P f u D N A ポリメラーゼ (S t r a
 t a g e n e) 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L B / 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。

30

40

【0267】

細胞を溶解して、T a q D N A ポリメラーゼ (A m e r s h a m P h a r m a c i a
 B i o t e c h) 及びP f u D N A ポリメラーゼ (S t r a t a g e n e) 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

50

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ไซด์(dimethylsulphoxide)(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。

【0268】

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。

【0269】

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: 22-42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、OLIGO4.06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 µCiの[γ-³²P]アデノシン三リン酸(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。

【0270】

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 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0.1x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。

【0271】

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ロ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, M

10

20

30

40

50

arshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16: 27-31. を参照)。

【0272】

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。

10

【0273】

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

20

30

【0274】

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、1~2 ngの初期量から5 μg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。

40

【0275】

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

50

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。

【0276】

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100 ng/μlのアレイエレメントDNA 1 μl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5 nlのアレイエレメントサンプルを分注する。

10

【0277】

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。

【0278】

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、5×SSC、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0.2 μg含む9 μl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1.8 cm²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ビティを有する防水チェンバーに移す。チェンバーの角に140 μlの5×SSCを加えて、チェンバー内を湿度100%に保持する。このアレイを含むチェンバーを、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 (1×SSC, 0.1% SDS) において45℃で10分間、第2洗浄緩衝液中 (0.1×SSC) 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。

20

【0279】

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 nm、及びCy3を励起するための632 nmのスペクトル線を生成し得るInnova 70混合ガス10 W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1.8 cm×1.8 cmのアレイは、20 μmの解像度でスキャンする。

30

【0280】

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565 nm、Cy5では650 nm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。

40

【0281】

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

50

により生成されるシグナル強度を用いて校正する。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、校正は2つの蛍光体を有する校正する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。

【0282】

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アナログディジタル(A/D)変換ボード(Analog Devices, Inc., Norwood MA)を用いてディジタル化される。ディ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。

【0283】

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、GEM TOOLS遺伝子発現分析プログラム(Incyte)である。

【0284】

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。

【0285】

1.2 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β-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

10

20

30

40

50

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. USA 91:3224-3227; Sandig, V. 他 (1996) Hum. Gene Ther. 7:1937-1945. を参照)。

【0286】

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。

【0287】

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ニングする。このようなベクターには、pCMV SPORTTM (Life Technologies.) 及びpCR 3.1 (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。5~10 µg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1~2 µg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。

【0288】

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か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。

【0289】

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。

10

【0290】

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)。

【0291】

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を用いてブロッキング処理し、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。

20

【0292】

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ー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 Biotech) のような活性化クロマトグラフィー用レジンと抗PRTS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。

30

【0293】

PRTSを含む培養液をイムノアフィニティークラムに通し、PRTS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PRTS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、PRTSを回収する。

40

【0294】

1.6 PRTSと相互作用する分子の同定

PRTSまたは生物学的に活性なその断片を、¹²⁵I Bolton-Hunter 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PRTSと共にインキュベートし、洗浄して、標識したPRTS複合体を有する全てのウェルをアッセイする。様々なPRTS濃度で得られたデータを用いて、候補分子と結合したPRTSの数量及び親和性、会合についての値を計算する。

50

【0295】

別法では、PRTS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。

【0296】

PRTSはまた、ハイスループット型の酵母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)。

10

【0297】

17 PRTS活性の実証

プロテアーゼ活性は、様々な色素分子と結合した適切な合成ペプチド基質の加水分解によって測定する。この測定では、加水分解の程度が、遊離した発色団の分光光度法 (または蛍光分析) による吸収により定量される (Beynon, R. J. 及びJ. S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 25-55)。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) としてプロテアーゼ活性のカテゴリに從ってデザインされる。一般的に用いられるクロモゲン は、2-ナフチルアミン、4-ニトロアニリン、及びフリルアクリル酸である。例えば、アルギニン-β-ナフチルアミンをSEQ ID NO: 3の基質として用いることができ (Fukasawa, K. M. ら (1996) J. Biol. Chem. 271: 30731-30735)、また、4-phenylazobenzoyloxycarbonyl-Pro-Leu-Gly-Pro-D-ArgをSEQ ID NO: 4の基質として用いることができる。別の例では、7-アミノ-4-トリフルオロメチル-クマリン-フェニールアラニン-プロリン-AFCをSEQ ID NO: 3の基質として用いることができる。アッセイは周囲温度で実施し、好適なバッファ中に一定量の酵素及び適切な基質を含む。反応は蛍光キュベット内で行い、ペプチド基質の加水分解中に遊離した色素原の吸光度の増減を測定する。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。

20

30

【0298】

ユビキチンヒドロラーゼ活性についての代替アッセイでは、ユビキチン前駆体の加水分解を測定する。このアッセイは周囲温度で行い、適当なバッファ中に一定量のPRTS及び適切な基質を含む。SEQ ID NO: 1の場合、化学合成されたヒトユビキチン-バリンを基質として用い得る。C-末端バリン残基の基質からの切断は、毛細管電気泳動 (Franklin, K. ら (1997) Anal. Biochem. 247: 305-309) でモニタリングする。

40

【0299】

別法では、SEQ ID NO: 5のユビキチンプロテアーゼ活性は、Sloper-Mould らの方法を用いて測定し得る ((1999) J. Biol. Chem. 274: 26878-26884)。一定分量のPRTSを、5 µl の [³⁵S] で標識したユビキチン/GST融合基質と共に好適なバッファにおいて37℃で1時間インキュベートする。サンプルは、12%のSDS-PAGEゲル上で電気泳動により分離する。ユビキチンの切断をゲルの蛍光光度分析によりモニタリングする。

【0300】

別法では、SEQ ID NO: 1の活性は、例えばCollige らの方法により測定し

50

得る(1999, J. Biol. Chem. 270: 16724-16730)。一定分量のPRTSを、プロペプチドのみに放射線標識したプロコラーゲンI型と共に、250 μ lの反応液即ち標準的なアッセイバッファ(50 mMカコジル酸ナトリウム, pH 7.5, 200 mM KCl, 2 mM CaCl, 2.5 mM NEM, 0.5 mM PMSF, 及び0.02% Brjを含む)において37℃で1時間インキュベートする。26℃で16時間インキュベートした後、50 μ lのEDTA溶液(0.2 M EDTA, pH 8, 0.5% SDS, 0.5 M DTT)及び300 μ lの99%エタノールを加えて反応を停止させる。サンプルを4℃に30分間保った後、9500 gで30分間遠心分離する。コラーゲン及び非切断放射活性pN-コラーゲン基質をペレットにすると、遊離アミノプロペプチドが溶液中に残る。上清のアリコットを液体シンチレーション分光法でアッセイする。 10

【0301】

別法では、プロテアーゼ活性についてのアッセイは、蛍光共鳴エネルギー移動(FRET)を利用する。このFRETは、適切にスペクトルがオーバーラップしている1つのドナーフルオロフォアと1つのアクセプターフルオロフォアが近接すると生じる。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、緑色蛍光タンパク質の赤色にシフトした変異体(red-shifted variant)(RSGFP4)と青色変異体(BFP5)との間に融合する。この融合タンパク質は、エネルギー移動がBFP5からRSGFP4へと起こったことを示すスペクトル特性を有する。融合タンパク質をPRTSと共にインキュベートすると、基質が切断され二つの蛍光タンパク質に分離する。これにはエネルギー移動の顕著な減少が伴う。この減少はPRTSを添加する前と後における発光スペクトルを比較することで測定できる(Mitra, R. Dら、(1996) Gene 173: 13-17)。アッセイは生細胞において行うこともできる。この場合、蛍光基質タンパク質は細胞中で恒常的に発現されるため、PRTSを誘導ベクターに導入してPRTSの存在及び非存在によりFRETのモニタリングができるようにする(Sagot, Iら: (1999) FEBS Lett. 447: 53-57)。 20

【0302】

更なる別法では、PRTSジペプチダーゼ活性についてのアッセイでは、ロイコトリエンD4などの様々なジペプチドにおけるPRTSの加水分解活性(Kozak, E. and S. Tate (1982) J. Biol. Chem. 257: 6322-6327)、またはペネムやカルバペネムなどの抗生物質の β ラクタム環の加水分解を測定する(Campbellら、(1984) J. Biol. Chem. 259: 14586-14590)。 30

【0303】

18 PRTS基質の同定

ファージディスプレイライブラリを用いて、PRTSの最適基質配列を識別し得る。リンカー及び既知の抗体エピトープが後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ニングする。遺伝子IIIがコートタンパク質をコードし、エピトープが各々のファージ分子表面上に表示され得る。ライブラリは蛋白分解条件下で、PRTSとインキュベートし、六量体がPRTS切断部位をコードする場合はエピトープが除去されるようにする。エピトープを認識する抗体を、固定したプロテインAと共に添加する。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く。次に表面のファージを増幅し、スクリーニングを更に数回行う。次に個々のファージクローンを単離し、配列決定する。これらペプチド基質の反応速度を、実施例17に於けるアッセイを用いて研究し、最適な切断配列を得ることが可能である(Ke, S. H.ら(1997) J. Biol. Chem 272: 16603-16609)。 40

【0304】

この方法は、in vivo PRTS基質をスクリーニングするために、ファージ粒子(50

T7 SELECTTM 10-3 ファージディスプレイベクター、Novagen, Madison, WI)、若しくは酵母細胞(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し得る。このような場合、全cDNAは、遺伝子IIIと適切なエピトープとの間に融合する。

【0305】

19 PRTS阻害剤の同定

実施例17のアッセイで説明したように、試験する化合物を、適切なバッファ及び基質と共に様々な濃度にし、マルチウェルプレートのウェルの中に配列する。各ウェルに対してPRTS活性を測定し、PRTS活性を阻害する各化合物の能力並びに容量反応速度を決定し得る。このアッセイは、PRTSの活性を向上させる分子を同定することにも用い得る。

【0306】

別法では、ファージディスプレイライブラリを用いてペプチドPRTS阻害剤をスクリーニングし得る。候補はプロテアーゼに強く結合するペプチドから見つける。この場合、マルチウェルプレートの各ウェルをPRTSでコーティングし、ランダムペプチドファージディスプレイライブラリ若しくはサイクリックペプチドライブラリと共にインキュベートする(Koivunen, E. ら、(1999) Nat. Biotechnol. 17: 768-774)。結合していないファージは洗い流し、残ったファージの増幅及びスクリーニングを数回繰り返す。実施例17に記載のアッセイを用いて、候補をPRTSの阻害活性について試験する。

【0307】

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の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。

【0308】

(表の簡単な説明)

表1は、本発明の完全長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。

【0309】

表2は、本発明の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。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。

【0310】

表3は、推定上のモチーフおよびドメインを含むポリペプチド配列の構造的な特徴、並びに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。

【0311】

表4は、ポリヌクレオチド配列の組み立てに用いたcDNA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。

【0312】

表5は、本発明のポリヌクレオチドの代表的なcDNAライブラリを示す。

【0313】

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。

【0314】

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、閾値パラメータを示す。

10

20

30

40

50

【表 1】

表1

インサイト プロジェクト ID	ポリベプチド SEQ ID NO:	インサイト ポリベプチド 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
275791	1	275791CD1	22	275791CB1
1389845	2	1389845CD1	23	1389845CB1
1726609	3	1726609CD1	24	1726609CB1
4503848	4	4503848CD1	25	4503848CB1
5544089	5	5544089CD1	26	5544089CB1
7474081	6	7474081CD1	27	7474081CB1
5281209	7	5281209CD1	28	5281209CB1
2256251	8	2256251CD1	29	2256251CB1
7160544	9	7160544CD1	30	7160544CB1
7477386	10	7477386CD1	31	7477386CB1
7473089	11	7473089CD1	32	7473089CB1
7604035	12	7604035CD1	33	7604035CB1
3473847	13	3473847CD1	34	3473847CB1
3750004	14	3750004CD1	35	3750004CB1
4904126	15	4904126CD1	36	4904126CB1
71268415	16	71268415CD1	37	71268415CB1
7473301	17	7473301CD1	38	7473301CB1
7473308	18	7473308CD1	39	7473308CB1
7478021	19	7478021CD1	40	7478021CB1
4333459	20	4333459CD1	41	4333459CB1
6817347	21	6817347CD1	42	6817347CB1

10

20

30

40

【表 2】

表2-1

ポリペプチド SEQ ID No:	インサイト ポリペプチド ID	GenBank ID NO:	種差スコア	GenBank 相同体
1	275791CD1	g9971757	1.00E-49	ユビキチン特異的プロセシングプロテアーゼ [ヒト]
2	1389845CD1	g1143194	1.30E-46	prostatasin [ヒト]
3	1726609CD1	g10719660	0	RNPEP様タンパク質 [ヒト] (Horikawa, Y. (2000) Nat. Genet. 26:163-175.)
		g1754515	3.30E-96	アミノペプチダーゼ-B1 (トブナズミ) (Prieto, I. (1998) Form. Metab. Res. 30:246-248.)
4	4503848CD1	g1783122	0	エンドペプチダーゼ24.16 M1型 [Sus scrofa]
5	5544089CD1	g5410230	5.20E-43	ユビキチン特異的プロテアーゼ3 [ヒト]
6	7474081CD1	g603903	2.90E-33	トリプシノーゲン [ニワトリ]
7	5281209CD1	g11071729	0	推定上のジベプチダーゼ [ヒト]
		g219600	1.40E-88	ジベプチダーゼ前駆体 [ヒト] (Satoh, S. (1993) Biochim. Biophys. Acta 1172:181-183.)
8	2256251CD1	g6103629	3.90E-66	膜貫通型トリプターゼ [ヒト] (Wong, G.W. (1999) J. Biol. Chem. 274:30784-30793.)
9	7160544CD1	g11095188	0	ジベプチルベプチダーゼ8 [ヒト] (Abbott, C.A. (2000) Eur. J. Biochem. 267:6140-6150.)
10	7477386CD1	g1865716	0	ジベプチルベプチダーゼ IV [Stenotrophomonas maltophilia] (Mentlein, R. (1999) Regul. Pept. 85:9-24; Kahne, T. Int. J. Mol. Med. (1999) 4:3-15.)
		g1865716	0	プロコラーゲンIN-プロテイナーゼ [Bos taurus] (Colige, A. (1999) Am. J. Hum. Genet. 65:308-317.)
11	7473089CD1	g7768706	3.60E-255	トロンボスポンジン1型モチーフを含むメタロプロテアーゼ [ヒト] (Vazquez, F. (1999) J. Biol. Chem. 274:23349-23357.)
12	7604035CD1	g6164595	4.70E-68	Lacunin [Manduca sexta]

10

20

30

40

表2-2

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 相関係
13	3473847CD1	g217172	9.20E-50	aqualysin 前駆体 (アミノ酸1から55) [Thermus aquaticus]
14	3750004CD1	g5923786	4.30E-51	亜鉛メタロプロテアーゼ ADAMTS6 [ヒト]
15	4904126CD1	g186286	3.90E-40	インターロイキン1-βコンバーターゼ [ヒト]
16	71268415CD1	g6651071	0	{Cerretti, D.P. (1992) Science 256:97-100.} disintegrin 及びメタロプロテイナーゼドメイン9 [ヒト]
17	7473301CD1	g6002714	5.10E-94	{Inoue, D. (1998) J. Biol. Chem. 273:4180-4187.} 膜型セリンプロテアーゼ [ヒト]
18	7473308CD1	g1552517	6.60E-77	{Takeuchi, T. (1999) Proc. Natl. Acad. Sci. USA 96:11054-11061.}
19	7478021CD1	g3211705	5.60E-189	トリプシン/ゲン E [ヒト]
20	4333459CD1	g1754714	2.30E-67	マトリックスメタロプロテイナーゼ [アフリカツメガエル] {Yang, M. (1997) J. Biol. Chem. 272:13527-13533.}
21	6817347CD1	g7673618	5.10E-283	Oviductin [アフリカツメガエル] {Lindsay, L.L. (1999) Biol. Reprod. 60:989-995.}
				コヒキナン特異的プロテアーゼ [マウス]

10

20

30

40

表3-1

SEQ ID No.	インサイト ホリベフチド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、モチーフ、 及びドメイン	分析方法及び データベース
1	275791CD1	232	T15 T17 S23 S43 S71 T90 S93 S100 S107 S111 T122 T174 S190 T10 S141 S190 T213 T227	N98 N99	ユビキチンC末端ヒドロラーゼファミリー-2 シグネチャ2 Dch_2: Y142-Y160 ユビキチンC末端ヒドロラーゼファミリー-2 シグネチャ2 DCH-2: L138-H203 ユビキチンC末端ヒドロラーゼファミリー-2 シグネチャ2 EL00972: Y142-D166, K169-S190	MOTIFS HMMER-PFAM BLIMPS-BLOCKS
2	1389845CD1	365	S120 S187 S225 S253 S82 T31 T37 T42 Y283		TRYPSIN DM00018 A57014 45-284: T123-Q314 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZIMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: I123-Q314 セリンプロテアーゼ、トリプシンファミリー BL00134A: C148-C164 Kringledメインタンパク質 EL00021: C148-F165, V231-G252 CHYMOTRYPSIN SERINE PROTEASE PR00722: G149-C164, G208-V222 V8セリンプロテアーゼ PR00839B: C148-F165 セリンプロテアーゼ、トリプシンファミリー、 活性部位トリプシン_his.prf: I140-F188 トリプシン: I123-Q314 トリプシン_His: L159-C164	BLAST_DOMO BLAST_PRODOR BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS PROFILESSCAN HMMER-PFAM MOTIFS

10

20

30

40

表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、モチーフ、 及びドメイン	分析方法及び データベース
3	1726609CD1	416	S244 S283 S30 S370 S408 T389 T59 T78 T87	N203 N413 N57	GO HYDROLASE; LEUKOTRIENE; A-4; ZINC; DM08707 P19602 7-609; M1-P354 AMINOPEPTIDASE HYDROLASE METALLOPROTEASE ZINC N GLYCOPROTEIN PROTEIN TRANSMEMBRANE SIGNAL ANCHOR MEMBRANE PD001134; R4-S177 中性亜鉛メタロペプチダーゼファミリー BL00142; D41-E51 MEMBRANE ALANYL DIPEPTIDASE PR00756; F14-L24, D41-T56, W60-Y72	BLAST_DOMO BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_PRINTS
4	4503848CD1	714	S124 S140 S147 S179 S200 S206 S226 S333 S551 S556 S592 T114 T133 T244 T252 T270 T308 T318 T322 T376 T406 T432 T528 T585 T602 T69 Y175 Y249 Y505	N425 N485 N601	シグナル切断: M1-S34 GO ZINC; METALLOPEPTIDASE; NEUTRAL; OLIGO-PEPTIDASE DM01184 Q02038 36-702; A46-A713 HYDROLASE METALLOPROTEASE ZINC OLIGOPEPTIDASE PRECURSOR MITOCHONDRIAL ENDOPEPTIDASE MITOCHONDRION TRANSIT PEPTIDE PD002945; W60-N529 膜貫通ドメイン: L14-M34 ペプチダーゼファミリー-M3ペプチダーゼ_M3 : C98-L711 中性亜鉛メタロペプチダーゼファミリー BL00142; T504-H514 亜鉛プロテアーゼ: T504-M513 シグナル切断: M1-G27	SFScan BLAST_DOMO BLAST_PRODUM HMMER HMMER_Pfam BLIMPS_BLOCKS MOTIFS SFScan

【表 6】

10

20

30

40

表3-3

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、モチーフ、 及びドメイン	分析方法及び データベース
5	5544089CD1	367	S108 S161 S197 S203 S235 S266 S361 S49 T180 T263 T316 T331	N139 N142 N308	UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659 P40818 782-1103: L36- L328 ユビキチンC末端ヒドロラーゼファミリー BL00972: Y74-L83, V120-C134, Y274-N298, G301-K322 ユビキチンC末端ヒドロラーゼファミリー UCH-2: E270-Q332 ユビキチンC末端ヒドロラーゼファミリー UCH_2_2: Y274-Y292 シグナル_切断: M1-S16 TRYPSIN DM00018 S55065 26-244: A27-T231 セリンプロテアーゼ、トリプシンファミリー シグネチャBL00134: C40-C56, V215-I228 I 型フィブロネクチンドメイン BL01253: C40-P53, I197-T231 Kringledメインタンパク質 BL00021: S159-Q164, C40-Y57 CHYMOTRYPSIN SERINE PROTEASE PR00722A: V41-C56, A95-A109 セリンプロテアーゼ、トリプシンファミリー、 活性部位 トリプシン_his.prf: S35-T76 トリプシン: G42-V178, G216-I228 ロイシンジッパー: I44-I65 シグナル_切断: M1-S19	BLAST_DOMO BLIMPS_BLOCKS HMMER_PFAM MOTIFS SPScan BLAST_DOMO BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_PRINTS PROFILES SCAN HMMER PFAM MOTIFS SPScan
6	7474081CD1	235	S134 S143 S159 S171 S226 T231	N90		

10

20

30

40

【表 7】

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	潜在断片化部位	潜在的 グリコシル化 部位	シグネチャ(signature)配列、モチーフ、 及びドメイン	分析方法及び データベース
7	5281209CD1	488	S13 T74 S186 S233 T363 T456 T4 T34 T125 S170 S172 T178 S249 T337 S387 S389 S419 T447	N119 N184 N243 N334	腎ジペプチダーゼタンパク質	BLIMPS-BLOCKS
					BL00869:P92-L247, E280-R412, S415-N457	
					DIPEPTIDASE MICROSOVAL PRECURSOR MDP	BLAST-PRODOM
					HYDROLASE MICROSOVAL SIGNAL GPI-ANCHOR	
					GLYCOPROTEIN ZINC PD005626: S143-E450	
					RENAL DIPEPTIDASE DM02775: T77-R410	BLAST-DOMO
					腎ジペプチダーゼ: V195-R217	MOTIFS
8	2256251CD1	346	S203 S210 S266 S45 S79 T131 T147 T216	N110	腎dipeptidase: A63-V475	HMME PFAM
					シグナルペプチド: M1-A36	HMME
					シグナル切断: M1-A36	SPSCAN
					TRYPSIN DM00018: P15944 31-270: I63-I294	BLAST-DOMO
					PROTEASE SERINE PRECURSOR SIGNAL	BLAST-PRODOM
					HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY	
					MULTIGENE FACTOR PD000046: I156-I290, I63-F178, N288-F314, P274-P305	
					セリンプロテアーゼ、トリプシンファミリー	BLIMPS_BLOCKS
					BL00134: C88-C104, D241-I264, P277-I290	
					I型フィブロネクチンドメイン BL01253: C88- A101, V158-E194, G240-C253, W259-H293	BLIMPS_BLOCKS
					Kringledメインタンパク質 BL00021: C88-F105, V169-G190, G249-I290	BLIMPS_BLOCKS
					CHYMOTRYPSIN SERINE PROTEASE PR00722: G89-C104, G146-V160, G240-V252	BLIMPS_PRINTS
					膜貫通ドメイン: P308-L328	HMME
					トリプシン: I63-I290	HMME PFAM
					トリプシン_His L99-C104	MOTIFS
					トリプシンファミリーセリンプロテアーゼ	PROFILES CAN
					活性部位 トリプシン_ser.prf: L80-H128 トリプシン_ser.prf: L229-R273	
					シグナル切断: M1-S45	SPSCAN

10

20

30

40

表3-5

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 糖コシ化 部位	シグネチャ(signature)配列、モチーフ、 及びドメイン	分析方法及び データベース
9	716054CD1	882	S115 S133 S293 S312 S412 S443 S479 S530 S587 S588 S723 S80 S850 T227 T234 T307 T326 T499 T52 T551 T594 T603 T615 T776 Y315 Y36 Y55 Y555 Y844		PROLYL ENDOPEPTIDASE FAMILY SERINE DM02461, P27487 192-765; F488-E870, G251-E370 DIPEPTIDYL IV HYDROLASE PROTEASE SERINE PEPTIDASE DIPEPTIDASE TRANSMEMBRANE GLYCOPROTEIN PD003048: L744-E870 DIPEPTIDYL IV HYDROLASE PROTEASE SERINE PEPTIDASE DIPEPTIDASE TRANSMEMBRANE GLYCOPROTEIN PD003086: Y423- V661, I212-T326 プロリルエンドペプチダーゼファミリー BL00708: G501-I513, Q714-L744 ジペプチルペプチダーゼ IV PF00930: I498- R508, F756-P783, R808-L828 PROLYL OLIGOPEPTIDASE SERINE PROTEASE PR00862: P647-F665, G737-R757 プロリルオリゴペプチダーゼファミリー ペプチダーゼ_S9: R672-L744 ジペプチルペプチダーゼ IV DPPIV_N-term: M88-N663	BLAST_DOMO BLAST_PRODUM BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_PFAM BLIMPS_PRINTS HMMER_PFAM HMMER_PFAM

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、モチーフ、 及びドメイン	分析方法及び データベース
10	7477386CD1	1189	S132 S169 S200 S32 S323 S350 S445 S480 S511 S626 S675 S699 S798 S1064 T247 T362 T521 T612 T718 T777 T946 T986 T1104 Y552	N109 N478 N944	do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN DM00368 Q05910 189-395:I261- P463 THROMBOSPONDIN TYPE 1 REPEAT DM00275 P07996 477-540: D555-C604 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE C02B4.1 A DISINTEGRIN METALLOPROTEASE PD013511: I474-E549 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q647-C716 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR PD007018: W854-Q974, W914- C1029, W558-K623 METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMERANE ADHESION PD000791:P256- P463 中性血結メタロペプチダーゼ BL00142: T398-N408 シグナルペプチド: M1-A22 Reprolysinファミリーペプチド Pep_M12B_propep: R120-V240 Reprolysin (M12B)ファミリー血結 メタロペプチダーゼ Reprolysin: I261-P463 トロンボスポンジン型ドメイン tsp_1: A973-C1024, S559-C609, Y852-C909, W914- C971 シグナル切断: M1-G23	BLAST_DOMO BLAST_DOMO BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLIMPS_BLOCKS HMMER HMMER_PFAM HMMER_PFAM HMMER_PFAM

10

20

30

40

表3-7

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、モチーフ、 及びドメイン	分析方法及び データベース
11	7473089CD1	952	S19 S203 S207 S303 S346 S432 S492 S575 S578 S611 S666 S682 S708 S745 S919 T171 T288 T317 T325 T337 T359 T471 T594 T687 T765	N141 N591 N623 N680	GO ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN DM00368 JC2550 1-201;R218-P427 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161; K684-E804 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654; V610-C683 METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791; V214- P427 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE C02B4.1 A DISINTEGRIN METALLOPROTEASE PD013511; L437-V505 中性垂給メタロペプチダーゼ BL00142; T358-N368 シグナル ペプチド : M1-G18 Reprolysinファミリープロペプチド Pep_M12B_propep: H67-N181 Reprolysin (M12B)ファミリー--垂給 メタロプロテアーゼ Reprolysin: R218-P427 トロンボスポンジン1型ドメイン tsp_1; A520-C570, W845-C896, W899-C952 垂給 プロテアーゼ: T358-F367 SpScanシグナル 切断: M1-G17	BLAST_DOMO BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLINDS_BLOCKS HMMER HMMER_PFAM HMMER_PFAM HMMER_PFAM MOTIFS SPSCAN

10

20

30

40

表3-8

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、モチーフ、 及びドメイン	分析方法及び データベース
12	7604035CD1	898	S187 S188 S258 S268 S285 S415 S467 S547 S696 S796 S819 S851 S892 T106 T198 T35 T434 T483 T492 T5	N3 N490 N773	PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE PD007018: Y726- C841, W786-A874, Y667-C778, W50-Q72, S368-Q383 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q416-C484 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: R485-I599 シグナル ペプチド: M1-D24 トロンボスポンジン[型ドメインEsp.1: G48-R87, W727-C783, E787-C841 シグナル切断: M1-D24	BLAST_PRODOR BLAST_PRODOR BLAST_PRODOR
13	3473847CD1	631	S117 S160 S174 S185 S188 S268 S28 S30 S358 S431 S503 S605 T142 T33 T346 T512 T606	N472	Subtilaseファミリーペプチダーゼ_S8: S86-N364 SERINE PROTEASES, SUBTILASE FAMILY, HISTIDINE DM00108 P80146 150-377: G116- T346 セリンプロテアーゼ, subtilaseファミリー、 BL00136: L123-I135, D163-G175, G323-G333 SUBTILISIN SERINE PROTEASE FAMILY PR00723: G116-I135, K161-S174, S322-M338 セリンプロテアーゼ, subtilaseファミリー、 活性部位 subtilase_ser.prf: A302-Q352	SPSCAN HMMER_PFBM BLAST_DOMO BLIMPS_BLOCKS BLIMPS_PRINTS PROFILESAN

表3-9

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、モチーフ、 及びドメイン	分析方法及び データベース
14	3750004CD1	470	S454 S51 S54 T104 T276 T386 T464	N182 N203	トロンボスポンジン1型ドメイン tsp_1: T34-C81 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q119-C185 シグナル_ペプチド: M1-G29 シグナル_切断: M1-G24	HMMER_PFBM ELAST_PRODUM
15	4904126CD1	110	S16 S36 T100 T49 N47		カスバーゼ補完ドメイン CARD: A2-A91 INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM07463 P29466 1-122: M1-S110	HMMER_PFBM ELAST_DOMO

表3-10

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、モチーフ、 及びドメイン	分析方法及び データベース
16	71268415CD1	879	S132 S14 S208 S288 S571 S711 S747 S754 S755 S827 T106 T118 T29 T30 T373 T412 T42 T428 T444 T55 T688 Y167 Y39	N368 N371 N569 N68	Reprolysinファミリーポリペプチド Pep_M12B_procp: R8-K119 Reprolysin (M12B)ファミリー-亜種 メタロプロテアーゼ Reprolysin: K134-P332 Disintegrin: E349-Q424 Go-ZINC; METALLOPEPTIDASE; NEUTRAL; ATFOLYSIN; DM00368 S60257 204-414: K126- D333 Go ZINC; REGULATED; EPIDIDYMAL; NEUTRAL; DM00591 S60257 492-628: F410-L549 MELTRIN, BETA METALLOPROTEASE DISINTEGRIN BETA INTEGRIN PROTEASE METALLOPROTEASE PD105322: P620-G812 METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791: K134- P332 MELTRIN, BETA METALLOPROTEASE DISINTEGRIN MELTRIN BETA INTEGRIN PROTEASE METALLOPROTEASE PD171676: K495-C567 CELL ADHESION PLATELET BLOOD COAGULATION VENOM DISINTEGRIN METALLOPROTEASE PRECURSOR SIGNAL PD000664: E349-Y423 中性亜種メタロプロテアーゼ BL00142: T266-G276 DISINTEGRIN SIGNATURE PR00289: C380- R399, E409-N421 NEPRILYSIN METALLOPROTEASE PR00786C: N259-P275	HMMER_PFAM HMMER_PFAM HMMER_PFAM ELAST_DOMO ELAST_DOMO ELAST_PRODOM ELAST_PRODOM ELAST_PRODOM ELAST_PRODOM ELIMPS_BLOCKS ELIMPS_PRINTS ELIMPS_PRINTS

10

20

30

40

表3-11

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、モチーフ、 及びドメイン	分析方法及び データベース
					Disintegrinsシグネチャdisintegrins.prf: E360-P419 中性亜鉛メタロプロテアーゼ、亜鉛結合領域 シグネチャ 亜鉛プロテアーゼ.prf: S249-G301 亜鉛ドメイン: V624-Y645 亜鉛結合領域 Zinc_Protease: T266-F275	PROFILES CAN PROFILES CAN HMMER MOTIFS
17	7473301CD1	850	S100 S275 S295 S358 S429 S448 S470 S474 S495 S536 S596 S64 S787 S802 S807 T117 T250 T312 T348 T382 T404 T426 T570 T714 T777	N19 N210 N422 N486 N533 N559 N568	TRYPSIN FAMILY SERINE PROTEASE trypsin: I613-I842 低密度リポタンパク質受容体ドメイン ldl_recept_a: Q489-S527, P530-Q562, I564-C603 TRYPSIN DM00018 P98072 800-1033: R612- V846 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: Q675-I842, I613-G809 セリンプロテアーゼ、トリプシンファミリー BL00134: C638-C654, D791-T814, P829-I842 I型フィブロネクチンドメインBL01253: C638-A651, R790-C803, W811-Y845 Kringletドメインタンパク質BL00021: I722-G743, L801-I842, C638-F655 LOW DENSITY LIPOPROTEIN RECEPTOR DOMAIN PRO0261: G501-E522 CHYMOTRYPSIN SERINE PROTEASE PRO0722: G639-C654, T697-W711, R790-S802	HMMER_PFAAM HMMER_PFAAM BLAST_DOMO BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS

10

20

30

40

表3-12

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、モチーフ、 及びドメイン	分析方法及び データベース
					トリプシンファミリー-セリンプロテアーゼ活性部位 trypsin_his.prf: I630-K679 trypsin_ser.prf: I776-R825	PROFESCAN
					共通ドメイン: I77-L95	HMMP
					トリプシンファミリー-セリンプロテアーゼ活性部位 Trypsin_His I649-C654 Trypsin_Ser D791-S802	MOTIFS
18	7473308CD1	254	S136 S14 S153 S195 S227 T230 T249		TRYPSIN FAMILY SERINE PROTEASE trypsin: I21-Q183 CHYMOTRYPSIN SERINE PROTEASE FAMILY PR00722B: T89-A103 TRYPSIN DM00018/P07478/24-242: I21-Q183 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZMOGEN GLYCOPROTEIN FAMILY MULTI-GENE FACTOR PD000046: G23-Q183	HMMP PFAM BLIMPS_PRINTS BLAST_DOMO BLAST_PRODOM

10

20

30

40

表3-13

SEQ ID NO:	インサイト ポリペプチドID	アミノ酸 残基数	潜在断リン酸化部位	潜在的 グリコシル化 部位	シグネチャ(signature)配列、モチーフ、 ヘモペグシドメイン	分析方法及び データベース
19	7478021CD1	568	S142 S145 S153 S172 S177 S190 S244 S316 S34 S420 S448 S552 T209 T22 T293 T334 T401 T427 T489 T79 Y509	N371	Matrixinペプチダーゼ M10: F56-T266 ヘモペグシドメイン: F332-T390, I393-S448, L450-Q498, I505-K548 MATRIXINS CYSTEINE SWITCH DM00558 P22757 100-337: A184-T334, P76- P124 MATRIXINS CYSTEINE SWITCH DM00558 P08254 29-274: Q158-T334, L85- M122 MATRIX METALLOPROTEINASE PD168921: S327- N392 MATRIX METALLOPROTEINASE PD169970: A494- M568 MATRIX PRECURSOR METALLOPROTEINASE HYDROLASE ZINC ZYMOGEN CALCIUM COLLAGEN DEGRADATION SIGNAL PD000673: F171-T266, P73-M122 Matrixins システインスイッチ BL00546: F92- D121, V224-P267, G273-Y304, L313-G326, F443-Y455, F409-E428 ヘモペグシドメインタンパク質 BL00024: M112- M122, G273-Y304, L313-G326, F443-Y455, Y408-D419 MATRIXIN SIGNATURE PRO0138: M112-P125, E198-F213, V224-W252, V279-Y304, L313- G326 Matrixins システインスイッチ システイン_switch.prf: A95-M204 中性重鎖メタロペプチダーゼ 重鎖結合領域シグネチャ 重鎖_プロテアーゼ.prf: D256-E312	HMMER_Pfam HMMER_Pfam BLAST_DOMO BLAST_DOMO BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_PRINTS PROFILESSCAN PROFILESSCAN

10

20

30

40

表3-14

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、モチーフ、 及びドメイン	分析方法及び データベース
					ヘモキシンドメインシグネチャ ヘモキシシ. prf: F409-R477 シグナルペプチド: M1-P21 亜結合領域 亜結合プロテアーゼ: V279-L288 シグナル切断: M1-P24 TRYPSIN FAMILY SERINE PROTEASE trypsin: I56-I298 TRYPSIN DM00018 Q05319 543-784; I56-I302 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: S117-I298, I56-G192 セリンプロテアーゼ、トリプシンファミリー BL00134: C81-C97, D238-G261, P285-I298 I 型フィブロネクサンドメイン BL01253: C81-A94, G154-E190, R237-C250 Kringledメインタンパク質 BL00021: C81-I98, I165-G186, S247-F288 CHYMOTRYPSIN SERINE PROTEASE PR00722: G82-C97, P142-F156, R237-M249 トリプシンファミリーセリンプロテアーゼ活性部位 trypsin_his. prf: I73-G122 trypsin_ser. prf: K225-R271 トリプシンファミリーセリンプロテアーゼ活性部位 Trypsin_His: I92-C97 Trypsin_Ser: D238-M249 シグナル切断: M1-S26	PROFILES CAN HMMER MOTIFS SPScan HMMER PFAM BLAST DOMG BLAST PRODOM BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_PRINTS PROFILES CAN MOTIFS SPScan
20	4333459CD1	306	S117 S138 S2 S223 S60 S72 T110 T139 T207 T217	N108		

10

20

30

40

表3-15

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、モチーフ、 及びドメイン	分析方法及び データベース
21	6817347CD1	953	S102 S114 S150 S172 S369 S429 S47 S623 S794 S804 S808 S831 S856 S919 S942 T289 T42 T455 T488 T544 T567 T568 T585 T59 T736 T777 T786 T839 Y929	N95	ユビキチンC末端ヒドロラゼファミリー 1 UCH-1: E593-D624 ユビキチンC末端ヒドロラゼファミリー 2 UCH-2: N875-K935 UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659 P40818 782-1103: T777-L931, L598-H709, I713-T753, V101- L128 PROTEASE UBIQUITIN HYDROLASE UBIQUITIN SPECIFIC ENZYME DEUBIQUITINATING C- TERMINAL THIOLESTERASE PROCESSING CONJUGATION PD017412: T777-B859 ユビキチンC末端ヒドロラゼファミリー 2 BL00972: G594-L611, Y675-L684, I714-C728, X878-H902, K904-D925 ユビキチンC末端ヒドロラゼファミリー 2 シグネチャ 1 Uch_2_1: G594-Q609 ユビキチンC末端ヒドロラゼファミリー 2 シグネチャ 2 Uch_2_2: Y879-Y896	HMER_PFAM HMER_PFAM BLAST_DOMO BLAST_PRODOM BLIMPS_BLOCKS MOTIFS MOTIFS

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	選択された断片	配列断片	5'位置	3'位置
22	275791CB1	2204	1168-1197, 1503-1522, 1-281, 1716-1738	6456514H1 (COLNDIC01)	890	1510
				7246467T8 (PROSTMY01)	692	1390
				4943009F8 (BRAIFEN05)	1319	1877
				55047202J1	1	811
				6053385H1 (BRAEDIR03)	1636	2204
23	1389845CB1	2036	1-392, 1460-1491, 1334-1400, 1974-2036	FL1389845_00001	6	2036
				1389845H1 (EOSINOT01)	1	244
24	1726609CB1	2185	1-44, 1804- 2185	71762189V1	1	562
				5426388F9 (PROSTMY07)	1370	1992
				71053940V1	744	1373
				71041539V1	1954	2185
				5968441H1 (BRAZNOT01)	1326	1824
25	4503848CB1	3486	1-1330	6756865J1 (SINTFER02)	642	1316
				2053131H1 (BEPINOT01)	2885	3136
				6440674H1 (BRAENOT02)	2576	3095
				95745066	1831	2254
				7191212H2 (BRATDIC01)	2295	2873
				5960039H1 (BRATNOT05)	1229	1797
				GBL.g7710158_edit	1	2015
				60200050D1	825	1146
				5969176H1 (BRAZNOT01)	1677	2104
				5649471H1 (BRAITUT23)	3157	3486
				2232143F6 (PROSNOT16)	2235	2670
				3022114H1 (PROSDIN01)	3116	3402
				60220456D1	1085	1452

10

20

30

40

表4-2

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	選択された断片	配列断片	5'位置	3'位置
26	5544089CB1	2847	1260-1631, 2332-2847, 408-914	55051688J1	810	1767
				2344450P6 (TESTTUT02)	2195	2847
				7658834H1 (OVARNOE02)	2130	2642
				71763578V1	1562	2266
				6576463H1 (COLHUTUS02)	1079	1838
27	7474081CB1	890	1-21	55051680J1.comp	1	958
				g2103202	1	493
28	5281209CB1	1577	1-629	g2142177	397	890
				FLS281209_g7712102_000	1	1467
				004_g436191		
				g3644494	1155	1577
				3142983R6 (HNT2AZS07)	1045	1576
29	2256251CB1	1958	1-399, 896-935	3220504T6 (COLNNON03)	1611	1958
				g4264312	400	848
				FL2256251_g7708357_000	910	1830
				002_g6103629		
				2256251R6 (OVARTUT01)	408	1003
30	7160544CB1	3106	1-540, 1166-1428	6471337H1 (PLACFBB01)	1654	2316
				6854305H1 (BRAIFEN08)	895	1561
				4443368H1 (SINDNOT01)	1332	1585
				6894004J1 (BRAITDR03)	470	1077
				7655990H1 (UTREDME06)	324	822
				7745974H1 (ADRETTUE04)	2378	3052
				7160544H1 (HNT2TXC01)	1	427
				70490289V1	2636	3106
				70748463V1	2269	2891
				7745974J1 (ADRETTUE04)	1566	2153

【表 2 1】

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	選択された断片	配列断片	5'位置	3'位置
31	7477386GB1	3567	1-971, 1953-2846, 3243-3567	GB1:g6682143_000029_ed it.20231-20345 GB1:g6682143_000023.co mp_edit.11365-11445 GB1:g6682143_000027_ed it.14110-14265 GB1:g6682143_000029_ed it.35032-35202 GB1:g6682143_000029_ed it.13651-13803 GB1:g6682143_000019_ed it.3461-3655 GB1:g6682143_000029_ed it.41513-41644 GB1:g6682143_000029_ed it.43912-44404 GB1:g6682143_000029_ed it.12846-12930 GB1:g6682143_000029_ed it.15804-15912 GB1:g6682143_000023.co mp_edit.9253-9693 GB1:g6682143_000029_ed it.27621-27746 GB1:g6682143_000029_ed it.18722-18853 GB1:g6682143_000029_ed it.17438-17581 GB1:g6682143_000029_ed it.35651-35783	1495 1 523 2272 958 679 2944 3076 874 1111 82 2068 1363 1219 2608	1608 81 678 2436 1110 873 3075 3567 957 1218 522 2193 1494 1362 2739

表4-4

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	選択された断片	配列断片	5'位置	3'位置
22	275791CB1	2204	1168-1197, 1503-1522, 1-281, 1716-1738	6456514H1 (COLNDIC01)	890	1510
				7246467T8 (PROSTMY01)	692	1390
				4943009F8 (BRAIFEN05)	1319	1877
				55047202J1	1	811
				6053385H1 (BRAEDIR03)	1636	2204
23	1389845CB1	2036	1-392, 1460-1491, 1334-1400, 1974-2036	FL1389845_00001	6	2036
				1389845H1 (EOSINOT01)	1	244
24	1726609CB1	2185	1-44, 1804- 2185	71762189V1	1	562
				5426388F9 (PROSTMY07)	1370	1992
				71053940V1	744	1373
				71041539V1	1954	2185
				5968441H1 (BRAZNOT01)	1326	1824
25	4503848CB1	3486	1-1330	6756865J1 (SINTFER02)	642	1316
				2053131H1 (BEPINOT01)	2885	3136
				6440674H1 (BRAENOT02)	2576	3095
				95745066	1831	2254
				7191212H2 (BRATDIC01)	2295	2873
				5960039H1 (BRATNOT05)	1229	1797
				GBL.g7710158_edit	1	2015
				60200050D1	825	1146
				5969176H1 (BRAZNOT01)	1677	2104
				5649471H1 (BRAITUT23)	3157	3486
				2232143F6 (PROSNOT16)	2235	2670
				3022114H1 (PROSDIN01)	3116	3402
				60220456D1	1085	1452

10

20

30

40

表4-5

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	選択された断片	配列断片	5'位置	3'位置
26	5544089CB1	2847	1260-1631, 2532-2847, 408-914	55051688J1 2344450P6 (TESTTUT02) 7658834H1 (OVARNOE02) 71763578V1 6576463H1 (COLHUTUS02) 55051680J1.comp g2103202 g2142177 FLS281209_g7712102_000 004_g436191 g3644494	810 2195 2130 1562 1079 1 1 397 1 1155	1767 2847 2642 2266 1838 958 493 890 1467 1577
27	7474081CB1	890	1-21	3142983R6 (HNT2AZS07) 3220504T6 (COLNNON03) g4264312 FL2256251_g7708357_000 002_g6103629	1045 1611 400 910	1576 1958 848 1830
28	5281209CB1	1577	1-629	2256251R6 (OVARTUT01) 6471337H1 (PLACFEB01) 6854305H1 (BRAIFEN08) 4443368H1 (SINDNOT01) 6894004J1 (BRAITDR03) 7745974H1 (UTREDME06) 7745974H1 (ADRETTUE04) 7160544H1 (HNT2TXC01) 70490289V1 70748463V1 7745974J1 (ADRETTUE04)	408 1654 895 1332 470 324 2378 1 2636 2269 1566	1003 2316 1561 1585 1077 822 3052 427 3106 2891 2153
29	2256251CB1	1958	1-399, 896-935			
30	7160544CB1	3106	1-540, 1166-1428			

10

20

30

40

表4-6

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	選択された断片	配列断片	5'位置	3'位置
31	7477386GB1	3567	1-971, 1953-2846, 3243-3567	GB1:g6682143_000029_ed it.20231-20345 GB1:g6682143_000023.co mp_edit.11365-11445 GB1:g6682143_000027_ed it.14110-14265 GB1:g6682143_000029_ed it.35032-35202 GB1:g6682143_000029_ed it.13651-13803 GB1:g6682143_000019_ed it.3461-3655 GB1:g6682143_000029_ed it.41513-41644 GB1:g6682143_000029_ed it.43912-44404 GB1:g6682143_000029_ed it.12846-12930 GB1:g6682143_000029_ed it.15804-15912 GB1:g6682143_000023.co mp_edit.9253-9693 GB1:g6682143_000029_ed it.27621-27746 GB1:g6682143_000029_ed it.18722-18853 GB1:g6682143_000029_ed it.17438-17581 GB1:g6682143_000029_ed it.35651-35783	1495 1 523 2272 958 679 2944 3076 874 1111 82 2068 1363 1219 2608	1608 81 678 2436 1110 873 3075 3567 957 1218 522 2193 1494 1362 2739

【表 2 5】

10

20

30

40

表4-7

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	選択された断片	配列断片	5'位置	3'位置
32	7473089CB1	2930	1-632, 1082-1138, 2453-2555, 1373-1615, 1716-1740	GBI:g6682143_000029_ed it.27099-27237	1759	1932
				GBI:g6682143_000029_ed it.24540-24713	1609	1758
				GBI:g6682143_000029_ed it.37355-37558	2740	2943
				GBI:g7387384_000011.co mp_edit.13491-13892	2529	2930
				GBI:g7387384_000010_ed it.2924-3211	1335	1619
				GBI:g7387384_000010_ed it.13479-13850	2157	2528
				GBI:g7387384_000010_ed it.11350-11535	1794	1979
				7988641H1 (UTRSTUC01)	1064	1587
				GBI:g7387384_000010_ed it.9694-9867	1620	1793
				GBI:g7387384_000012.co mp_edit.9639-10595	75	1031
				GBI:g7387384_000010_ed it.1917-2074	1032	1163
				GBI:g7387384_000010_ed it.2514-2684	1164	1334
				7631548J1 (BRAFTUE03)	21	619
				GBI:g7387384_000010_ed it.11639-11815	1980	2156
				GBI:g7387384_edit	1	2930

【表 2 6】

10

20

30

40

表4-8

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	選択された断片	配列断片	5'位置	3'位置
33	7604035CB1	4230	4185-4230, 894-2774	6254235H1 (LUNPTUT02)	3308	3897
				6213818H1 (MUSCDIT06)	3900	4230
				6314348H1 (NERUTDN03)	3426	3994
				7195502H1 (LUNGER04)	2758	3376
				6800634J1 (COLENOH03)	2678	3323
				8113675H1 (OSTEUNC01)	1661	2049
				6804059H1 (COLENOH03)	478	1050
				7750274H1 (HEAONOE01)	1995	2503
				3843227F6 (DENDNOT01)	2094	2707
				7632961H1 (BLADTUE01)	1418	2024
				5509797J1	683	1535
				7716357J1 (SINTFEE02)	1	678
				71906145V1	1340	2118
34	3473847CB1	3699	1-2631	7101935F8 (BRAWTR02)	166	760
				70857826V1	2757	3441
				70855756V1	2650	3239
				820867R1 (KERANOT02)	2166	2732
				8055446J1 (ESOGTUE01)	579	1013
				70857738V1	3156	3699
				70858612V1	1998	2671
				GNN.g7208751_000002_00 2.edit	555	1850
				7101935R8 (BRAWTR02)	1	473

【表 2 7】

10

20

30

40

表4-9

ポリスクレオナド SEQ ID NO:	インサイト ポリスクレオナドID	配列長	選択された断片	配列断片	5'位置	3'位置
35	3750004CB1	2410	1-264, 2116-2410, 1057-1167, 1590-1649	g7712021_ed1t	1	246
				7680089J1 (BRAFTUE01)	1327	1911
				6804411H1 (COLENOR03)	1088	1618
				71909368V1	536	968
				g1187194	1655	2127
				g2241985	706	1144
				6314962H1 (NBRDTN03)	973	1135
				6823371J1 (SINTNOR01)	65	855
				7655009J1 (UTREDNE06)	1407	1990
				g1272147	1754	2410
36	4904126CB1	549		71620969V1	1	549
37	71268415CB1	2755	1-1097, 2326-2755	7715927J1 (SINTFEE02)	590	1340
				7372052H2 (BRAIPEE04)	2044	2514
				g6651070_CD	102	2755
				7723192J2 (THYRDIE01)	905	1500
				GBI:g7709257_000011.ed 1t	1	139
				8037549H1 (SMCRUNE01)	206	819
				7720289J1 (THYRDIE01)	1596	2263
				8037549J1 (SMCRUNE01)	1456	2120

【表 2 8】

表4-10

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	選択された断片	配列断片	5'位置	3'位置
38	7473301CBI	2553	1-2394	GBI.g7272157_000017.ed it	2001	2553
				71704195V1	1713	2016
				5544473H1 (TESTNOC01)	622	680
				GNN.g7272157_000017_00 2.edit	1688	2382
				5544473T8 (TESTNOC01)	2246	2550
				71703469V1	1163	1746
				GNN.g8571511_000004_00 2.edit	981	1468
39	7473308CBI	1041	826-1041, 1-299	GNN.g6624046_000008_00 4	1	1111
				GNN.g1552511_035	1	1041
40	7478021CBI	1707	1-1186	g8176728_edit	979	1083
				g7684439_edit	1	978
41	4333459CBI	1262	1-1262	g7684439_edit_2	1084	1707
				71571956V1	704	1262
				5634861R8 (PLACFER01)	1	266
				71571988V1	256	937
42	6817347CBI	3067	1-2270	71573159V1	247	928
				55022864H1	2314	3067
				55022792H2	1392	2091
				55022814H1	2080	2886
				55022795J2	2044	2726
				GNN.g7417337_004_edit	1	3067

【表 2 9】

10

20

30

40

表5

ポリスクレオサド SEQ ID NO:	インサイト プロジェクトID	代表的ライブラリ
22	275791CB1	TESTNOT03
23	1389845CB1	ECITXT01
24	1726609CB1	ERAFUT02
25	4503848CB1	PROSNOT16
26	5544089CB1	ERAFEC01
28	5281209CB1	HNT2AZS07
29	2256251CB1	OVARFUT01
30	7160544CB1	ERAFNOT02
32	7473089CB1	UTRSTUC01
33	7604035CB1	PLACNOR01
34	3473847CB1	KERANOT02
35	3750004CB1	ERAFUE01
36	4904126CB1	TYMNOT08
37	71268415CB1	THYRDIE01
38	7473301CB1	TESTNOC01
41	4333459CB1	KIDCTMT01
42	6817347CB1	ADRETUR01

表 6-1

ライブラリの説明		
ライブラリ	ベクター	
ADRETUR01	PCDNA2.1	このランダムプライムライブラリは、52歳白人男性の腎尿管切除および腎病変の局部破壊の際に採取した副腎腫瘍組織の左上極から単離したRNAを用いて作製した。病理学的には、副腎実質のほぼ全体および隣接した脂肪組織に及ぶ腫瘍を形成したグレード3の副腎皮質癌であった。転移性腫瘍結節が肺門領域に見られた。腎静脈にも腫瘍が及び、腎静脈の切除した縁にも腫瘍の発生が見られた。副腎皮質癌および血管の断片が下大静脈に見られた。病歴には、過度の体重の減少があった。家族歴には、皮膚癌、1型糖尿病、および神経性の抑うつ症が含まれていた。
BRAFNOT02	pINcy	このライブラリは、心不全で死亡した35歳の白人男性から採取した上前頭皮質から単離したRNAを用いて作製した。病理学的には、中程度の軟膜線維症及び大脳新皮質の複数の微小梗塞症であった。大脳半球全体に亘る収縮したやや好酸性の錐体ニューロンの証拠が見つかった。加えて、大脳皮質全体に亘って、回りがグリオオシスの複数の顕微鏡的空洞化小領域が見られた。病歴には、拡張型心筋症、うっ血性心不全、心臓肥大症、並びに脾臓及び肝臓の肥大があった。
BRAFTUE01	PCDNA2.1	この5' バイアスランダムプライムライブラリは、58歳白人男性の脳髓膜病変の切除術の際に前頭葉から採取した脳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、グレード2の転移性副腎腫であった。患者には偏頭痛があった。患者に脳溢血及び肺水腫が起こり、この入院中に死亡した。患者の病歴には、グレード2の腎細胞癌、不眠症、及び慢性的な気道閉塞があった。また、以前に腎尿管切除術を受けている。投薬には、Decadron及びジラサンチンが含まれていた。家族歴には、患者の父の腎臓の悪性新生物があった。
BRAIFEC01	pINcy	この大きいサイズに断片化されたライブラリは、左心低形成症により妊娠23週で死産した白人男子胎児から採取した脳組織から単離したRNAを用いて作製した。

【表 3 1】

10

20

30

40

表6-2

ライブラリ	ベクター	ライブラリの説明
BRA1TUT02	PSPORT1	このライブラリは、58歳の白人男性の前葉から脳腫瘍の病巣切除の際に採取した脳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、グレート2転移性副腎腫を示していた。患者の病歴には、グレート2腎細胞癌、不眠症、及び慢性気道閉塞があった。家族歴には、腎臓の悪性新生物があった。
EOS1TX01	pINCY	このライブラリは、IL-5を用いて刺激した好酸球から単離したRNAを用いて作製した。
HNT2AZS07	PSPORT1	このライブラリは、0.35 μ molのAZで3日間処理したHNT2細胞系（委任ニコロロン前駆体の特性を示すヒト奇形癌由来）より単離したRNAを用いて作製した。サブトラククション用のハイブリダイゼーションプロトコルは、未処理のhNT2細胞から同様にして作製したライブラリから得た。AZ処理ライブラリからの3.08 $\times 10^6$ 個のクローンを、未処理のライブラリからの3.04 $\times 10^6$ 個のクローンでサブトラククションハイブリダイゼーションを3回行った。サブトラククションハイブリダイゼーションの条件は、Swaroopら (NAR (1991) 19:1954) および Bonaldoら (Genome Research (1996) 6:791)の方法に基づいて決定した。
KERAN0102	PSPORT1	このライブラリは、乳房の表皮細胞 (NHEK) から単離したRNAを用いて作製した。NHEK (Clontech #CC-2501) は、30歳の黒人女性の乳房縮小手術の際に採取されたヒト乳房表皮細胞系である。
KIDCTMT01	pINCY	このライブラリは、65歳の男性の腎尿管切除術の際に腎皮質組織より単離したRNAを用いて作製した。病理学的には、関連する腫瘍組織は、腎臓及び腎臓の中央部内における腎細胞癌を示していた。
OVARTUT01	PSPORT1	このライブラリは、43歳の白人女性から卵巣及び卵巣の摘出の際に採取した卵巣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、左卵巣全体に係わるグレート2粘液性嚢胞腺癌を示していた。患者の病歴には、僧帽弁疾患、肺炎、及びウイルス性肝炎があった。家族歴には、アテローム性冠動脈疾患、膵臓癌、ストレス反応、脳血管障害、乳癌、及び子宮癌が含まれていた。

10

20

30

40

表 6-3

ライブラリ	ベクター	ライブラリの説明
PLACNOR01	PCDNA2.1	このランダムプライムライブラリは、2人のドナーからのプールされたcDNAを用いて作製した。cDNAは、妊娠16週で子宮内死亡及び水頭症により死亡した白人胎児（ドナーA）から採取した胎盤組織と、妊娠18週で子宮内死亡により死亡した白人女子胎児（ドナーB）から採取した胎盤組織とから単離したmRNAを用いて作製した。ドナーAの病歴には、臍帯の巻き付き（頭3回及び肩1回）が含まれていた。血清学的には、抗CMVに対しては陽性、他は陰性であった。家族歴には、ドナーの母の多胎妊娠及び出産、中絶が含まれていた。ドナーBは、血清学的には陰性であった。
PROSNOT16	pINCY	このライブラリは、68歳の白人男性の根治前立腺切除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、腺癌（Gleasonグレード3+4）を示していた。関連する腫瘍組織は病理学的には、腺癌（Gleasonグレード3+4）を示していた。患者は、前立腺特異抗原（PSA）の値が高く、入院中に重症筋無力症と診断された。患者の病歴には、変形性関節症及びII型糖尿病が含まれていた。家族歴には、良性高血圧、急性心筋梗塞、高脂血症、及びアテローム硬化性冠動脈疾患が含まれていた。
TESTNOC01	PBLUESCRIPT	この大きいサイズに断片化されたライブラリは、10歳から61歳までの11人の白人男性のプールから採取した精巢組織から単離したRNAを用いて作製した。
TESTN0T03	PBLUESCRIPT	このライブラリは、肝臓病で死亡した37歳の白人男性から採取した精巢組織から単離したRNAを用いて作製した。患者の病歴には、肝硬変、黄疸及び肝不全が含まれる。
THYRD1B01	PCDNA2.1	この5'バイアスランダムプライムライブラリは、22歳白人女性の非切開甲状腺生検、甲状腺部分切除、及び所属リンパ節切除の際に採取した病変甲状腺組織から単離したRNAを用いて作製した。病理学的には、腺癌形成であった。患者には、甲状腺の悪性新生物があった。患者の病歴には、正常出産、アルコール及び喫煙の乱用があった。また、以前に鼓膜切開術を受けていた。投薬の中には、種類が不明の避妊ピルが含まれていた。家族歴には、母の高脂血症及び抑うつ障害、祖父（母）の良性高血圧症、うつ病、心不全、及び慢性白血病があった。

10

20

30

40

表6-4

ライブラリの説明		
ライブラリ	ベクター	このライブラリは、成人（40～50歳）の白人男性から採取したアネルギー性同種 T-リンパ球組織から単離した RNA を用いて作製した。細胞は、1 μg/ml の OKT3 mAb 及び 5% ヒト血清の存在下で 3 日間インキュベートしたものである。
TYMN0708	pINCY	
UTRSTUC01	PSPORT1	この大きいサイズに断片化されたライブラリは、2 人のドナーからのプールされた cDNA を用いて作製した。cDNA は、37 歳の黒人女性（ドナー A）の筋腫摘出術、子宮内膜癌術、右卵管部生検、及び随伴性虫垂切除術の際に採取した子宮腫瘍組織、並びに 49 歳白人女性（ドナー B）の腔式子宮摘出術及び両卵管卵巣摘出術の際に採取した子宮平滑筋腫であることがわかった。卵管採取細胞が見られた。子宮内膜は、ホルモンが作用している分泌期であった。ドナー A は、欠乏性貧血、膈ヘルニア、及び閉経前月経過多の状態であった。病歴には閉経前月経過多及び肺のサルコイドーシスがあった。また、以前にヒステロスコーピー、子宮内膜癌術、及び内視鏡肺生検を受けていた。ドナー B は病理検査により、子宮底内、並びに前子宮壁及び近接する子宮内膜ポリープに及ぶ腫瘍を形成しているグレード 3 の腺癌平上皮癌であることが分かった。腫瘍は最大 7mm（侵入していない壁部の厚みは 2.2cm）侵入していた。近接する子宮内膜は不活性であった。エストロゲン受容体及びプロゲステロン受容体のパラフィン切片免疫染色は陽性であった。患者の病歴には、悪性乳房新生物があった。また、以前に片側単純拡大乳房切除術及び両卵管破壊術を受けていた。投薬には、Megace 及び CAF（シクロホスファミド、アドリアマイシン、Fluorouracil）が含まれていた。

【表 3 4】

10

20

30

40

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、アミノ酸または核酸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 BLAST	核酸配列を構築するプログラム。	Applied Biosystems, Foster City, CA.	
FASTA	Basic Local Alignment Search Tool は、アミノ酸及び核酸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。 Pearson 及び Lipman アルゴリズムは、両方合わせた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。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. 他 (1997) Nucleic Acids Res. 25: 3389-3402. Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 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: 確率値=1.0E-8 以下 完全長配列: 確率値=1.0E-10 以下
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、構造的フィンガンブリーント領域を探索する。	Henikoff, S. and J.G. Henikoff, Nucl. Acid Res., 19:5565-72, 1991. J.C.; Henikoff and S. Henikoff (1996) Methods Enzymol. 265:88-105; Artwood, T.K. 他 (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	ESTs: fasta E 値=1.0E-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 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 シングルペルプザドヒット: スコア=0 以上

10

20

30

40

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。	Gribkov, M. 他 (1988) CABIOS 4:61-66; Gribkov, 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と特定の Prosite モチーフに対する GCG 指定 "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短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リメンテーションに基づく SWAT や CrossMatch を含む Phils 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ャ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	重み付けマトリックスを用いたタンパク質配列上の膜貫通セグメントの明確化および向きを決定するためのプログラム	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	隠れマルコフモデル (HMM) を用いたタンパク質配列上の膜貫通セグメントの明確化および向きを決定するためのプログラム。	Sonnhammer, E.L. 他 (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow 他 eds., 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 Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

10

20

30

40

【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
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/98468 A2

(51) International Patent Classification: C12N 9/00

(21) International Application Number: PCT/US01/19178

(22) International Filing Date: 13 June 2001 (13.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/212,336	16 June 2000 (16.06.2000)	US
60/213,955	22 June 2000 (22.06.2000)	US
60/215,396	29 June 2000 (29.06.2000)	US
60/216,821	7 July 2000 (07.07.2000)	US
60/218,946	14 July 2000 (14.07.2000)	US

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

(72) Inventors: and

(75) Inventors/Applicants (for US only): YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). GANDHI, Ameena, R. [US/US]; 837 Robie Avenue, #1, Menlo Park, CA 94025 (US). LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). DELEGEANE, Angelo, M. [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). NGUYEN, Dannie, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). LEE, Ernestine, A. [US/US]; 624 Kains Street, Albany, CA 94706 (US). HAFALIA, April [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). KHAN, Farrab, A. [INA/US]; 3617 Central Road #102, Glenview, IL 60025 (US). WALLA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). LU, Dyung, Aina, M. [US/US];

233 Coy Drive, San Jose, CA 95123 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). TANG, Ya, Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). WALSH, Roderick, T. [IE/US]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ (GB). AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Avenue #3, Sunnyvale, CA 94086 (US). DAS, Debopriya [IN/US]; 1179 Bonita Avenue, Apt. 3, Mountain View, CA 94040 (US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). KALLICK, Deborah, A. [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter 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, CO, 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, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, 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, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

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

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

(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.



WO 01/98468 A2

WO 01/98468

PCT/US01/19178

PROTEASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteases and to the use of these sequences in hydrolysis of peptide bonds and 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.

10 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 01/98468

PCT/US01/19178

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: trypases (after arginine or lysine), aspases (after aspartate), chymases (after phenylalanine or leucine), metases (methionine), and serases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) *Methods 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*). The prolyl oligopeptidase (S9) family includes enzymes from prokaryotes and eukaryotes with greatly differing specificities. Dipeptidyl peptidase IV (DPP-IV) is identical to CD26 and is implicated in the inactivation of peptide hormones, as well as in regulating T-cell growth (reviewed in Kahne, T. et al. (1999) *Int. J. Mol. Med.* 4:3-15; Mentlein, R. (1999) *Regul. Pept.* 85:9-24). Inhibition of DPP-IV has been suggested as a treatment for type 2 diabetes (Holst, J.J. and C.F. Deacon (1998) *Diabetes* 47:1663-1670), and lowered serum DPP-IV activity has been measured in anorexia and bulimia patients (van West, D. et al. (2000) *Eur. Arch. Psych. Clin. Neurosci.* 250:86-92).

WO 01/98468

PCT/US01/19178

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 (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).

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.

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

WO 01/98468

PCT/US01/19178

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).

The proteasome is an intracellular protease complex found in some bacteria and in all
 5 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 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
 10 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
 15 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, Unp, 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).
 20 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
 25 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

30 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

WO 01/98468

PCT/US01/19178

papain-like enzymes (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

5 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 Rawlings, N.D. and A.J. Barrett (1994) Methods 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

10 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

15 the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium (Keszler, 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

20 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).

25 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

30 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 thiol 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

WO 01/98468

PCT/US01/19178

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
 5 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) 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.
 10 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
 15 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-
 20 1b 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).

25 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
 polypeptide. 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
 30 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

WO 01/98468

PCT/US01/19178

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 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).

Metalloproteases

Metalloproteases require a metal ion for activity, usually manganese or zinc. Examples of manganese metalloenzymes include aminopeptidase P and human proline dipeptidase (PEPD). Aminopeptidase P can degrade bradykinin, a nonapeptide activated in a variety of inflammatory responses. Aminopeptidase P has been implicated in coronary ischemia/reperfusion injury. Administration of aminopeptidase P inhibitors has been shown to have a cardioprotective effect in rats (Ersahin, C. et al (1999) J. Cardiovasc. Pharmacol. 34:604-611).

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 aminopeptidases B and N, angiotensin-converting enzyme, neurolysin, the matrix metalloproteases and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which 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/neurolysin can hydrolyze 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 components of the extracellular matrix (ECM). They are Zn^{+2} endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain which

WO 01/98468

PCT/US01/19178

- 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 to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In the inactive form, the Zn^{+2} ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the Zn^{+2} -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 stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is very important in tissue homeostasis (reviewed in Yong, V.W. et al. (1998) Trends Neurosci. 21:75). Ehlers-Danlos syndrome type VII C is caused by mutations in the procollagen I N-proteinase gene (Colige, A. et al. (1999) Am. J. Hum. Genet. 65:308-317).
- 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 (Saarialho-Kore, 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 (Steen, B. et al. (1998) 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; Tarabotti, G. et al. (1995) J. NCI 87:293; Davies, B. et al. (1993) Cancer Res. 53:2087). MMPs may be active in Alzheimer's disease. 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 (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing 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 SVMPs. 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.

WO 01/98468

PCT/US01/19178

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 (Schlöndorff, J. and C.P. Blobel (1999) J. Cell. Sci. 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), activating the program for lateral inhibition in *Drosophila* neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff 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 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. (1997) J. Biol. Chem. 272:556-562). To date eleven members are recognized by the Human Genome Organization (HUGO; <http://www.gene.ucl.ac.uk/users/hester/adamts.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).

The discovery of new proteases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in hydrolysis of peptide bonds and 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

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,"

WO 01/98468

PCT/US01/19178

"PRTS-8," "PRTS-9," "PRTS-10," "PRTS-11," "PRTS-12," "PRTS-13," "PRTS-14," "PRTS-15," "PRTS-16," "PRTS-17," "PRTS-18," "PRTS-19," "PRTS-20," and "PRTS-21." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a

5 polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the invention provides an isolated polypeptide

10 comprising the amino acid sequence of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-

15 21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-21. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:22-42.

20 Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically

25 active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. 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.

30 The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

WO 01/98468

PCT/US01/19178

consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. 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 selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of 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 selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of 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 target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a

WO 01/98468

PCT/US01/19178

polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of 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.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. 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 of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. 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 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 selected from the group consisting of a) a polypeptide comprising an

WO 01/98468

PCT/US01/19178

amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and
 5 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method 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
 10 treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a
 15 naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under
 20 suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, b) a polypeptide comprising a
 25 naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21. The method comprises a) combining the polypeptide with at least one test compound under
 30 conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test 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.

WO 01/98468

PCT/US01/19178

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:22-42, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered
 5 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 selected from the group consisting of i) a polynucleotide
 10 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions
 15 whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42, iii) a
 20 polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of 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
 25 biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

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

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

WO 01/98468

PCT/US01/19178

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

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble
5 polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides 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.

10 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

Before the present proteins, nucleotide sequences, and methods are described, it is understood
15 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.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an,"
20 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.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings
25 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
30 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.

WO 01/98468

PCT/US01/19178

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 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.

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

WO 01/98468

PCT/US01/19178

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.

- 5 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
10 directly interacting with PRTS or by acting on components of the biological pathway in which PRTS 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
15 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.

- 20 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
25 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
30 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

WO 01/98468

PCT/US01/19178

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'.

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 Phrap (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
Asp	Asn, Glu

WO 01/98468

PCT/US01/19178

	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
5	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
10	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
15	Val	Ile, Leu, Thr

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.

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.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

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,

WO 01/98468

PCT/US01/19178

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
 5 defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:22-42 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:22-42, for example, as distinct from any other sequence in the
 10 genome from which the fragment was obtained. A fragment of SEQ ID NO:22-42 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:22-42 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:22-42 and the region of SEQ ID NO:22-42 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:1-21 is encoded by a fragment of SEQ ID NO:22-42. A fragment of SEQ ID NO:1-21 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-21. For example, a fragment of SEQ ID NO:1-21 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-21. The precise length of a fragment of SEQ ID NO:1-21 and the region of SEQ ID NO:1-21 to which the fragment
 20 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.

25 "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
 30 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

WO 01/98468

PCT/US01/19178

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 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:

20 *Matrix: BLOSUM62*
 Reward for match: 1
 Penalty for mismatch: -2
 Open Gap: 5 and Extension Gap: 2 penalties
 Gap x drop-off: 50
 25 *Expect: 10*
 Word Size: 11
 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.

WO 01/98468

PCT/US01/19178

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.

5 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
10 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: Ktuple=1, gap penalty=3, window=5, and
15 "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.

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
20 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

25 *Word Size: 3*

Filter: on

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,
30 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 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.

WO 01/98468

PCT/US01/19178

"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.

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 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 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.

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

WO 01/98468

PCT/US01/19178

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., C₀t or R₀t 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 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

WO 01/98468

PCT/US01/19178

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 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.

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 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).

WO 01/98468

PCT/US01/19178

Oligonucleotides for use as primers are selected using software known in the art for such 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

5 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 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,

10 Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for 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

15 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. 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

20 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 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.

25 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.

30 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.

WO 01/98468

PCT/US01/19178

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.

5 "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.

10 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, 15 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 20 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 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 25 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 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.

30 "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, 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.

WO 01/98468

PCT/US01/19178

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 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" 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 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, 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), *supra*.

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 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% 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

WO 01/98468

PCT/US01/19178

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 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 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

15

THE INVENTION

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 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 Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) 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 polypeptides 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

WO 01/98468

PCT/US01/19178

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 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.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteases. For example, SEQ ID NO:1 is a ubiquitin carboxyl terminal hydrolase. SEQ ID NO:1 is 48% identical, from residue M1 to residue G225, to human ubiquitin-specific processing protease (GenBank ID g9971757) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.00e-49, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 contains a ubiquitin carboxyl terminal hydrolase catalytic site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. The score is 53.4 bits and the E-value is 4.9e-12, which indicates the probability of obtaining the observed structural motif by chance. The presence of this motif was corroborated by BLIMPS (probability score=2.6e-4) and MOTIFS analyses. This provides further evidence that SEQ ID NO:1 is a ubiquitin carboxyl-terminal hydrolase. In an alternative example, SEQ ID NO:2 is 45% identical to amino acids 15-235 of human prostatic, a serine protease (GenBank ID g1143194) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.3e-46, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a trypsin family serine protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. This match has a probability score of 2.7e-58. BLIMPS, MOTIFS, and PROFILESCAN analyses confirm the presence of this domain. (See Table 3.) BLIMPS analysis also reveals a kringle domain, providing further corroborative evidence that SEQ ID NO:2 is a serine protease of the trypsin family. In an alternative example, SEQ ID NO:7 is a dipeptidase which hydrolyses a variety of peptides (Kozak, E. and S. Tate (1982) J. Biol. Chem. 257:6322-6327), and is responsible for the hydrolysis of the beta

WO 01/98468

PCT/US01/19178

lactam rings of antibiotics such as penem and carbapenem (Campbell et al., (1984) J. Biol. Chem. 259:14586-14590). SEQ ID NO:7 shows 48% amino acid sequence identity over 377 amino acids (total length equals 411 amino acids) to human dipeptidase precursor (GenBank ID g219600) as determined by Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 1.1e-88, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Additionally, the protease of the invention demonstrates a renal dipeptidase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. The HMM score for the renal dipeptidase PFAM hit is 412.7. Data from BLIMPS, MOTIFS, BLAST-DOMO, and BLAST-PRODOM analyses provide further corroborative evidence that SEQ ID NO:7 is a renal dipeptidase. The BLIMPS-BLOCKS hit scores for localized regions range from 1040-1537. The BLAST-DOMO hit probability score is 5.2e-85. The BLAST-PRODOM hit probability score is 4.7e-73. In an alternative example, SEQ ID NO:8 is 86% identical to human transmembrane trypsin (GenBank ID g6103629) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.9e-166, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 contains a trypsin family protease active site domain with a probability score of 5.3e-89 as determined by searching for matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. BLIMPS, MOTIFS, and PROFILESCAN analyses confirm the presence of this motif. BLIMPS analysis also shows that SEQ ID NO:8 contains a kringle domain and a type I fibronectin domain. HMMER-based analysis reveals the presence of a transmembrane domain (See Table 3.). Taken together, these analyses show that SEQ ID NO:8 is a transmembrane member of the trypsin family of serine proteases. In an alternative example, SEQ ID NO:17 shares 44% local identity with human membrane-type serine protease 1 (MT-SP1, GenBank ID g6002714) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.1e-94, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 contains a trypsin family serine protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) HMM-based analysis also reveals a transmembrane domain near the N-terminus of SEQ ID NO:17. A domain found in the low-density lipoprotein receptor and other proteins, including MT-SP1 (PDOC00929) was also identified in this way. The presence of the trypsin active site motif is confirmed by PROFILESCAN, BLIMPS, and MOTIFS analyses. BLIMPS analysis revealed the presence of kringle and type I fibronectin domains. Taken together, these data provide further corroborative evidence that SEQ ID NO:17 is a transmembrane member of the trypsin family of serine proteases. SEQ ID NO:3-6, SEQ ID NO:9-16,

WO 01/98468

PCT/US01/19178

and SEQ ID NO:18-21 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-21 are described in Table 7.

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:22-42 or that distinguish between SEQ ID NO:22-42 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/or 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, 7246467T8 is the identification number of an Incyte cDNA sequence, and PROSTMY01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71041539V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5745066) 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.g7208751_000002_002.edit is the identification number of a Genscan-predicted coding sequence, with g7208751 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. For example, FL1389845_00001 represents a "stitched" sequence in which 1389845 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the 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. For example, FL2256251_g7708357_000002_g6103629 is the identification

WO 01/98468

PCT/US01/19178

number of a "stretched" sequence, with 2256251 being the Incyte project identification number, g7708357 being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, and g6103629 being the GenBank identification number of the nearest GenBank protein homolog. (See Example V.) In some cases, Incyte cDNA coverage
 5 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
 10 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
 15 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:22-42, which encodes PRTS. The polynucleotide sequences of
 20 SEQ ID NO:22-42, 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
 25 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:22-42 which has at 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:22-42. Any one of the polynucleotide variants described above can encode an amino acid sequence which
 30 contains at least one functional or structural characteristic of PRTS.

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

WO 01/98468

PCT/US01/19178

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.

Although nucleotide sequences which encode PRTS and its variants are generally capable of
5 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
10 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.

The invention also encompasses production of DNA sequences which encode PRTS and PRTS
15 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.

Also encompassed by the invention are polynucleotide sequences that are capable of
20 hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:22-42 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 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
30 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,

WO 01/98468

PCT/US01/19178

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.)

- 5 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., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)
- 10 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. 15 (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 20 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 25 72°C.

- When screening for full length cDNAs, it is preferable to use libraries that have been 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' 30 non-transcribed regulatory regions.

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

WO 01/98468

PCT/US01/19178

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 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

5 which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which 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

10 amino acid sequence may be produced and used to express PRTS.

The nucleotide sequences of the present invention can be engineered using methods generally 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

15 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, 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 MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number

20 5,837,458; 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 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

25 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 DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" 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

30 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 maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

WO 01/98468

PCT/US01/19178

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, PRTS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide
 5 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 sequence of PRTS, or any part thereof, may be altered during direct synthesis and/or combined with sequences
 10 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. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing.
 15 (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 suitable host. These elements include regulatory sequences, such as enhancers, constitutive and
 20 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 sequences encoding PRTS and its initiation codon and upstream regulatory sequences are inserted into
 25 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 and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate
 30 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*

WO 01/98468

PCT/US01/19178

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 Heeke, 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; Buller, 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 Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 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.

WO 01/98468

PCT/US01/19178

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) *BioTechnology* 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 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

WO 01/98468

PCT/US01/19178

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 *ap^r* 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 chlorsulfuron 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 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

WO 01/98468

PCT/US01/19178

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) Immunochemical 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 radionuclides, 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 WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing 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

WO 01/98468

PCT/US01/19178

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. coli*. Cells expressing PRTS or cell membrane fractions which contain PRTS are then contacted with

WO 01/98468

PCT/US01/19178

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 01/98468

PCT/US01/19178

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

PRTS are useful for hydrolyzing peptide bonds. 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 hemic, neurological, reproductive, endocrine, urogenital, diseased, teratocarcinoma, and tumorous 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-

WO 01/98468

PCT/US01/19178

- 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
- 5 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, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins,
- 10 thrombophlebitis and phlebothrombosis, 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,
- 15 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 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,
- 20 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 lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable
- 25 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 syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral,
- 30 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 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal

WO 01/98468

PCT/US01/19178

- 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, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,
- 5 bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, 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
- 10 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,
- 15 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
- 20 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,
- 25 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, 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,
- 30 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, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including

WO 01/98468

PCT/US01/19178

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 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 polyneuropathy, 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

WO 01/98468

PCT/US01/19178

directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PRTS.

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
5 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
10 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
15 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,
20 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,
25 polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (*Bacilli Calmette-Guerin*) and *Cornebacterium 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
30 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

WO 01/98468

PCT/US01/19178

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; Kozbor, D. et al. (1985) *J. 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.)

5 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
10 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 idiotype 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
15 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
20 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
25 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
30 (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

WO 01/98468

PCT/US01/19178

- 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 a preparation of monoclonal antibodies, which are monospecific for a particular PRTS epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about
- 5 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_a ranging from about 10^6 to 10^7 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
- 10 (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
- 15 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 Cofigan 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
- 20 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.)
- 25 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
- 30 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

WO 01/98468

PCT/US01/19178

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 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 vivo 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, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PRTS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus

WO 01/98468

PCT/US01/19178

(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) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the
 5 ecdysone-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
 10 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
 15 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
 20 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) 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
 25 receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, 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
 30 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

WO 01/98468

PCT/US01/19178

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 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 Antinuzzi, P.A. et al. (1999) *Annu. 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

WO 01/98468

PCT/US01/19178

the SFV genome (Garoff, 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.,
 5 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
 10 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
 15 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,
 20 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.E. and B.I. 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.

25 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.

30 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme 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

WO 01/98468

PCT/US01/19178

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
 5 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 complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

10 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 extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and
 15 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 endonucleases.

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 which may be effective in altering expression of a specific polynucleotide may include, but are not
 20 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 polynucleotide expression. Thus, in the treatment of disorders associated with increased PRTS
 25 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
 30 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 based on chemical and/or structural properties of the target polynucleotide; and selection from a

WO 01/98468

PCT/US01/19178

- 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
- 5 any method commonly known in the art. Typically, the expression of a specific nucleotide is detected 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
- 10 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.
- 15 268:8-13). 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
- 20 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.)
- 25 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.
- 30 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.

WO 01/98468

PCT/US01/19178

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.

5 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 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
10 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 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
15 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 macromolecules comprising PRTS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the
20 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 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,
25 or pigs. An animal model may also be used to determine the appropriate concentration range and route 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
30 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 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

WO 01/98468

PCT/US01/19178

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 preferably within a range of circulating concentrations that includes the ED_{50} 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 moiety 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. 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 μ g to 100,000 μ g, 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 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 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 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 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,

WO 01/98468

PCT/US01/19178

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 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.

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 5' 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:22-42 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 ^{32}P or ^{35}S , 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,

WO 01/98468

PCT/US01/19178

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

5 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, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose

10 veins, thrombophlebitis and phlebothrombosis, 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,

15 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 autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia,

20 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 lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's

25 thyroiditis, hypereosinophilia, 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 syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and

30 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 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

WO 01/98468

PCT/US01/19178

- 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, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and
- 5 Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, 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,
- 10 anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dysidrotic 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,
- 15 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,
- 20 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
- 25 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, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
- 30 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, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion

WO 01/98468

PCT/US01/19178

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 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, 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 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. 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.

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 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.

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

WO 01/98468

PCT/US01/19178

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 values from an experiment in which a known amount of a substantially purified polynucleotide is used.

- 5 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, 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.

- 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 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 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 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 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

WO 01/98468

PCT/US01/19178

using gel electrophoresis in non-denaturing gels. In fSCCP, 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 (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual
 5 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 throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

10 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
 15 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
 20 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
 25 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
 30 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

WO 01/98468

PCT/US01/19178

given time. (See Seifhamer 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, 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, E.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

WO 01/98468

PCT/US01/19178

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 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; Mendoz, 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 proteome 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. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the

WO 01/98468

PCT/US01/19178

analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. 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 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

WO 01/98468

PCT/US01/19178

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 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

WO 01/98468

PCT/US01/19178

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, including U.S. Ser. No. 60/212,336, U.S. Ser. No. 60/213,995, U.S. Ser. No. 60/215,396, U.S. Ser. No. 60/216,821, and U.S. Ser. No. 60/218,946, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incute cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incute Genomics, Palo Alto CA) and shown in Table 4, column 5. 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 (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 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 libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript 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

WO 01/98468

PCT/US01/19178

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the 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., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-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.

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 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 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 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 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

WO 01/98468

PCT/US01/19178

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 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, BLIMPS, 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 Consed, 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 multisequence 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

WO 01/98468

PCT/US01/19178

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

5 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:22-42. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

10 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., gbpr 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
15 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 models for proteases. Potential proteases were also identified by homology to Incyte cDNA sequences that had
20 been annotated as proteases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr 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
25 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
30 sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"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

WO 01/98468

PCT/US01/19178

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 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.

- 5 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
- 10 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 genpept and gbpr public
- 15 databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. 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
- 20 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
- 25 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
- 30 determine whether it contained a complete gene.

VI. Chromosomal Mapping of PRTS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:22-42 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

WO 01/98468

PCT/US01/19178

SEQ ID NO:22-42 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 Genome Research (WIGR), and Généthon 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, of 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 Généthon 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.

In this manner, SEQ ID NO:25 was mapped to chromosome 5 within the interval from 69.60 to 76.50 centiMorgans. SEQ ID NO:28 was mapped to chromosome 16 within the interval from 81.80 to 84.40 centiMorgans.

20 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.)

25 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESQ (Incyte 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:

$$30 \quad \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length(Seq. 1)}, \text{length(Seq. 2)} \}}$$

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

WO 01/98468

PCT/US01/19178

as follows: the BLAST score is multiplied by the percent nucleotide identity and the 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 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.

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; 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, 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).

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 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.

WO 01/98468

PCT/US01/19178

Selected human cDNA libraries were used to extend the sequence. If more than one extension 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 Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (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 T7 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 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l 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 μ l to 10 μ l 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

WO 01/98468

PCT/US01/19178

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).

- 5 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 designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:22-42 are employed to screen cDNAs, genomic

- 10 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 Ci of [γ - ^{32} P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston
- 15 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).
- 20 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
- 25 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., Baldeschweiler, *supra.*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned
- 30 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

WO 01/98468

PCT/US01/19178

contain any appropriate number of elements. (See, e.g., Schena, 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
- 5 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 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
- 10 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.

15 **Tissue or Cell Sample Preparation**

- 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
- 20 dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse 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
- 25 incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified 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
- 30 resuspended in 14 μ l 5X SSC/0.2% SDS.

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

WO 01/98468

PCT/US01/19178

amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. 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 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and 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.

WO 01/98468

PCT/US01/19178

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.

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-835H 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).

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.

WO 01/98468

PCT/US01/19178

XII. Expression of PRTS

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.

- 5 Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic 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
- 10 recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PRTS 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
- 15 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 91:3224-3227; Sandig, 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,
- 20 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 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
- 25 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 (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 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

WO 01/98468

PCT/US01/19178

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 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 optics-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.)

WO 01/98468

PCT/US01/19178

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

5 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.

XV. Purification of Naturally Occurring PRTS Using Specific Antibodies

Naturally occurring or recombinant PRTS is substantially purified by immunoaffinity

10 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

15 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

25 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)

30 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

WO 01/98468

PCT/US01/19178

- 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. For example, arginine- β -naphthylamide can be used as a substrate for SEQ ID NO:3 (Fukasawa, K.M. et al. (1996) J. Biol. Chem. 271:30731-30735) and 4-phenylazobenzoyloxycarbonyl-Pro-Leu-Gly-Pro-D-Arg can be used as a substrate for SEQ ID NO:4. In an alternative example, a substrate for SEQ ID NO:9 would be 7-amino-4-trifluoromethyl coumarin-Phe-Pro-AFC. Assays are 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.
- 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. For SEQ ID NO:1, 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).
- Alternatively, the ubiquitin protease activity of SEQ ID NO:5 can be measured using the method of Sloper-Mould et al. ((1999) J. Biol. Chem. 274:26878-26884). Aliquots of PRTS are incubated with 5 μ l [^{35} S]-labeled ubiquitin-GST fusion substrate for 1 hour at 37 °C in an appropriate buffer. Samples are resolved by electrophoresis on a 12% SDS-PAGE gel. Ubiquitin cleavage is monitored by fluorography of the gel.
- Alternatively, the activity of SEQ ID NO:10, for example, can be measured by the method of Colige et al. (1999, J. Biol. Chem. 270:16724-16730). An aliquot of PRTS is incubated with amino procollagen type I substrate radioactively labeled only in the propeptide in a 250 μ l reaction containing 50 mM sodium cacodylate, pH 7.5, 200 mM KCl, 2 mM CaCl₂, 2.5 mM NEM, 0.5 mM PMSF, and 0.02% Brij (standard assay buffer). After 16 h at 26 °C, the reaction is stopped by adding 50 μ l of EDTA solution (0.2 M EDTA, pH 8, 0.5% SDS, 0.5 M DTT) and 300 μ l of 99% ethanol. The samples are kept for 30 min at 4 °C and centrifuged for 30 min at 9500 g. Collagen and uncleaved

WO 01/98468

PCT/US01/19178

radioactive pN-collagen substrate are pelleted, whereas the freed amino propeptides remained in solution. An aliquot of the supernatant is assayed by liquid scintillation spectrometry.

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 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 that FRET can be monitored in the presence and absence of PRTS (Sagot, I. et al. (1999) *FEBS Lett.* 447:53-57).

In yet another alternative, an assay for PRTS dipeptidase activity measures the hydrolysis activity of PRTS on a variety of dipeptides such as leukotriene D4 (Kozak, E. and S. Tate (1982) *J. Biol. Chem.* 257:6322-6327), or hydrolysis of the beta-lactam ring of antibiotics such as penum and carbapenem (Campbell et al., (1984) *J. Biol. Chem.* 259:14586-14590).

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 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 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 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.

WO 01/98468

PCT/US01/19178

XIX. Identification of PRTS Inhibitors

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. PRTS activity is measured for each well and the ability of each compound to inhibit PRTS activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PRTS activity.

In the alternative, phage display libraries can be used to screen for peptide PRTS inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PRTS 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 PRTS 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 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 01/98468

PCT/US01/19178

Table 1

Incyte Project ID	Polypeptide Seq ID NO:	Incyte Polypeptide ID	Polynucleotide Seq ID NO:	Incyte Polynucleotide ID
275791	1	275791CD1	22	275791CB1
1389845	2	1389845CD1	23	1389845CB1
1726609	3	1726609CD1	24	1726609CB1
4503848	4	4503848CD1	25	4503848CB1
5544089	5	5544089CD1	26	5544089CB1
7474081	6	7474081CD1	27	7474081CB1
5281209	7	5281209CD1	28	5281209CB1
2256251	8	2256251CD1	29	2256251CB1
7160544	9	7160544CD1	30	7160544CB1
7477386	10	7477386CD1	31	7477386CB1
7473089	11	7473089CD1	32	7473089CB1
7604035	12	7604035CD1	33	7604035CB1
3473847	13	3473847CD1	34	3473847CB1
3750004	14	3750004CD1	35	3750004CB1
4904126	15	4904126CD1	36	4904126CB1
71268415	16	71268415CD1	37	71268415CB1
7473301	17	7473301CD1	38	7473301CB1
7473308	18	7473308CD1	39	7473308CB1
7478021	19	7478021CD1	40	7478021CB1
4333459	20	4333459CD1	41	4333459CB1
6817347	21	6817347CD1	42	6817347CB1

WO 01/98468

PCT/US01/19178

Table 2

Polypeptide SEQ ID NO:	Incyte ID	GenBank ID NO:	Probability score	GenBank homolog
1	275791CD1	g9971757	1.00E-49	ubiquitin-specific processing protease [Homo sapiens]
2	1389645CD1	g1143194	1.30E-46	prolactin [Homo sapiens]
3	1726609CD1	g10719660	0	ENPEP-like protein [Homo sapiens]
		g1754515	3.30E-96	(Horikawa, Y. et al. (2000) Nat. Genet. 26:163-175.)
				aminopeptidase-B [Rattus norvegicus]
				(Prieto, I. et al. (1998) Horm. Metab. Res. 30:246-248.)
4	4503848CD1	g1783122	0	endopeptidase 24.16 type M1 [Sus scrofa]
5	5540489CD1	g5410230	5.20E-43	ubiquitin-specific protease 3 [Homo sapiens]
6	7474081CD1	g603903	2.90E-33	Trypsinogen [Gallus gallus]
7	5281209CD1	g11071729	0	Putative dipeptidase [Homo sapiens]
		g219600	1.40E-88	dipeptidase precursor [Homo sapiens]
				(Sato, S. et al. (1993) Biochim. Biophys. Acta 1172:181-183.)
8	2256251CD1	g6103629	3.90E-166	transmembrane tryptase [Homo sapiens]
				(Wong, G.W. et al. (1999) J. Biol. Chem. 274:30794-30798.)
9	7160544CD1	g11095188	0	dipeptidyl peptidase 8 [Homo sapiens]
				(Abbott, C.A. et al. (2000) Eur. J. Biochem. 267:6140-6150.)
		g1753197	6.80E-64	dipeptidyl peptidase IV [Stenotrophomonas maltophilia]
				(Mentem, R. (1995) Regul. Pept. 55:9-24; Kahne, P. Zuck. Biol. Med. (1999) 43-55.)
10	7477386CD1	g1865716	0	Procollagen I N-protease [Bos taurus]
				(Collige, A. et al. (1999) Am. J. Hum. Genet. 65:308-317.)
11	7473089CD1	g7768706	3.60E-255	metalloprotease with thrombospondin type 1 motifs [Homo sapiens]
				(Vazquez, F. et al. (1999) J. Biol. Chem. 274:23349-23357.)
12	7604035CD1	g6164595	4.70E-68	Lacunin [Manduca sexta]

WO 01/98468

PCT/US01/19178

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
13	3473847CD1	g217172	9.20E-50	aqualysin precursor (aa 1 to 513) [Thermus aquaticus]
14	3750004CD1	g5923786	4.30E-51	zinc metalloprotease ADAMTS6 [Homo sapiens]
15	4904126CD1	g186286	3.90E-40	interleukin 1-beta convertase [Homo sapiens] (Cerretti, D.P. et al. (1992) Science 256:97-100.)
16	74268415CD1	g6651071	0	disintegrin and metalloproteinase domain 19 [Homo sapiens] (Inoue, D. et al. (1998) J. Biol. Chem. 273:4180-4187.)
17	7473301CD1	g6002714	5.10E-94	membrane-type serine protease 1 [Homo sapiens] (Takeuchi, T. et al. (1995) Proc. Natl. Acad. Sci. USA 96:11054-11061.)
18	7473308CD1	g1552517	6.60E-77	trypsinogen E [Homo sapiens]
19	7478021CD1	g3211705	5.60E-189	matrix metalloproteinase [Xenopus laevis] (Yang, M. (1997) J. Biol. Chem. 272:13527-13533.)
20	4333459CD1	g1754714	2.30E-67	Oviductin [Xenopus laevis] (Lindsay, L.L. et al. (1999) Biol. Reprod. 60:989-995.)
21	6817347CD1	g7673618	5.10E-283	ubiquitin specific protease [Mus musculus]

Table 3

SEQ ID NO.	Tripeptide Index	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	275791CD1	232	T15 T17 S23 S43 S71 T30 S93 S100 S107 S111 T122 T174 S190 T210 S241 S190 T213 T227		Ubiquitin carboxyl-terminal hydrolase family 2 signature 2 UCH-2-2: Y142-Y160 Ubiquitin carboxyl-terminal hydrolase family 2 signature 2 UCH-2: L138-H203 Ubiquitin carboxyl-terminal hydrolase family 2 signature 2 BU00972: Y142-D166, Y169-S190 TRYPSIN DM000181 A570-145-284; T123-Q314 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTISERINE FACTOR, PD000046: I123-Q314 Serine proteases, trypsin family BL00134A: C148-C164 Kringlike domain proteins BL00021: C148-F165, V231-G252 CHYMOTRYPSIN SERINE PROTEASE PR00722: C149-C164, G208-V222 V8 Serine proteases PR00839B: C148-F165 Serine proteases, trypsin family, active sites trypsin_his.prpf: I140-F168 Trypsin: I123-Q314 Trypsin His: L159-CL64	HMW-PPAM BLIMPS-BLOCKS BLIMPS-BLOCKS BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS HMMER-PPAM MOTIFS
2	1389845CD1	365	S120 S187 S225 S253 S82 T31 T37 T42 Y283		BLAST_DOMO BLAST_PRODOM	

Table 3 (cont.)

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Potential phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1726609CD1	416	S244 S283 S30 S370 S408 T389 T59 T78 T87	R203 N413 N57	DO HYDROLASE; LEUKOTRIENE; A-4; ZINC; DM87071P1960217-609; M1-P354 AMINOPEPTIDASE HYDROLASE METALLOPROTEASE ZINC N GLYCOPROTEIN PROTEIN TRANSMEMBRANE SIGNAL ANCHOR MEMBRANE PD001134; R4-S177 Neutral zinc metalloprotease family BL00142; D41-F51 MEMBRANE ANCHOR DIPEPTIDASE PR00756; F14-L24, D41-T56, W60-Y72 signal cleavage: M1-S34	BLAST_DOMO ELAST_PRODUM BLIMPS_BLOCKS BLIMPS_PRINTS
4	4503848CD1	714	S124 S140 S147 S179 S200 S206 S226 S233 S251 S268 S321 T114 S555 S592 T114 T133 T214 T252 T270 T306 T318 T322 T376 T406 T432 T526 T585 T502 T69 T175 T249 T505	N485 M601	DO ZINC; METALLOPEPTIDASE; NEUTRAL; OLIGO-PEPTIDASE DM11184 [Q02038]36-702; A46-N4713 HYDROLASE METALLOPROTEASE ZINC OLIGOPEPTIDASE PRECURSOR MITOCHONDRIAL ENDOPPEPTIDASE MITOCHONORION TRANSIT PEPTIDE PR007945; W60-N529 Transmembrane domain: L14-W34 Peptidase family M3 Peptidase_X3: C98-7/11 Neutral zinc metalloprotease family BL00142; T504-H514 Zinc Protease: T504-N513 signal cleavage: M1-G27	SSScan ELAST_DOMO ELAST_PRODUM BLAST_PRODUM HMHR HAHR_PPAM BLIMPS_BLOCKS MOTTIS

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	5544089CD1	367	S108 S161 S197 S203 S235 S266 S361 S49 T180 T263 T316 T331	N139 N142 N308	UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DMO0659 P40818 782-1103: L36-L328 Ubiquitin carboxyl-terminal hydrolase family BL00972: Y74-L83, V120-CL34, Y274-N298, G301-K322 Ubiquitin carboxyl-terminal hydrolase family UCH-2: E270-Q332 Ubiquitin carboxyl-terminal hydrolase family Uch 2.2: Y274-Y292 signal cleavage: M1-S16	BLAST_DOMO BLIMPS_BLOCKS HMMER_Pfam MOTIFS SPScan
6	7474081CD1	235	S134 S143 S159 S171 S226 T231	N90	TRYPSIN DMO0018 S55065 26-244: A27-T231 Serine proteases, trypsin family signature BL00134: C40-C56, V115-I228 Type 1 fibronectin domain BL01253: C40-P53, I197-T231 Kringles domain proteins BL00021: S159-Q164, C40-Y57 CHYMOTRYPSIN SERINE PROTEASE PRO0722A: V41-C56, A95-A109 Serine proteases, trypsin family, active site trypsin.His.pri: S55-T76 Trypsin: G42-V176, G216-I228 Leucine zipper: L44-L65 signal cleavage: M1-S19	BLAST_DOMO BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS HMMER_Pfam MOTIFS SPScan

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	5281209CD1	488	S13 T74 S186 S233 T363 T456 T4 T34 Y125 S170 S172 Y178 S249 T337 S387 S389 S419 T447	M119 N184 N243 N334	Renal dipeptidase proteins BL00869:P92-L247, E280-R412, S415-N457 DIPEPTIDASE MICROSOMAL PRECURSOR MOP HYDROLASE MICROSOMAL SIGNAL GPI-ANCHOR GLYCOPROTEIN ZINC DQ02526:S143-E450 RENAL DIPEPTIDASE DQ02775:T777-R410 Renal dipeptidase:V195-R217 Renal dipeptidase:A63-V475 Signal peptide:MI-A36 HMMER-PPFAM HMMER MOTIFS	BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO MOTIFS HMMER-PPFAM HMMER SPSCAN
8	22562615CD1	346	S203 S210 S266 S45 S79 T131 T147 T216	M110	TRYPSIN DQ00018:P15944:131-270, I63-I294 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: L156-I290, I63-F178, N288-F314, P274-P305 Serine proteases, trypsin family BL00134: C88-C104, D241-I264, E277-I290 Type I fibronectin domain BC01253: C88-A101, V158-E194, G240-C253, W259-H293 Xingde domain protease BL00021: C88-F105, V169-GL90, G249-I290 CHROMATOPHSIN SERINE PROTEASE PR00722: G49-C104, G146-V160, G240-V252 transmembrane domain: P308-L328 Trypsin: I63-I290 Trypsin_His I99-C104 Trypsin family serine protease active sites trypsin_his.prf: L80-H128 trypsin_ser.prf: I229-I273 signal cleavage: M1-S45	BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_PRINTS HMMER HMMER-PPFAM MOTIFS PROFILESSCAN SPSCAN

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	7160544CD1	882	S115 S133 S293 S312 S412 S443 S479 S530 S587 S588 S723 S80 S850 T227 T234 T307 T326 T499 T52 T551 T594 T603 T615 T776 Y315 Y336 Y55 Y555 Y844		PROLYL ENDOPEPTIDASE FAMILY SERINE DMQ2461 P27487192-765: F488-E870, G251-E370 DIPEPTIDYL IV HYDROLASE PROTEIN SERINE PEPTIDASE DIPEPTIDASE TRANSMEMBRANE GLYCOPROTEIN PROTEIN PD003048: L744-E870 DIPEPTIDYL IV HYDROLASE PROTEIN SERINE PEPTIDASE DIPEPTIDASE TRANSMEMBRANE GLYCOPROTEIN PROTEIN PD003086: Y423- V661, I212-T326 Prolyl endopeptidase family BL00708: G501-I513, Q714-I744 Dipeptidyl peptidase IV PF00930: Y498- R508, F756-F783, R808-L828 PROLYL OLIGOPEPTIDASE SERINE PROTEASE PRO0862: P647-F665, G737-R757 Prolyl oligopeptidase family Peptidase_S9: S672-L744 Dipeptidyl peptidase IV DPP4V_N-term: M85-R663	BLAST_DOMO BLAST_PRODUM BLAST_PRODUM BLIMPS_BLOCKS BLIMPS_PPFAM BLIMPS_PRINTS HMMER_PPFAM HMMER_PPFAM

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7477386CD1	1189	S132 S169 S200 S32 S323 S350 S445 S480 S511 S626 S675 S699 S798 S1064 T247 T362 T521 T612 T718 T777 T946 T986 T1104 Y552	N109 N478 N944	do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN DM00368 Q05910 189-395:I261- P463 THROMBOSPONDIN TYPE 1 REPEAT DM00275 P07996 477-540: P555-C604 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE C0284.1 A DISINTEGRIN METALLOPROTEASE PD013511: L474-E549 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q647-C716 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR PD007018: W854-Q974, W914- C1029, W558-K623 METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791:P256- P463 Neutral zinc metalloproteinase BL00142: T398-N408 signal peptide: M1-A22 Reprolysin family propeptide Pep_M128 DROMER: R420-V240 Reprolysin (M128) family zinc metalloprotease Reprolysin type 1 Thrombospondin type 1 domain 1 A973-C1024, S559-C609, Y852-C909, W914- C971 signal cleavage: M1-G23	ELAST_DOMO ELAST_DOMO ELAST_PRODUM ELAST_PRODUM ELAST_PRODUM ELAST_PRODUM ELIMPS_BLOCKS HMMER HMMER_FFAM HMMER_FFAM HMMER_FFAM SPSCAN

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	7473089CD1	952	S19 S203 S207 S303 S346 S432 S492 S575 S578 S611 S666 S682 S708 S745 S819 T171 T288 T317 T325 T337 T359 T471 T594 T687 T765	N141 N591 N623 N680	do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN DM00366 JC25501-201:R218-P427 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: K664-E804 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: V610-C683 METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD00791: V214-P427 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE C02B4.1 A DISINTEGRIN METALLOPROTEASE PD013511: L437-V505 Neutral zinc metalloproteinase BL00142: T358-N368 signal peptide: M1-G18 Reprolysin family propeptide Rep. M12B propep: H67-N181 Reprolysin (M12B) family zinc metalloproteinase Reprolysin: E218-P427 Thrombospondin type 1 domain tsp_1: A520-C570. H645-C896. H899-C952 Zinc Protease: T358-F367 Specan signal_cleavage: M1-G17	BLAST_POMO BLAST_PRODOR BLAST_PRODOR BLAST_PRODOR BLAST_PRODOR BLAST_PRODOR BLAST_PRODOR BLIMPS_BLOCKS HMMER HMMER_FFAM HMMER_FFAM HMMER_FFAM MOTIFS SPECAN

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO.	Incyte ID	Polypeptide	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7604035CD1		898	S187 S188 S258 S268 S285 S415 S467 S547 S696 S796 S819 S851 S892 T106 T198 T35 T434 T483 T492 T5	N3 N490 N773	PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE PD007018: Y726-C841, W786-A874, Y667-C778, W50-Q72, S368-Q383 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q416-C484 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEPROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: R485-I599 signal peptide: M1-D24 Thrombospondin type 1 domain tsp_1: G48-R87, W227-C783, E787-C841 signal cleavage: M1-D24	BLAST_PRODOM
13	3473847CD1		631	S117 S160 S174 S185 S188 S268 S28 S30 S358 S431 S503 S605 T142 T33 T346 T512 T606	N472	Subtilase family Peptidase S8, S86-N364 SERINE PROTEASES, SUBTILASE FAMILY HISTIDINE DM00108/P80146 150-377: G116-T346 Serine proteases, subtilase family BL00136: L123-I135, D163-Q175, G323-G333 SUBTILASIN SERINE PROTEASE FAMILY PRO0723: G16-I195, K161-S174, S122-M338 Serine proteases, subtilase family, active site subtilase_ser.pef: A302-Q352	HMMER HMMER_PFM SPSCAN HMMER_PFM BLAST_DOMO BLIPS_BLOCKS BLIPS_PRINTS PROFILESAN

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO.	Uncyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	3750004CD1	470	S454 S51 S54 T104 T276 T386 T464	N182 N203	Thrombospondin type 1 domain tse_1: T34-C81 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS NEFROTICINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: Q119-C185 signal_peptide: M1-G29 signal_cleavage: M1-G24 Caspase recruitment domain CARD: A2-A91	HMMER_PPM BLAST_PRODOM HMMER SPScan HMMER_PPM BLAST_POMO
15	4904126CD1	110	S16 S36 T100 T49 N47		INTERLEUKIN-1 BETA CONVERTING ENZYME FAMILY HISTIDINE DM07463 P29466 1-122: M1-S110	

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	71268415CD1	879	S132 S14 S208 S288 S571 S711 S747 S754 S755 S827 T106 T118 T29 T30 T373 T412 T42 T428 T444 T55 T688 Y167 Y39	N368 N371 N569 N68	Reprolysin family propeptide Pep_M12B_Prop: R8-K119 Reprolysin (M12B) family zinc metalloprotease Disintegrin: E349-Q424 do-ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368 S60257 204-414; K126-D333 do ZINC; REGULATED; BRIDYVAL; NEUTRAL; DM00591 S60257 492-628; F410-I549 MELTRIN; BETA METALLOPROTEASE DISINTEGRIN BETA INTEGRIN PROTEASE METALLOPROTEASE PD105322; P620-G812 METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL PROTEIN TRANSMEMBRANE ADHESION PD000791; K134-P332 MELTRIN; BETA METALLOPROTEASE DISINTEGRIN MELTRIN BETA INTEGRIN PROTEASE METALLOPROTEASE PD171676; K495-C567 CELL ADHESION PLATELET BLOOD COAGULATION VENOM DISINTEGRIN METALLOPROTEASE PRECURSOR SIGNAL PD000664; E349-Y423 Neutral zinc metalloprotease E100142; P286-G276 DISINTEGRIN SIGNATURE PR00289; C380-K397; E403-N421 NEPRILYSIN METALLOPROTEASE PR00786C; N259-F273	HMMER_FFAM HMMER_FFAM HMMER_FFAM HMMER_FFAM BLAST_DOMO BLAST_DOMO BLAST_PRODOR BLAST_PRODOR BLAST_PRODOR BLAST_PRODOR BLAST_PRODOR BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Disintegrins signature disintegrins.prf:	Analytical Methods and Databases
17	7473301CD1	850	S100 S275 S295 S258 S429 S448 S470 S474 S495 S536 S596 S64 S787 S802 S807 T117 T250 T312 T348 T382 T404 T426 T570 T714 T777	N19 N210 N422 N486 N533 N559 N568	TRYPSIN FAMILY SERINE PROTEASES TRYPSIN: I613-1842 LOW-DENSITY LIPOPROTEIN RECEPTOR DOMAIN LDL_RECEPTOR_A: Q489-S527, P530-Q562, I564-C563 TRYPSIN DMO0018 P98072 800-1033: R612-V846 PROTEASE SERINE PROTEASOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIOMER FACTOR PD000046: Q675-I842, I613-6809 Serine proteases, trypsin family BL00134: C638-C654, D791-T814, P829-I842 Type I fibronectin domain BL01253: C638-A651, R790-C803, W811-Y845 Kringlike domain proteins BL00021: I722-G743, L801-I844, C638-P655 LOW DENSITY LIPOPROTEIN RECEPTOR DOMAIN PF00261: G501-E522 CHYMOTRYPSIN SERINE PROTEASE PF00722: G619-C654, *G697-W711, R790-S802	ES360-P419 Neutral zinc metalloproteinases, zinc-binding region signature zinc protease.prf: S249-G301 transmembrane domain: V624-Y645 Zn binding region Zinc_Protease: T2566-P275 HMMER MOTIFS HMMER_PPFAM HMMER_PPFAM BLAST_DMOO BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS	

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Trypsin family serine protease active sites trypsin_his.prf: L630-K679 trypsin_ser.prf: I776-R825 transmembrane domain: I77-L95 Trypsin family serine protease active sites Trypsin_His L649-C654 Trypsin_Ser D791-S802	PROFILERSCAN HMMER MOTIFS
18	7473308CD1	254	S136 S14 S153 S195 S227 T230 T249		TRYPSIN FAMILY SERINE PROTEASE TRYPSIN: I21-Q183 CHYMOTRYPSIN SERINE PROTEASE FAMILY PRO0722B: T89-A103 TRYPSIN DM00018 P07478 24-242: I21-Q183 PROTEASE SERINE PROTEASIN SIGNAL HYDROLASE ZINCEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PF000046: G23-Q183	HMMER_FFAM BLIPS_PRINTS BLAST_DOMO BLAST_PRODOR

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	74780210D1	568	S142 S145 S153 S172 S177 S190 S244 S316 S34 S420 S448 S552 T209 T22 T283 T334 T401 T427 T489 T79 Y509	N371	Matrixin Peptidase_M10: F56-T266 Hemopexin domain: F332-T390, I393-S448, L450-Q498, I505-K548 MATRIXINS CYSTEINE SWITCH DMQ0558 P22757 100-337: A184-T334, P76-P124 MATRIXINS CYSTEINE SWITCH DMQ0558 P08254 29-274: Q158-T334, L85-M122 MATRIX METALLOPROTEINASE PDI68921: S327-N392 MATRIX METALLOPROTEINASE PDI69970: A494-M568 MATRIX PRECURSOR METALLOPROTEINASE HYDROLASE ZINC ZYMOMEN CALCIUM COLLAGEN DEGRADATION SIGNAL PD000673: F171-T266, P73-M122 Matrixins cysteine switch BL00546: F92-D121, Y224-P267, G273-Y304, L313-G326, F443-Y455, F409-E428 Hemopexin domain Protein BL00024: M12-M122, G273-Y304, L313-G326, F443-Y455, Y408-D419 MATRIXIN SIGNATURE PRO0138: M12-P125, F198-P213, V224-W252, V279-Y304, L313-G326 Matrixins cysteine switch cysteine_switch.pif: A95-M204 Neutral zinc metalloproteinases, zinc-binding region signature zinc_protease.pif: D256-E312	HMMER_PPAM HMMER_PPAM ELASTO_DOMO ELASTO_DOMO ELASTO_PRODROM ELASTO_PRODROM ELASTO_PRODROM ELIMPS_BLOCKS ELIMPS_BLOCKS ELIMPS_PRINTS PROFILESSCAN PROFILESSCAN

Table 3 (cont.)

SEQ ID NO.	Incyte polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	4333459CD1	306	S117 S138 S2 S223 S60 S72 T110 T139 T207 T217	M108	Hemopexin domain signature hemopexin.prf: F409-R477 signal peptide: M1-P21 Zn binding region Zinc-Protease: V279-L248 signal cleavage: M1-P24 TRYPSIN FAMILY SERINE PROTEASE trypsin.i56-L298 TRYPSIN DM00018 Q05319 543-784: I56-L302 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZINCEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: S117-I298, I56-G192 Serine proteases, trypsin family BL00134: C81-C97, D238-G261, S283-L298 Type I fibronectin domain BL01253: C81-A94, G154-E190, R237-C250 Kringlike domain proteins BL00021: C81-I98, I165-G186, S247-F288 CHYMOTRYPSIN SERINE PROTEASE PR00722: G82-C97, P142-F156, R237-M249 Trypsin family serine protease active sites trypsin.his.prf: L73-G122 trypsin_ser.prf: K225-R271 Trypsin family serine protease active sites Trypsin.his: I92-C97 Trypsin_Ser: D238-M249 signal cleavage: M1-S26 SPSCAN HMMER_Pfam BLAST_DOMO BLAST_PRODOM BLINPS_BLOCKS BLINPS_BLOCKS BLINPS_BLOCKS BLINPS_PRINTS PROFILESCAN MOTIFS SPSCAN	PROFILESKAN HMMER MOTIFS SPSCAN

WO 01/98468

PCT/US01/19178

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	6817347CD1	953	S102 S114 S150 S172 S369 S429 S47 S623 S794 S804 S808 S831 S856 S919 S942 T289 T42 T455 T488 T544 T567 T568 T585 T59 T736 T777 T786 T839 Y929	N95	Ubiquitin carboxyl-terminal hydrolase family 1 UCH-1: R593-D624 Ubiquitin carboxyl-terminal hydrolase family 2 UCH-2: N875-K935 UBIQUITIN CARBOXYL-TERMINAL HYDROLASES FAMILY 2 DM00659[P40818 782-1103: T777-L931, L598-H709, I713-T753, V101-L128 PROTEASE UBIQUITIN HYDROLASE UBIQUITIN SPECIFIC ENZYME DEUBIQUITINATING C-TERMINAL THIOLESTERASE PROCESSING CONJUGATION PD017412: T777-E859 Ubiquitin carboxyl-terminal hydrolase family 2 BL00972: G594-L611, Y675-L684, I714-C728, K878-H902, K904-D925 Ubiquitin carboxyl-terminal hydrolase family 2 signature 1 Uch_2.1: G594-Q609 Ubiquitin carboxyl-terminal hydrolase family 2 signature 2 Uch_2.2: Y879-Y896	HMMER_FFAM HMMER_FFAM BLAST_DOMO BLAST_PRODOR BLIMPS_BLOCKS MOTIFS MOTIFS

WO 01/98468

PCT/US01/19178

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	275791CB1	2204	1168-1197, 1503-1522, 1-281, 1716-1738	6456514H1 (COLMDIC01) 7246467T8 (PROSTMY01) 4943009F8 (BRALFEN05) 55047202J1 6053385H1 (BRADIR03)	890 692 1319 1 1636	1510 1390 1877 811 2204
23	1389845CB1	2036	1-392, 1468-1491, 1334-1400, 1974-2036	F61389845 00001 1389845H1 (EOSINGT01)	6 1	2036 244
24	1726609CB1	2185	1-44, 1804- 2185	71762189Y1 5426388F9 (PROSTMY07) 71053940V1 71041539Y1 5968414H1 (BRANZNGT01) 6756865J1 (SINTFER02) 2953131H1 (98FINT01) 6440674H1 (BRANZNGT02) 8745066 7191212H2 (BRATDIC01) 5960039H1 (BRATNOT05) GB1.g7710158 -edit 60200050D1	1 1370 744 1954 1326 542 2885 2576 1831 2295 1229 1 825	662 1992 1373 2185 1624 1316 3136 3095 2254 2873 1797 2015 1146
25	4503848CB1	3486	1-1330	5969176H1 (BRANZNGT01) 5649471H1 (BRALFUT23) 2322143F6 (PROSNOT16) 3022114H1 (PROSIND01) 60220456D1	1677 3157 2235 3116 1085	2104 3486 2670 3402 1462

WO 01/98468

PCT/US01/19178

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	5544089CB1	2847	1260-1631, 2532-2847, 408-914	55051688J1 2344450F6 (TESTTUT02) 7658834H1 (OVARNOE02) 71763578V1 6576463H1 (COLTUS02) 55051680J1.comp	810 2195 2130 1562 1079 1	1767 2847 2642 2266 1838 958
27	7474081CB1	890	1-21	92103202 82142177	1 397	493 890
28	5281209CB1	1577	1-629	FL5281209.g7712102_000 004.g436191 G3644194	1 1 1155	1467 1467 1577
29	2256251CB1	1958	1-399, 896-935	3142938R6 (HNP2AS07) 3220594T6 (COLANON03) 92264312 F2256251.g7708357_000 002.g6105629	1045 1611 400 910 408	1376 1358 848 1830 1003
30	7160344CB1	3106	1-540, 1166-1428	256251R6 (OVARNOE01) 6471337H1 (PLACFEB01) 6854305H1 (BRATFEN08) 4443368H1 (SINNO701) 6894004J1 (BRATFEN03) 7655990H1 (UTREDE06) 7745974H1 (ADRETTUB04) 7160544H1 (HNP2TXC01) 70490289V1 70748463V1 7745974J1 (ADRETTUB04)	1654 895 1332 470 324 2378 1 2636 2269 1566	2316 1561 1585 1077 822 3052 427 3106 2891 2153

WO 01/98468

PCT/US01/19178

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	7477386CB1	3567	1-971, 1953-2846, 3243-3567	GBI:96682143_000029_ed it.20231-20345 GBI:96682143_000023.co mp_edit.11365-11445 GBI:96682143_000027_ed it.14110-14265 GBI:96682143_000029_ed it.35032-35202 GBI:96682143_000029_ed it.13651-13803 GBI:96682143_000019_ed it.3461-3655 GBI:96682143_000029_ed it.41513-41644 GBI:96682143_000029_ed it.43912-44404 GBI:96682143_000029_ed it.12846-12930 GBI:96682143_000029_ed it.15809-15912 mp_edit.3253-3653 it.27621-27746 GBI:96682143_000029_ed it.18722-18853 GBI:96682143_000029_ed it.17438-17581 GBI:96682143_000029_ed it.35651-35783	1495 1 523 2272 958 679 2944 3076 874 1111 82 2068 1363 1219 2608	1608 81 678 2436 1110 873 3075 3567 957 1218 522 2193 1494 1362 2739

WO 01/98468

PCT/US01/19178

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	275791CB1	2204	1168-1197,	6456514HL (COLANDIC01)	890	1510
			1503-1522,	724646778 (PROSTW01)	692	1390
			1-281,	4943009F8 (BRAIFN05)	1319	1877
			1716-1738	55047202J1	1	811
				6053385H1 (BRADIR03)	1636	2204
23	1389845CB1	2036	1-392,	FL1389845_00001	6	2036
			1468-1491, 1334-1400, 1974-2036	1389845H1 (EOSINOT01)	1	244
24	1726609CB1	2185	1-44, 1804-2185	71762189V1	1	662
				5426388P9 (PROSTW07)	1370	1992
				71063940V1	744	1373
				71041639V1	1954	2185
				5868414H1 (BRANOT01)	1326	1824
25	4503848CB1	3486	1-1330	675686501 (SINTFR02)	682	1316
				2053131H1 (REFINOT01)	2885	3136
				6440674H1 (BRAENOT02)	2576	3095
				95745066	1831	2254
				7191212H2 (BRATDIC01)	2295	2873
				5960039H1 (BRATNOT05)	1229	1797
				GB1.g7710158_edit	1	2015
				602005001	825	1146
				5969176H1 (BRANOT01)	1677	2104
				5649471H1 (BRATUT23)	3157	3486
				2322143F6 (PROSNOT16)	2235	2670
				3022114H1 (PROSDIN01)	3116	3402
				60220456D1	1085	1462

Table 4 (cont.)

Polynucleotide Seq ID NO:	Incyte Polynucleotide ID	Sequence length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	5544089C81	2847	1260-1631, 2532-2847, 408-914	55031688J1 234445056 (TESVTTT02) 765983481 (VARNOE02) 71763578V1 657646381 (COLHFTU02) 55051580J1.comp	810 2195 2130 1562 1079 1	1767 2847 2642 2266 1838
27	7474081C81	890	1-21	321030202 321042177 F05281209.g7712102_000 004_5436191 93644494	1 397 1 1155 1045	453 890 1467
28	5281209C81	1577	1-629	3142938R6 (HMT2AZS07) 3320504T6 (COLANON03) 54264312 FL3256251.g7708357_000 002_96103629	1611 1611 910	1577 1958 848 1830
29	2256251C81	1958	1-399, 896-935	2256251R6 (OVAKVTU01) 647133781 (PLACFES01) 695430551 (RDAIFEN08) 4443368E1 (SINDNFT03) 6894004J1 (GRATHTU03) 7655990H1 (UTREDE06) 7745974H1 (ADREPTU04) 7160544H1 (HNT2TXC01) 70490289V1 70748463V1	408 1654 895 2332 470 324 2378 1 2636 2269	1003 2136 1563 1585 1077 822 3052 427 3106 2891
30	7160544C81	3106	1-540, 1166-1428	7745974J1 (ADREPTU04)	1566	2153

WO 01/98468

PCT/US01/19178

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	7477386C81	3567	1-971, 1953-2846, 3243-3567	GBR:G6682143_000029_ed it.20231-20345 GBR:G6682143_000023.co mp_edit.11365-11445 GBR:G6682143_000027_ed it.14110-14265 GBR:G6682143_000029_ed it.35032-35202 GBR:G6682143_000029_ed it.13651-13803 GBR:G6682143_000019_ed it.3461-3655 GBR:G6682143_000029_ed it.41513-41644 GBR:G6682143_000029_ed it.43912-44404 GBR:G6682143_000029_ed it.12846-12930 GBR:G6682143_000029_ed it.15804-15912 GBR:G6682143_000023.co mp_edit.9253-96933 it.27828-27746 GBR:G6682143_000029_ed it.18722-18853 GBR:G6682143_000029_ed it.17458-17581 GBR:G6682143_000029_ed it.35651-35783	1495 1 523 2272 958 679 2944 3076 874 1111 82 2068 1363 1219 2608	1608 81 678 2436 1110 873 3075 3567 957 1218 522 2193 1494 1362 2739

WO 01/98468

PCT/US01/19178

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
32	7473089CB1	2930	1-632, 1082-1138, 2453-2555, 1373-1615, 1716-1740	GB1:g6682143_000029_ed	1759	1932
				it:27099-27217		
				GB1:g6682143_000029_sd	1609	1758
				it:24540-24713		
				GB1:g6682143_000029_sd	2740	2943
				it:37355-37558		
				GB1:g7387384_000011_co	2529	2930
				mp:edit:13491-13892		
				GB1:g7387384_000010_sd	1335	1619
				it:2924-3211		
				GB1:g7387384_000010_sd	2157	2528
				it:13479-13850		
				GB1:g7387384_000010_sd	1794	1979
				it:11350-11535		
				7986541H1 (UTRSTUC01)	1064	1587
				it:9694-9867	1620	1793
				GB1:g7387384_000012_co	75	1031
				mp:edit:9639-10595		
				GB1:g7387384_000010_sd	1032	1163
				it:1917-2074		
				GB1:g7387384_000010_sd	1164	1334
				it:2514-2664		
				7615482L (BBAFTUE03)	21	619
				GB1:g7387384_000010_sd	1380	2156
				it:11639-11815		
				GB1:g7387384_edit	1	2930

WO 01/98468

PCT/US01/19178

Table 4 (cont.)

Polynucleotide Seq ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
33	7604035CB1	4230	4185-4230, 894-2774	6254235H1 (LUNFTT02) 6213818H1 (MUSCDIT06) 6314348H1 (NERITD03) 7195502H1 (LUNGPER04) 6806634J1 (COLENO03) 8113675H1 (QSTEDNC01) 6804059H1 (COLENO03) 7750274H1 (HRAONOB01) 3843227F6 (DENDNGT01) 7632961H1 (BLADVTUE01) 5509797J1 7716357J1 (STNTPFE02)	3108 3900 3426 2758 2678 1661 478 1995 2094 1418 688 1 1	3897 4230 3994 3376 3323 2049 1050 2503 2707 2024 1535 678 2118
34	3473847CB1	3699	1-2631	71906145V1 7101935F8 (BRANFTDR02) 70857826V1 70855756V1 820867R1 (KRRANOT02) 8055446G1 (ESOGFTUE01) 70857738V1 70858612V1 GNN_g7208751_000002_00 2. edit 7101935R8 (BRANFTDR02)	168 168 2757 2650 2166 579 3156 1998 555 1	760 3441 3239 2732 1013 3699 2671 1850 473

WO 01/98468

PCT/US01/19178

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
35	3750004CB1	2410	1-264, 2116-2410, 1057-1167, 1590-1649	g7712021.ed	1	246
				7680089J1 (BRAFTUE01)	1327	1911
				6804411H1 (COLNOR03)	1088	1618
				71909368V1	536	968
				g1187194	1655	2127
				g2241985	706	1144
				6314962H1 (NERVTN03)	973	1135
				6823371J1 (SINROR01)	65	855
				7655009J1 (UTREME06)	1407	1990
				g1272147	1754	2410
36	4904126CB1	549		71620969V1	1	549
37	71268415CB1	2755	1-1097, 2326-2755	7715927J1 (SINPFE02)	590	1340
				772052H2 (BRAFFEE04)	2044	2514
				g6631070.cd	102	2755
				7723192J2 (HYHDE01)	905	1500
				g81:g7709257_000011.ed	1	139
				it		
				8037549H1 (SMCUNE01)	206	819
				7720289J1 (HYHDE01)	1596	2263
				8037549J1 (SMCUNE01)	1456	2120

WO 01/98468

PCT/US01/19178

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
38	7473301CB1	2553	1-2394	GBL.g7272157_000017.ed it 71704195V1 5544473H1 (TESTNOC01) GNN.g7272157_000017_00 2.edit 5544473H8 (TESTNOC01) 71703469V1 GNN.g8571511_000004_00 2.edit GNN.g6624046_000008_00 4	2001 1713 622 1688 2246 1163 981 1 1	2553 2016 680 2382 2550 1746 1468 1111 1041
39	7473308CB1	1041	826-1041, 1-289	GNN.g1552511_035	1	1041
40	7478021CB1	1707	1-1188	g8176728.edit g7684439.edit g7684439.edit_2	979 1 1	1083 978 1084
41	4333459CB1	1262	1-1262	71571956V1 5534851R8 (PLACFER01) 71571988V1 71573159V1 55022864H1 55022792H2 55022814H1 5502279502	704 1 256 247 2314 1392 2080 2044	1707 286 937 928 3067 2091 2886 2726
42	6817347CB1	3067	1-2270	GNN.g7417337_004.edit	1	3067

WO 01/98468

PCT/US01/19178

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
22	275791CB1	TESTNOF03
23	1389845CB1	ECSTXT01
24	1726609CB1	BRATUT02
25	4501848CB1	PROSN0T16
26	5544089CB1	BRATFC01
28	5281209CB1	HMTZAS07
29	2256251CB1	OVARCT01
30	7160544CB1	BRANGT02
32	7473089CB1	UTRSTUC01
33	7604035CB1	PLACN0R01
34	3473847CB1	KRRAN0T02
35	3750004CB1	BRAPT0R01
36	4904126CB1	TUTYMN0T08
37	71688415CB1	THYADIE01
38	7473501CB1	TESTNOC01
41	4333459CB1	KIDCTMT01
42	6817347CB1	ADRETUR01

WO 01/98468

PCT/US01/19178

Table 6

Library	Vector	Library Description
ADRETR01	PCDNA2.1	This random primed library was constructed using RNA isolated from left upper pole, adrenal gland tumor tissue removed from a 52-year-old Caucasian male during nephroureterectomy and local destruction of renal lesion. Pathology indicated grade 3 adrenal cortical carcinoma forming a mass that infiltrated almost the whole adrenal parenchyma and extended to adjacent adipose tissue. A metastatic tumor nodule was identified in the hilar region. The renal vein was infiltrated by tumor and the neoplastic process was present at the resection margin of the renal vein. Fragments of adrenal cortical carcinoma and thrombus were found in the inferior vena cava. Patient history included abnormal weight loss. Family history included skin cancer, type I diabetes, and neurotic depression.
BRAINOT02	pINCY	Library was constructed using RNA isolated from superior frontal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. In addition, scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver.
BRAPTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. The patient presented with migraine headache. The patient developed a cerebral hemorrhage and pulmonary edema, and died during this hospitalization. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Previous surgeries included a nephroureterectomy. Patient medications included Decadron and Dilantin. Family history included a malignant neoplasm of the kidney in the father.
BRAIFRC01	pINCY	This large size-fractionated library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.

WO 01/98468

PCT/US01/19178

Table 6 (cont.)

Library	Vector	Library Description
BRATWT02	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 59-year-old Caucasian male during excision of a cerebral meningial lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
HGSTVXT01	PINCY	Library was constructed using RNA isolated from eosinophils stimulated with IL-5.
HNTXAZS07	PSPORT1	This subtracted library was constructed from RNA isolated from an hMT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor) treated for three days with 0.35 micromolar A2. The hybridization probe for subtraction was derived from a similarly constructed library from untreated hMT2 cells. 3.08M clones from the A2-treated library were subjected to three rounds of subtractive hybridization with 3.04M clones from the untreated library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (NAR (1991) 19:1954) and Bonaldo et al. (Genome Research (1996) 6:791).
KERANOT02	PSPORT1	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK) NHEK (Clontech MCC-2501) is human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
KIDCTWT01	PINCY	Library was constructed using RNA isolated from kidney cortex tissue removed from a 65-year-old male during nephroureterectomy. Pathology for the associated tumor tissue indicated grade 3 renal cell carcinoma within the mid-portion of the kidney and the renal capsule.
OVANTUT01	PSPORT1	Library was constructed using RNA isolated from ovarian tumor tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology indicated grade 2 mucinous cystadenocarcinoma involving the entire left ovary. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.

WO 01/98468

PCT/US01/19178

Table 6 (cont.)

Library	Vector	Library Description
PLACNOR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from placental tissue removed from a Caucasian fetus (donor A), who died after 16 weeks' gestation from fetal demise and hydrocephalus and from placental tissue removed from a Caucasian male fetus (donor B), who died after 18 weeks' gestation from fetal demise. Patient history for donor A included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV and remaining serologies were negative. Family history included multiple pregnancies and live births, and an abortion in the mother. Serology was negative for donor B.
PROSNOT16	PINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
TESTNOC01	PELUESCRIPT	This large size fractionated library was constructed using RNA isolated from testicular tissue removed from a pool of eleven, 10 to 61-year-old Caucasian males.
TESTNOT03	PELUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
THYROIDEM1	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased thyroid tissue removed from a 22-year-old Caucasian female during closed thyroid biopsy, partial thyroidectomy, and regional lymph node excision. Pathology indicated adenomatous hyperplasia. The patient presented with malignant neoplasm of the thyroid. Patient history included normal delivery, alcohol abuse, and tobacco abuse. Previous surgeries included myringotomy. Patient medications included an unspecified type of birth control pills. Family history included hyperlipidemia and depressive disorder in the mother, and benign hypertension, congestive heart failure, and chronic leukemia in the grandparent(s).

WO 01/98468

PCT/US01/19178

Table 6 (cont.)

Library	Vector	Library Description
MTXANOT08	PBCV	The library was constructed using RNA isolated from anergic/allogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male. The cells were incubated for 3 days in the presence of 1 microgram/ml OKT3 mAb and 5% human serum.
UTRSTUC01	PSPORI1	This large size fractionated library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from uterus tumor tissue removed from a 37-year-old Black female (donor A) during myomectomy, dilation and curettage, right fimbrial region biopsy, and incidental appendectomy; and from endometrial tumor tissue removed from a 49-year-old Caucasian female (donor B) during vaginal hysterectomy and bilateral salpingo-oophorectomy. For donor A, pathology indicated multiple uterine leiomyomata. A fimbrial cyst was identified. The endometrium was in secretory phase with hormonal effect. The patient presented with deficiency anemia, an umbilical hernia, and premenopausal menorrhagia. Patient history included premenopausal menorrhagia and sarcoidosis of the lung. Previous surgeries included hysterectomy, dilation and curettage, and endoscopic lung biopsy. Patient medications included Chromagen and Claritin. For donor B, pathology indicated grade 3 adenocarcinoma forming a mass within the uterine fundus and involving the anterior uterine wall, as well as focally involving an adjacent endometrial polyp. The tumor invaded to a maximum depth of 7mm (uninvolved wall thickness, 2.2cm). The adjacent endometrium was inactive. Paraffin section immunostains for estrogen receptors and progesterone receptors were positive. Patient history included malignant breast neoplasm. Previous surgeries included unilateral extended simple mastectomy and bilateral tubal destruction. Patient medications included Megace and CAF (Cyclophosphamide, Adriamycin, Fluorouracil).

WO 01/98468

PCT/US01/19178

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL PDF	A Fast Data Finder useful in comparing and annotating 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: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, ifasta, fastx, tfastx, and search.	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; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks 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) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:89-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	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) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View in a Nishell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less. Signal peptide hits: Score= 0 or greater

WO 01/98468

PCT/US01/19178

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.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores: GCG-specified "HICG" value for that particular Prosite motif Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phila 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) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Conced	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 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) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol. 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.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetec Computer Group, Madison, WI.	

WO 01/98468

PCT/US01/19178

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of
 - 5 SEQ ID NO:1-21,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-21,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, and
 - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-21.
- 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:22-42.
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 01/98468

PCT/US01/19178

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting
 - 5 of SEQ ID NO:22-42,
 - b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:22-42,
 - c) a polynucleotide complementary to a polynucleotide of a),
 - d) a polynucleotide complementary to a polynucleotide of 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.
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.
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 a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

WO 01/98468

PCT/US01/19178

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

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.

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.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

20 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.

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.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

30 a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

WO 01/98468

PCT/US01/19178

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a 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:

- 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 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 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 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 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
- 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 01/98468

PCT/US01/19178

29. A diagnostic test for a condition or disease associated with the expression of PRTS in a biological sample comprising the steps of:

- 5 a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

- 10 a) a chimeric antibody,
b) a single chain antibody,
c) a Fab fragment,
d) a F(ab')₂ fragment, or
15 e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim

20 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim

25 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

- 30 a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
b) isolating antibodies from said animal; and

WO 01/98468

PCT/US01/19178

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

5 36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim
10 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

15 c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

20

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25 41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

30

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

WO 01/98468

PCT/US01/19178

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 in the sample.

- 5 44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-21 from a sample, the method comprising:
- a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an
- 10 amino acid sequence selected from the group consisting of SEQ ID NO:1-21.

 45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

15 46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

 47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

 48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

20 49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

 50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

25 51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

 52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

 53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30 54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

 55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

WO 01/98468

PCT/US01/19178

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.
67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.
68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.
69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:25.
70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.

WO 01/98468

PCT/US01/19178

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
5 NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:29.

10 74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:31.

15 76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
20 NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:34.

25 79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:36.

30 81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:37.

WO 01/98468

PCT/US01/19178

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:38.

83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
5 NO:39.

84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:40.

10 85. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:41.

86. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:42.

WO 01/98468

PCT/US01/19178

<110> INCYTE GENOMICS, INC.
 YUE, Henry
 ELLIOTT, Vicki
 GANDHI, Anesha R.
 LAL, Preeti
 AU-YOUNG, Janice
 TRIBOULEY, Catherine M.
 DELEGEANE, Angelo M.
 BAUGHN, Mariah R.
 NGUYEN, Dannie B.
 LEE, Ernestine A.
 HAPALIA, April
 KHAN, Farrah A.
 WALIA, Narinder K.
 YAO, Monique G.
 LU, Dyung Aina M.
 PATTERSON, Chandra
 TANG, Y. Tom
 WALSH, Roderick T.
 AZIMZAI, Yalda
 LU, Yan
 RAMKUMAR, Jayalaximi
 XU, Yuming
 REDDY, Roopa
 DAS, Depopriya
 KEARNEY, Liam
 KALLICK, Deborah A.

<120> Proteases

<130> PI-0123 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/212,336; 60/213,955; 60/215,396; 60/216,821; 60/218,946
 <151> 2000-06-16; 2000-06-22; 2000-06-29; 2000-07-07; 2000-07-14

<160> 42
 <170> PERL Program

<210> 1
 <211> 232
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 275791CD1

<400> 1
 Met Pro Glu Asn Pro Asp Thr Met Glu Thr Glu Lys Pro Lys Thr
 1 5 10 15
 Ile Thr Glu Leu Asp Pro Ala Ser Phe Thr Glu Ile Thr Lys Asp
 20 25 30
 Cys Asp Glu Asn Lys Glu Asn Lys Thr Pro Glu Gly Ser Gln Gly
 35 40 45
 Glu Val Asp Trp Leu Gln Gln Tyr Asp Met Glu Arg Glu Arg Glu
 50 55 60
 Glu Gln Glu Leu Gln Gln Ala Leu Ala Gln Ser Leu Gln Glu Gln
 65 70 75
 Glu Ala Trp Glu Gln Lys Glu Asp Asp Asp Leu Lys Arg Ala Thr
 80 85 90
 Glu Leu Ser Leu Gln Glu Phe Asn Asn Ser Phe Val Asp Ala Leu
 95 100 105
 Gly Ser Asp Glu Asp Ser Gly Asn Glu Asp Val Phe Asp Met Glu
 110 115 120

WO 01/98468

PCT/US01/19178

Tyr Thr Glu Ala Glu Ala Glu Glu Leu Lys Arg Asn Ala Glu Thr
 125 130 135
 Gly Asn Leu Pro His Ser Tyr Arg Leu Ile Ser Val Val Ser His
 140 145 150
 Ile Gly Ser Thr Ser Ser Ser Gly His Tyr Ile Ser Asp Val Tyr
 155 160 165
 Asp Ile Lys Lys Gln Ala Trp Phe Thr Tyr Asn Asp Leu Glu Val
 170 175 180
 Ser Lys Ile Gln Glu Ala Ala Val Gln Ser Asp Arg Asp Arg Ser
 185 190 195
 Gly Tyr Ile Phe Phe Tyr Met His Lys Glu Ile Phe Asp Glu Leu
 200 205 210
 Leu Glu Thr Glu Lys Asn Ser Gln Ser Leu Ser Thr Glu Val Gly
 215 220 225
 Lys Thr Thr Arg Gln Ala Ser
 230

<210> 2
 <211> 365
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1389845CD1

<400> 2
 Met Pro Lys Tyr Leu Gly Gly Gly Cys Cys Ile Pro Gly Pro Trp
 1 5 10 15
 Ala Glu Arg Arg Val Tyr Ser Leu Gly His Gln Asp Lys Ser Arg
 20 25 30
 Thr His Gln Glu Leu Arg Thr Asp Arg Arg Thr Thr Glu Gly Val
 35 40 45
 Thr Gly Trp Cys Glu Asp Trp Cys Pro Trp Ala Arg Thr Leu Leu
 50 55 60
 Ser Ser Pro Cys Trp Leu Gln Thr Arg Val Gln Ala Leu Gly Ser
 65 70 75
 Ala Thr Leu Thr Gln Pro Ser Leu Glu Asp Arg Met Arg Gly Val
 80 85 90
 Ser Cys Leu Gln Val Leu Leu Leu Leu Val Leu Gly Ala Ala Gly
 95 100 105
 Thr Gln Gly Arg Lys Ser Ala Ala Cys Gly Gln Pro Arg Met Ser
 110 115 120
 Ser Arg Ile Val Gly Gly Arg Asp Gly Arg Asp Gly Glu Trp Pro
 125 130 135
 Trp Gln Ala Ser Ile Gln His Arg Gly Ala His Val Cys Gly Gly
 140 145 150
 Ser Leu Ile Ala Pro Gln Trp Val Leu Thr Ala Ala His Cys Phe
 155 160 165
 Pro Arg Arg Ala Leu Pro Ala Glu Tyr Arg Val Arg Leu Gly Ala
 170 175 180
 Leu Arg Leu Gly Ser Thr Ser Pro Arg Thr Leu Ser Val Pro Val
 185 190 195
 Arg Arg Val Leu Leu Pro Pro Asp Tyr Ser Glu Asp Gly Ala Arg
 200 205 210
 Gly Asp Leu Ala Leu Leu Gln Leu Arg Arg Pro Val Pro Leu Ser
 215 220 225
 Ala Arg Val Gln Pro Val Cys Leu Pro Val Pro Gly Ala Arg Pro
 230 235 240
 Pro Pro Gly Thr Pro Cys Arg Val Thr Gly Trp Gly Ser Leu Arg
 245 250 255
 Pro Gly Val Pro Leu Pro Glu Trp Arg Pro Leu Gln Gly Val Arg
 260 265 270
 Val Pro Leu Leu Asp Ser Arg Thr Cys Asp Gly Leu Tyr His Val
 275 280 285
 Gly Ala Asp Val Pro Gln Ala Glu Arg Ile Val Leu Pro Gly Ser
 290 295 300

WO 01/98468

PCT/US01/19178

Leu Cys Ala Gly Tyr Pro Gln Gly His Lys Asp Ala Cys Gln Val
 305 310 315
 Cys Thr Gln Pro Pro Gln Pro Pro Glu Ser Pro Pro Cys Ala Gln
 320 325 330
 His Pro Pro Ser Leu Asn Ser Arg Thr Gln Asp Ile Pro Thr Gln
 335 340 345
 Ala Gln Asp Pro Gly Leu Gln Pro Arg Gly Thr Thr Pro Gly Val
 350 355 360
 Trp Asn Pro Glu Asn
 365

<210> 3
 <211> 416
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1726609CD1

<400> 3
 Met Trp Gly Arg Tyr Asp Ile Val Phe Leu Pro Pro Ser Phe Pro
 1 5 10 15
 Ile Val Ala Met Glu Asn Pro Cys Leu Thr Phe Ile Ile Ser Ser
 20 25 30
 Ile Leu Glu Ser Asp Glu Phe Leu Val Ile Asp Val Ile His Glu
 35 40 45
 Val Ala His Ser Trp Phe Gly Asn Ala Val Thr Asn Ala Thr Trp
 50 55 60
 Glu Glu Met Trp Leu Ser Glu Gly Leu Ala Thr Tyr Ala Gln Arg
 65 70 75
 Arg Ile Thr Thr Glu Thr Tyr Gly Ala Ala Phe Thr Cys Leu Glu
 80 85 90
 Thr Ala Phe Arg Leu Asp Ala Leu His Arg Gln Met Lys Leu Leu
 95 100 105
 Gly Glu Asp Ser Pro Val Ser Lys Leu Gln Val Lys Leu Glu Pro
 110 115 120
 Gly Val Asn Pro Ser His Leu Met Asn Leu Phe Thr Tyr Glu Lys
 125 130 135
 Gly Tyr Cys Phe Val Tyr Tyr Leu Ser Gln Leu Cys Gly Asp Pro
 140 145 150
 Gln Arg Phe Asp Asp Phe Leu Arg Ala Tyr Val Glu Lys Tyr Lys
 155 160 165
 Phe Thr Ser Val Val Ala Gln Asp Leu Leu Asp Ser Phe Leu Ser
 170 175 180
 Phe Phe Pro Glu Leu Lys Glu Gln Ser Val Asp Cys Arg Ala Gly
 185 190 195
 Leu Glu Phe Glu Arg Trp Leu Asn Ala Thr Gly Pro Pro Leu Ala
 200 205 210
 Glu Pro Asp Leu Ser Gln Gly Ser Ser Leu Thr Arg Pro Val Glu
 215 220 225
 Ala Leu Phe Gln Leu Trp Thr Ala Glu Pro Leu Asp Gln Ala Ala
 230 235 240
 Ala Ser Ala Ser Ala Ile Asp Ile Ser Lys Trp Arg Thr Phe Gln
 245 250 255
 Thr Ala Leu Phe Leu Asp Arg Leu Leu Asp Gly Ser Pro Leu Pro
 260 265 270
 Gln Glu Val Val Met Ser Leu Ser Lys Cys Tyr Ser Ser Leu Leu
 275 280 285
 Asp Ser Met Asn Ala Glu Ile Arg Ile Arg Trp Leu Gln Ile Val
 290 295 300
 Val Arg Asn Asp Tyr Tyr Pro Asp Leu His Arg Val Arg Arg Phe
 305 310 315
 Leu Glu Ser Gln Met Ser Arg Met Tyr Thr Ile Pro Leu Tyr Glu
 320 325 330
 Asp Leu Cys Thr Gly Ala Leu Lys Ser Phe Ala Leu Glu Val Phe
 335 340 345

WO 01/98468

PCT/US01/19178

Tyr Gln Thr Gln Gly Arg Leu His Pro Asn Leu Arg Arg Ala Ile
 350 355 360
 Gln Gln Ile Leu Ser Gln Gly Leu Gly Ser Ser Thr Glu Pro Ala
 365 370 375
 Ser Glu Pro Ser Thr Glu Leu Gly Lys Ala Glu Ala Asp Thr Asp
 380 385 390
 Ser Asp Ala Gln Ala Leu Leu Leu Gly Asp Glu Ala Pro Ser Ser
 395 400 405
 Ala Ile Ser Leu Arg Asp Val Asn Val Ser Ala
 410 415

<210> 4
 <211> 714
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4503848CD1

<400> 4
 Met His Ile His Met Leu Thr Leu Asp Gln Gln Lys Ser Leu Ile
 1 5 10 15
 Leu Ile Leu Phe Leu Ile Leu Phe Arg Val Gly Gly Ser Arg Ile
 20 25 30
 Leu Leu Arg Met Thr Leu Gly Arg Glu Val Met Ser Pro Leu Gln
 35 40 45
 Ala Met Ser Ser Tyr Thr Val Ala Gly Arg Asn Val Leu Arg Trp
 50 55 60
 Asp Leu Ser Pro Glu Gln Ile Lys Thr Arg Thr Glu Glu Leu Ile
 65 70 75
 Val Gln Thr Lys Gln Val Tyr Asp Ala Val Gly Met Leu Gly Ile
 80 85 90
 Glu Glu Val Thr Tyr Glu Asn Cys Leu Gln Ala Leu Ala Asp Val
 95 100 105
 Glu Val Lys Tyr Ile Val Glu Arg Thr Met Leu Asp Phe Pro Gln
 110 115 120
 His Val Ser Ser Asp Lys Glu Val Arg Ala Ala Ser Thr Glu Ala
 125 130 135
 Asp Lys Arg Leu Ser Arg Phe Asp Ile Glu Met Ser Met Arg Gly
 140 145 150
 Asp Ile Phe Glu Arg Ile Val His Leu Gln Glu Thr Cys Asp Leu
 155 160 165
 Gly Lys Ile Lys Pro Glu Ala Arg Arg Tyr Leu Glu Lys Ser Ile
 170 175 180
 Lys Met Gly Lys Arg Asn Gly Leu His Leu Pro Glu Gln Val Gln
 185 190 195
 Asn Glu Ile Lys Ser Met Lys Lys Arg Met Ser Glu Leu Cys Ile
 200 205 210
 Asp Phe Asn Lys Asn Leu Asn Glu Asp Asp Thr Phe Leu Val Phe
 215 220 225
 Ser Lys Ala Glu Leu Gly Ala Leu Pro Asp Asp Phe Ile Asp Ser
 230 235 240
 Leu Glu Lys Thr Asp Asp Asp Lys Tyr Lys Ile Thr Leu Lys Tyr
 245 250 255
 Pro His Tyr Phe Pro Val Met Lys Lys Cys Cys Ile Pro Glu Thr
 260 265 270
 Arg Arg Arg Met Glu Met Ala Phe Asn Thr Arg Cys Lys Glu Glu
 275 280 285
 Asn Thr Ile Ile Leu Gln Gln Leu Leu Pro Leu Arg Thr Lys Val
 290 295 300
 Ala Lys Leu Leu Gly Tyr Ser Thr His Ala Asp Phe Val Leu Glu
 305 310 315
 Met Asn Thr Ala Lys Ser Thr Ser Arg Val Thr Ala Phe Leu Asp
 320 325 330
 Asp Leu Ser Gln Lys Leu Lys Pro Leu Gly Glu Ala Glu Arg Glu
 335 340 345

WO 01/98468

PCT/US01/19178

Phe Ile Leu Asn Leu Lys Lys Lys Glu Cys Lys Asp Arg Gly Phe
 350 355 360
 Glu Tyr Asp Gly Lys Ile Asn Ala Trp Asp Leu Tyr Tyr Tyr Met
 365 370 375
 Thr Gln Thr Glu Glu Leu Lys Tyr Ser Ile Asp Gln Glu Phe Leu
 380 385 390
 Lys Glu Tyr Phe Pro Ile Glu Val Val Thr Glu Gly Leu Leu Asn
 395 400 405
 Thr Tyr Gln Glu Leu Leu Gly Leu Ser Phe Glu Gln Met Thr Asp
 410 415 420
 Ala His Val Trp Asn Lys Ser Val Thr Leu Tyr Thr Val Lys Asp
 425 430 435
 Lys Ala Thr Gly Glu Val Leu Gly Gln Phe Tyr Leu Asp Leu Tyr
 440 445 450
 Pro Arg Glu Gly Lys Tyr Asn His Ala Ala Cys Phe Gly Leu Gln
 455 460 465
 Pro Gly Cys Leu Leu Pro Asp Gly Ser Arg Met Met Ala Val Ala
 470 475 480
 Ala Leu Val Val Asn Phe Ser Gln Pro Val Ala Gly Arg Pro Ser
 485 490 495
 Leu Leu Arg His Asp Glu Val Arg Thr Tyr Phe His Glu Phe Gly
 500 505 510
 His Val Met His Gln Ile Cys Ala Gln Thr Asp Phe Ala Arg Phe
 515 520 525
 Ser Gly Thr Asn Val Glu Thr Asp Phe Val Glu Val Pro Ser Gln
 530 535 540
 Met Leu Glu Asn Trp Val Trp Asp Val Asp Ser Leu Arg Arg Leu
 545 550 555
 Ser Lys His Tyr Lys Asp Gly Ser Pro Ile Ala Asp Asp Leu Leu
 560 565 570
 Glu Lys Leu Val Ala Ser Arg Leu Val Asn Thr Gly Leu Leu Thr
 575 580 585
 Leu Arg Gln Ile Val Leu Ser Lys Val Asp Gln Ser Leu His Thr
 590 595 600
 Asn Thr Ser Leu Asp Ala Ala Ser Glu Tyr Ala Lys Tyr Cys Ser
 605 610 615
 Glu Ile Leu Gly Val Ala Ala Thr Pro Gly Thr Asn Met Pro Ala
 620 625 630
 Thr Phe Gly His Leu Ala Gly Gly Tyr Asp Gly Gln Tyr Tyr Gly
 635 640 645
 Tyr Leu Trp Ser Glu Val Phe Ser Met Asp Met Phe Tyr Ser Cys
 650 655 660
 Phe Lys Lys Glu Gly Ile Met Asn Pro Glu Val Gly Met Lys Tyr
 665 670 675
 Arg Asn Leu Ile Leu Lys Pro Gly Gly Ser Leu Asp Gly Met Asp
 680 685 690
 Met Leu His Asn Phe Leu Lys Arg Glu Pro Asn Gln Lys Ala Phe
 695 700 705
 Leu Met Ser Arg Gly Leu His Ala Pro
 710

<210> 5
 <211> 367
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5544089CD1

<400> 5
 Met Phe Ala Pro Ser Val Leu Ser Ser Gly Leu Ser Gly Gly Ala
 1 5 10 15
 Ser Lys Gly Arg Lys Met Glu Leu Ile Gln Pro Lys Glu Pro Thr
 20 25 30
 Ser Gln Tyr Ile Ser Leu Cys His Glu Leu His Thr Leu Phe Gln
 35 40 45

WO 01/98468

PCT/US01/19178

```

Val Met Trp Ser Gly Lys Trp Ala Leu Val Ser Pro Phe Ala Met
50 55 60
Leu His Ser Val Trp Arg Leu Ile Pro Ala Phe Arg Gly Tyr Ala
65 70 75
Gln Gln Asp Ala Gln Glu Phe Leu Cys Glu Leu Leu Asp Lys Ile
80 85 90
Gln Arg Glu Leu Glu Thr Thr Gly Thr Ser Leu Pro Ala Leu Ile
95 100 105
Pro Thr Ser Gln Arg Lys Leu Ile Lys Gln Val Leu Asn Val Val
110 115 120
Asn Asn Ile Phe His Gly Gln Leu Leu Ser Gln Val Thr Cys Leu
125 130 135
Ala Cys Asp Asn Lys Ser Asn Thr Ile Glu Pro Phe Trp Asp Leu
140 145 150
Ser Leu Glu Phe Pro Glu Arg Tyr Gln Cys Ser Gly Lys Asp Ile
155 160 165
Ala Ser Gln Pro Cys Leu Val Thr Glu Met Leu Ala Lys Phe Thr
170 175 180
Glu Thr Glu Ala Leu Glu Gly Lys Ile Tyr Val Cys Asp Gln Cys
185 190 195
Asn Ser Lys Arg Arg Arg Phe Ser Ser Lys Pro Val Val Leu Thr
200 205 210
Glu Ala Gln Lys Gln Leu Met Ile Cys His Leu Pro Gln Val Leu
215 220 225
Arg Leu His Leu Lys Arg Phe Arg Trp Ser Gly Arg Asn Asn Arg
230 235 240
Glu Lys Ile Gly Val His Val Gly Phe Glu Glu Ile Leu Asn Met
245 250 255
Glu Pro Tyr Cys Cys Arg Glu Thr Leu Lys Ser Leu Arg Pro Glu
260 265 270
Cys Phe Ile Tyr Asp Leu Ser Ala Val Val Met His His Gly Lys
275 280 285
Gly Phe Gly Ser Gly His Tyr Thr Ala Tyr Cys Tyr Asn Ser Glu
290 295 300
Gly Gly Phe Trp Val His Cys Asn Asp Ser Lys Leu Ser Met Cys
305 310 315
Thr Met Asp Glu Val Cys Lys Ala Gln Ala Tyr Ile Leu Phe Tyr
320 325 330
Thr Gln Arg Val Thr Glu Asn Gly His Ser Lys Leu Leu Pro Pro
335 340 345
Glu Leu Leu Leu Gly Ser Gln His Pro Asn Glu Asp Ala Asp Thr
350 355 360
Ser Ser Asn Glu Ile Leu Ser
365

```

<210> 6

<211> 235

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7474081CD1

<400> 6

```

Met Lys Tyr Val Phe Tyr Leu Gly Val Leu Ala Gly Thr Phe Phe
1 5 10 15
Phe Ala Asp Ser Ser Val Gln Lys Glu Asp Pro Ala Pro Tyr Leu
20 25 30
Val Tyr Leu Lys Ser His Phe Asn Pro Cys Val Gly Val Leu Ile
35 40 45
Lys Pro Ser Trp Val Leu Ala Pro Ala His Cys Tyr Leu Pro Asn
50 55 60
Leu Lys Val Met Leu Gly Asn Phe Lys Ser Arg Val Arg Asp Gly
65 70 75
Thr Glu Gln Thr Ile Asn Pro Ile Gln Ile Val Arg Tyr Trp Asn
80 85 90

```

WO 01/98468

PCT/US01/19178

```

Tyr Ser His Ser Ala Pro Gln Asp Asp Leu Met Leu Ile Lys Leu
95 100 105
Ala Lys Pro Ala Met Leu Asn Pro Lys Val Gln Pro Leu Thr Leu
110 115 120
Ala Thr Thr Asn Val Arg Pro Gly Thr Val Cys Leu Leu Ser Gly
125 130 135
Leu Asp Trp Ser Gln Glu Asn Ser Gly Arg His Pro Asp Leu Arg
140 145 150
Gln Asn Leu Glu Ala Pro Val Met Ser Asp Arg Glu Cys Gln Lys
155 160 165
Thr Glu Gln Gly Lys Ser His Arg Asn Ser Leu Cys Val Lys Phe
170 175 180
Val Lys Val Phe Ser Arg Ile Phe Gly Glu Val Ala Val Ala Thr
185 190 195
Val Ile Cys Lys Asp Lys Leu Gln Gly Ile Glu Val Gly His Phe
200 205 210
Met Gly Gly Asp Val Gly Ile Tyr Thr Asn Val Tyr Lys Tyr Val
215 220 225
Ser Trp Ile Glu Asn Thr Ala Lys Asp Lys
230 235

```

```

<210> 7
<211> 488
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 5281209CD1

```

```

<400> 7
Met Gln Pro Thr Gly Arg Glu Gly Ser Arg Ala Leu Ser Arg Arg
1 5 10 15
Tyr Leu Arg Arg Leu Leu Leu Leu Leu Leu Leu Leu Leu Arg
20 25 30
Gln Pro Val Thr Arg Ala Glu Thr Thr Pro Gly Ala Pro Arg Ala
35 40 45
Leu Ser Thr Leu Gly Ser Pro Ser Leu Phe Thr Thr Pro Gly Val
50 55 60
Pro Ser Ala Leu Thr Thr Pro Gly Leu Thr Thr Pro Gly Thr Pro
65 70 75
Lys Thr Leu Asp Leu Arg Gly Arg Ala Gln Ala Leu Met Arg Ser
80 85 90
Phe Pro Leu Val Asp Gly His Asn Asp Leu Pro Gln Val Leu Arg
95 100 105
Gln Arg Tyr Lys Asn Val Leu Gln Asp Val Asn Leu Arg Asn Phe
110 115 120
Ser His Gly Gln Thr Ser Leu Asp Arg Leu Arg Asp Gly Leu Val
125 130 135
Gly Ala Gln Phe Trp Ser Ala Ser Val Ser Cys Gln Ser Gln Asp
140 145 150
Gln Thr Ala Val Arg Leu Ala Leu Glu Gln Ile Asp Leu Ile His
155 160 165
Arg Met Cys Ala Ser Tyr Ser Glu Leu Glu Leu Val Thr Ser Ala
170 175 180
Glu Gly Leu Asn Ser Ser Gln Lys Leu Ala Cys Leu Ile Gly Val
185 190 195
Glu Gly Gly His Ser Leu Asp Ser Ser Leu Ser Val Leu Arg Ser
200 205 210
Phe Tyr Val Leu Gly Val Arg Tyr Leu Thr Leu Thr Phe Thr Cys
215 220 225
Ser Thr Pro Trp Ala Glu Ser Ser Thr Lys Phe Arg His His Met
230 235 240
Tyr Thr Asn Val Ser Gly Leu Thr Ser Phe Gly Glu Lys Val Val
245 250 255
Glu Glu Leu Asn Arg Leu Gly Met Met Ile Asp Leu Ser Tyr Ala
260 265 270

```

WO 01/98468

PCT/US01/19178

```

Ser Asp Thr Leu Ile Arg Arg Val Leu Glu Val Ser Gln Ala Pro
275 280 285
Val Ile Phe Ser His Ser Ala Ala Arg Ala Val Cys Asp Asn Leu
290 295 300
Leu Asn Val Pro Asp Asp Ile Leu Gln Leu Leu Lys Lys Asn Gly
305 310 315
Gly Ile Val Met Val Thr Leu Ser Met Gly Val Leu Gln Cys Asn
320 325 330
Leu Leu Ala Asn Val Ser Thr Val Ala Asp His Phe Asp His Ile
335 340 345
Arg Ala Val Ile Gly Ser Glu Phe Ile Gly Ile Gly Gly Asn Tyr
350 355 360
Asp Gly Thr Gly Arg Phe Pro Gln Gly Leu Glu Asp Val Ser Thr
365 370 375
Tyr Pro Val Leu Ile Glu Glu Leu Leu Ser Arg Ser Trp Ser Glu
380 385 390
Glu Glu Leu Gln Gly Val Leu Arg Gly Asn Leu Leu Arg Val Phe
395 400 405
Arg Gln Val Glu Lys Val Arg Glu Glu Ser Arg Ala Gln Ser Pro
410 415 420
Val Glu Ala Glu Phe Pro Tyr Gly Gln Leu Ser Thr Ser Cys His
425 430 435
Ser His Leu Val Pro Gln Asn Gly His Gln Ala Thr His Leu Glu
440 445 450
Val Thr Lys Gln Pro Thr Asn Arg Val Pro Trp Arg Ser Ser Asn
455 460 465
Ala Ser Pro Tyr Leu Val Pro Gly Leu Val Ala Ala Ala Thr Ile
470 475 480
Pro Thr Phe Thr Gln Trp Leu Cys
485

```

<210> 8

<211> 346

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2256251CD1

<400> 8

```

Met Leu Leu Gly Arg Val Trp Gln Thr Arg Glu Leu Lys Ser Lys
1 5 10
Val Pro Lys Lys Ala Gly Arg Cys Gly Gln Gly Arg Leu His Gly
20 25 30
Gly Ser Ala Val Gly Phe Leu Gly Ser Pro Pro Gly Thr Pro Ser
35 40 45
Ser Phe Asp Leu Gly Cys Gly Arg Pro Gln Val Ser Asp Ala Gly
50 55 60
Gly Arg Ile Val Gly Gly His Ala Ala Pro Ala Gly Ala Trp Pro
65 70 75
Trp Gln Ala Ser Leu Arg Leu Arg Arg Val His Val Cys Gly Gly
80 85 90
Ser Leu Leu Ser Pro Gln Trp Val Leu Thr Ala Ala His Cys Phe
95 100 105
Ser Gly Ser Leu Asn Ser Ser Asp Tyr Gln Val His Leu Gly Glu
110 115 120
Leu Glu Ile Thr Leu Ser Pro His Phe Ser Thr Val Arg Gln Ile
125 130 135
Ile Leu His Ser Ser Pro Ser Gly Gln Pro Gly Thr Ser Gly Asp
140 145 150
Ile Ala Leu Val Glu Leu Ser Val Pro Val Thr Leu Phe Ser Arg
155 160 165
Ile Leu Pro Val Cys Leu Pro Glu Ala Ser Asp Asp Phe Cys Pro
170 175 180
Gly Ile Arg Cys Trp Val Thr Gly Trp Gly Tyr Thr Arg Glu Gly
185 190 195

```

WO 01/98468

PCT/US01/19178

Glu Pro Leu Pro Pro Pro Tyr Ser Leu Arg Glu Val Lys Val Ser
 200 205 210
 Val Val Asp Thr Glu Thr Cys Arg Arg Asp Tyr Pro Gly Pro Gly
 215 220 225
 Gly Ser Ile Leu Gln Pro Asp Met Leu Cys Ala Arg Gly Pro Gly
 230 235 240
 Asp Ala Cys Gln Asp Asp Ser Gly Gly Pro Leu Val Cys Gln Val
 245 250 255
 Asn Gly Ala Trp Val Gln Ala Gly Ile Val Ser Trp Gly Glu Gly
 260 265 270
 Cys Gly Arg Pro Asn Arg Pro Gly Val Tyr Thr Arg Val Pro Ala
 275 280 285
 Tyr Val Asn Trp Ile Arg Arg His Ile Thr Ala Ser Gly Gly Ser
 290 295 300
 Glu Ser Gly Tyr Pro Arg Leu Pro Leu Leu Ala Gly Leu Phe Leu
 305 310 315
 Pro Gly Leu Phe Leu Leu Leu Val Ser Cys Val Leu Leu Ala Lys
 320 325 330
 Cys Leu Leu His Pro Ser Ala Asp Gly Thr Pro Phe Pro Ala Pro
 335 340 345
 Asp

<210> 9
 <211> 882
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7160544CD1

<400> 9
 Met Ala Ala Ala Met Glu Thr Glu Gln Leu Gly Val Glu Ile Phe
 1 5 10 15
 Glu Thr Ala Asp Cys Glu Glu Asn Ile Glu Ser Gln Asp Arg Pro
 20 25 30
 Lys Leu Glu Pro Phe Tyr Val Glu Arg Tyr Ser Trp Ser Gln Leu
 35 40 45
 Lys Lys Leu Leu Ala Asp Thr Arg Lys Tyr His Gly Tyr Met Met
 50 55 60
 Ala Lys Ala Pro His Asp Phe Met Phe Val Lys Arg Asn Asp Pro
 65 70 75
 Asp Gly Pro His Ser Asp Arg Ile Tyr Tyr Leu Ala Met Ser Gly
 80 85 90
 Glu Asn Arg Glu Asn Thr Leu Phe Tyr Ser Glu Ile Pro Lys Thr
 95 100 105
 Ile Asn Arg Ala Ala Val Leu Met Leu Ser Trp Lys Pro Leu Leu
 110 115 120
 Asp Leu Phe Gln Ala Thr Leu Asp Tyr Gly Met Tyr Ser Arg Glu
 125 130 135
 Glu Glu Leu Leu Arg Glu Arg Lys Arg Ile Gly Thr Val Gly Ile
 140 145 150
 Ala Ser Tyr Asp Tyr His Gln Gly Ser Gly Thr Phe Leu Phe Gln
 155 160 165
 Ala Gly Ser Gly Ile Tyr His Val Lys Asp Gly Gly Pro Gln Gly
 170 175 180
 Phe Thr Gln Gln Pro Leu Arg Pro Asn Leu Val Glu Thr Ser Cys
 185 190 195
 Pro Asn Ile Arg Met Asp Pro Lys Leu Cys Pro Ala Asp Pro Asp
 200 205 210
 Trp Ile Ala Phe Ile His Ser Asn Asp Ile Trp Ile Ser Asn Ile
 215 220 225
 Val Thr Arg Glu Glu Arg Arg Leu Thr Tyr Val His Asn Glu Leu
 230 235 240
 Ala Asn Met Glu Glu Asp Ala Arg Ser Ala Gly Val Ala Thr Phe
 245 250 255

WO 01/98468

PCT/US01/19178

Val Leu Gln Glu Glu Phe Asp Arg Tyr Ser Gly Tyr Trp Trp Cys 260 270
 Pro Lys Ala Glu Thr Thr Pro Ser Gly Gly Lys Ile Leu Arg Ile 265 275
 Leu Tyr Glu Glu Asn Asp Glu Ser Glu Val Glu Ile Ile His Val 280 285
 Thr Ser Pro Met Leu Glu Thr Arg Arg Ala Asp Ser Phe Arg Tyr 290 295
 Pro Lys Thr Gly Thr Ala Asn Pro Lys Val Thr Phe Lys Met Ser 300 305
 Glu Ile Met Ile Asp Ala Glu Gly Arg Ile Ile Asp Val Ile Asp 310 315
 Lys Glu Leu Ile Gln Pro Phe Glu Ile Leu Phe Glu Gly Val Glu 320 325
 Tyr Ile Ala Arg Ala Gly Trp Thr Pro Glu Gly Lys Tyr Ala Trp 330 335
 Ser Ile Leu Leu Asp Arg Ser Gln Thr Arg Leu Gln Ile Val Leu 340 345
 Ile Ser Pro Glu Leu Phe Ile Pro Val Glu Asp Asp Val Met Glu 350 355
 Arg Gln Arg Leu Ile Glu Ser Val Pro Asp Ser Val Thr Pro Leu 360 365
 Ile Ile Tyr Glu Glu Thr Thr Asp Ile Trp Ile Asn Ile His Asp 370 375
 Ile Phe His Val Phe Pro Gln Ser His Glu Glu Glu Ile Glu Phe 380 385
 Ile Phe Ala Ser Glu Phe Cys Lys Thr Gly Arg His Leu Tyr Lys 390 395
 Ile Thr Ser Ile Leu Lys Glu Ser Lys Tyr Lys Arg Ser Ser Gly 400 405
 Gly Leu Pro Ala Pro Ser Asp Phe Lys Cys Pro Ile Lys Glu Glu 410 415
 Ile Ala Ile Thr Ser Gly Glu Trp Glu Val Leu Gly Arg His Gly 420 425
 Ser Asn Ile Gln Val Asp Glu Val Arg Arg Leu Val Tyr Phe Glu 430 435
 Gly Thr Lys Asp Ser Pro Leu Glu His His Leu Tyr Val Val Ser 440 445
 Tyr Val Asn Pro Gly Glu Val Thr Arg Leu Thr Asp Arg Gly Tyr 450 455
 Ser His Ser Cys Cys Ile Ser Gln His Cys Asp Phe Phe Ile Ser 460 465
 Lys Tyr Ser Asn Gln Lys Asn Pro His Cys Val Ser Leu Tyr Lys 470 475
 Leu Ser Ser Pro Glu Asp Asp Pro Thr Cys Lys Thr Lys Glu Phe 480 485
 Trp Ala Thr Ile Leu Asp Ser Ala Gly Pro Leu Pro Asp Tyr Thr 490 495
 Pro Pro Glu Ile Phe Ser Phe Glu Ser Thr Thr Gly Phe Thr Leu 500 505
 Tyr Gly Met Leu Tyr Lys Pro His Asp Leu Gln Pro Gly Lys Lys 510 515
 Tyr Pro Thr Val Leu Phe Ile Tyr Gly Gly Pro Gln Val Gln Leu 520 525
 Val Asn Asn Arg Phe Lys Gly Val Lys Tyr Phe Arg Leu Asn Thr 530 535
 Leu Ala Ser Leu Gly Tyr Val Val Val Val Ile Asp Asn Arg Gly 540 545
 Ser Cys His Arg Gly Leu Lys Phe Glu Gly Ala Phe Lys Tyr Lys 550 555
 Met Gly Gln Ile Glu Ile Asp Asp Gln Val Glu Gly Leu Gln Tyr 560 565
 Leu Ala Ser Arg Tyr Asp Phe Ile Asp Leu Asp Arg Val Gly Ile 570 575
 His Gly Trp Ser Tyr Gly Gly Tyr Leu Ser Leu Met Ala Leu Met 580 585
 Gln Arg Ser Asp Ile Phe Arg Val Ala Ile Ala Gly Ala Pro Val 590 595

WO 01/98468

PCT/US01/19178

Thr	Leu	Trp	Ile	Phe	Tyr	Asp	Thr	Gly	Tyr	Thr	Glu	Arg	Tyr	Met	755
Gly	His	Pro	Asp	Gln	Asn	Glu	Gln	Gly	Tyr	Tyr	Leu	Gly	Ser	Val	760
Ala	Met	Gln	Ala	Glu	Lys	Phe	Pro	Ser	Glu	Pro	Asn	Arg	Leu	Leu	765
Leu	Leu	His	Gly	Phe	Leu	Asp	Glu	Asn	Val	His	Phe	Ala	His	Thr	770
Ser	Ile	Leu	Leu	Ser	Phe	Leu	Val	Arg	Ala	Gly	Lys	Pro	Tyr	Asp	775
Leu	Gln	Ile	Tyr	Pro	Gln	Glu	Arg	His	Ser	Ile	Arg	Val	Pro	Glu	780
Ser	Gly	Glu	His	Tyr	Glu	Leu	His	Leu	Leu	His	Tyr	Leu	Gln	Glu	785
Asn	Leu	Gly	Ser	Arg	Ile	Ala	Ala	Leu	Lys	Val	Ile				790

<210> 10

<211> 1189

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477386CD1

<400> 10

Met	Ala	Pro	Leu	Arg	Ala	Leu	Leu	Ser	Tyr	Leu	Leu	Pro	Leu	His	1
Cys	Ala	Leu	Cys	Ala	Ala	Ala	Gly	Ser	Arg	Thr	Pro	Glu	Leu	His	5
Leu	Ser	Gly	Lys	Leu	Ser	Asp	Tyr	Gly	Val	Thr	Val	Pro	Cys	Ser	20
Thr	Asp	Phe	Arg	Gly	Arg	Phe	Leu	Ser	His	Val	Val	Ser	Gly	Pro	35
Ala	Ala	Ala	Ser	Ala	Gly	Ser	Met	Val	Val	Asp	Thr	Pro	Pro	Thr	50
Leu	Pro	Arg	His	Ser	Ser	His	Leu	Arg	Val	Ala	Arg	Ser	Pro	Leu	65
His	Pro	Gly	Gly	Thr	Leu	Trp	Pro	Gly	Arg	Val	Gly	Arg	His	Ser	80
Leu	Tyr	Phe	Asn	Val	Thr	Val	Phe	Gly	Lys	Glu	Leu	His	Leu	Arg	95
Leu	Arg	Pro	Asn	Arg	Arg	Leu	Val	Val	Pro	Gly	Ser	Ser	Val	Glu	110
Trp	Gln	Glu	Asp	Phe	Arg	Glu	Leu	Phe	Arg	Gln	Pro	Leu	Arg	Gln	125
Glu	Cys	Val	Tyr	Thr	Gly	Gly	Val	Thr	Gly	Met	Pro	Gly	Ala	Ala	140
Val	Ala	Ile	Ser	Asn	Cys	Asp	Gly	Leu	Ala	Gly	Leu	Ile	Arg	Thr	155
Asp	Ser	Thr	Asp	Phe	Phe	Ile	Glu	Pro	Leu	Glu	Arg	Gly	Gln	Gln	170
Glu	Lys	Glu	Ala	Ser	Gly	Arg	Thr	His	Val	Val	Tyr	Arg	Arg	Glu	185
Ala	Val	Gln	Gln	Glu	Trp	Ala	Glu	Pro	Asp	Gly	Asp	Leu	His	Asn	200
Glu	Ala	Phe	Gly	Leu	Gly	Asp	Leu	Pro	Asn	Leu	Leu	Gly	Leu	Val	215
Gly	Asp	Gln	Leu	Gly	Asp	Thr	Glu	Arg	Lys	Arg	Arg	His	Ala	Lys	230
Pro	Gly	Ser	Tyr	Ser	Ile	Glu	Val	Leu	Leu	Val	Val	Asp	Asp	Ser	245
Val	Val	Arg	Phe	His	Gly	Lys	Glu	His	Val	Gln	Asn	Tyr	Val	Leu	260
Thr	Leu	Met	Asn	Ile	Val	Val	Asp	Glu	Ile	Tyr	His	Asp	Glu	Ser	275

WO 01/98468

PCT/US01/19178

290	295	300
Leu Gly Val His Ile Asn Ile Ala Leu Val Arg Leu Ile Met Val		
305	310	315
Gly Tyr Arg Gln Gln Ser Leu Ser Leu Ile Glu Arg Gly Asn Pro		
320	325	330
Ser Arg Ser Leu Glu Gln Val Cys Arg Trp Ala His Ser Gln Gln		
335	340	345
Arg Gln Asp Pro Ser His Ala Glu His His Asp His Val Val Phe		
350	355	360
Leu Thr Arg Gln Asp Phe Gly Pro Ser Gly Gly Tyr Ala Pro Val		
365	370	375
Thr Gly Met Cys His Pro Leu Arg Ser Cys Ala Leu Asn His Glu		
380	385	390
Asp Gly Phe Ser Ser Ala Phe Val Ile Ala His Glu Thr Gly His		
395	400	405
Val Leu Gly Met Glu His Asp Gly Gln Gly Asn Gly Cys Ala Asp		
410	415	420
Glu Thr Ser Leu Gly Ser Val Met Ala Pro Leu Val Gln Ala Ala		
425	430	435
Phe His Arg Phe His Trp Ser Arg Cys Ser Lys Leu Glu Leu Ser		
440	445	450
Arg Tyr Leu Pro Ser Tyr Asp Cys Leu Leu Asp Asp Pro Phe Asp		
455	460	465
Pro Ala Trp Pro Gln Pro Pro Glu Leu Pro Gly Ile Asn Tyr Ser		
470	475	480
Met Asp Glu Gln Cys Arg Phe Asp Phe Gly Ser Gly Tyr Gln Thr		
485	490	495
Cys Leu Ala Phe Arg Thr Phe Glu Pro Cys Lys Gln Leu Trp Cys		
500	505	510
Ser His Pro Asp Asn Pro Tyr Phe Cys Lys Thr Lys Lys Gly Pro		
515	520	525
Pro Leu Asp Gly Thr Glu Cys Ala Pro Gly Lys Trp Cys Phe Lys		
530	535	540
Gly His Cys Ile Trp Lys Ser Pro Glu Gln Thr Tyr Gly Gln Asp		
545	550	555
Gly Gly Trp Ser Ser Trp Thr Lys Phe Gly Ser Cys Ser Arg Ser		
560	565	570
Cys Gly Gly Gly Val Arg Ser Arg Ser Arg Ser Cys Asn Asn Pro		
575	580	585
Ser Pro Ala Tyr Gly Gly Arg Leu Cys Leu Gly Pro Met Phe Glu		
590	595	600
Tyr Gln Val Cys Asn Ser Glu Glu Cys Pro Gly Thr Tyr Glu Asp		
605	610	615
Phe Arg Ala Gln Gln Cys Ala Lys Arg Asn Ser Tyr Tyr Val His		
620	625	630
Gln Asn Ala Lys His Ser Trp Val Pro Tyr Glu Pro Asp Asp Asp		
635	640	645
Ala Gln Lys Cys Glu Leu Ile Cys Gln Ser Ala Asp Thr Gly Asp		
650	655	660
Val Val Phe Met Asn Gln Val Val His Asp Gly Thr Arg Cys Ser		
665	670	675
Tyr Arg Asp Pro Tyr Ser Val Cys Ala Arg Gly Glu Cys Val Pro		
680	685	690
Val Gly Cys Asp Lys Glu Val Gly Ser Met Lys Ala Asp Asp Lys		
695	700	705
Cys Gly Val Cys Gly Gly Asp Asn Ser His Cys Arg Thr Val Lys		
710	715	720
Gly Thr Leu Gly Lys Ala Ser Lys Gln Ala Gly Ala Leu Lys Leu		
725	730	735
Val Gln Ile Pro Ala Gly Ala Arg His Ile Gln Ile Glu Ala Leu		
740	745	750
Glu Lys Ser Pro His Arg Ile Val Val Lys Asn Gln Val Thr Gly		
755	760	765
Ser Phe Ile Leu Asn Pro Lys Gly Lys Glu Ala Thr Ser Arg Thr		
770	775	780
Phe Thr Ala Met Gly Leu Glu Trp Glu Asp Ala Val Glu Asp Ala		
785	790	795

WO 01/98468

PCT/US01/19178

Lys Glu Ser Leu Lys Thr Ser Gly Pro Leu Pro Glu Ala Ile Ala
 800 805 810
 Ile Leu Ala Leu Pro Pro Thr Glu Gly Gly Pro Arg Ser Ser Leu
 815 820 825
 Ala Tyr Lys Tyr Val Ile His Glu Asp Leu Leu Pro Leu Ile Gly
 830 835 840
 Ser Asn Asn Val Leu Leu Glu Glu Met Asp Thr Tyr Glu Trp Ala
 845 850 855
 Leu Lys Ser Trp Ala Pro Cys Ser Lys Ala Cys Gly Gly Gly Ile
 860 865 870
 Gln Phe Thr Lys Tyr Gly Cys Arg Arg Arg Arg Asp His His Met
 875 880 885
 Val Gln Arg His Leu Cys Asp His Lys Lys Arg Pro Lys Pro Ile
 890 895 900
 Arg Arg Arg Cys Asn Gln His Pro Cys Ser Gln Pro Val Trp Val
 905 910 915
 Thr Glu Glu Trp Gly Ala Cys Ser Arg Ser Cys Gly Lys Leu Gly
 920 925 930
 Val Gln Thr Arg Gly Ile Gln Cys Leu Leu Pro Leu Ser Asn Gly
 935 940 945
 Thr His Lys Val Met Pro Ala Lys Ala Cys Ala Gly Asp Arg Pro
 950 955 960
 Glu Ala Arg Arg Pro Cys Leu Arg Val Pro Cys Pro Ala Gln Trp
 965 970 975
 Arg Leu Gly Ala Trp Ser Gln Cys Ser Ala Thr Cys Gly Glu Gly
 980 985 990
 Ile Gln Gln Arg Gln Val Val Cys Arg Thr Asn Ala Asn Ser Leu
 995 1000 1005
 Gly His Cys Glu Gly Asp Arg Pro Asp Thr Val Gln Val Cys Ser
 1010 1015 1020
 Leu Pro Ala Cys Gly Ala Glu Pro Cys Thr Gly Asp Arg Ser Val
 1025 1030 1035
 Phe Cys Gln Met Glu Val Leu Asp Arg Tyr Cys Ser Ile Pro Gly
 1040 1045 1050
 Tyr His Arg Leu Cys Cys Val Ser Cys Ile Lys Lys Ala Ser Gly
 1055 1060 1065
 Pro Asn Pro Gly Pro Asp Pro Gly Pro Thr Ser Leu Pro Pro Phe
 1070 1075 1080
 Ser Thr Pro Gly Ser Pro Leu Pro Gly Pro Gln Asp Pro Ala Asp
 1085 1090 1095
 Ala Ala Glu Pro Pro Gly Lys Pro Thr Gly Ser Glu Asp His Gln
 1100 1105 1110
 His Gly Arg Ala Thr Gln Leu Pro Gly Ala Leu Asp Thr Ser Ser
 1115 1120 1125
 Pro Gly Thr Gln His Pro Phe Ala Pro Glu Thr Pro Ile Pro Gly
 1130 1135 1140
 Ala Ser Trp Ser Ile Ser Pro Thr Thr Pro Gly Gly Leu Pro Trp
 1145 1150 1155
 Gly Trp Thr Gln Thr Pro Thr Pro Val Pro Glu Asp Lys Gly Gln
 1160 1165 1170
 Pro Gly Glu Asp Leu Arg His Pro Gly Thr Ser Leu Pro Ala Ala
 1175 1180 1185
 Ser Pro Val Thr

<210> 11
 <211> 952
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7473089CD1

<400> 11
 Met Leu Leu Leu Gly Ile Leu Thr Leu Ala Phe Ala Gly Arg Thr
 1 5 10 15

WO 01/98468

PCT/US01/19178

Ala Gly Gly Ser Glu Pro Glu Arg Glu Val Val Val Pro Ile Arg
 20 25 30
 Leu Asp Pro Asp Ile Asn Gly Arg Arg Tyr Tyr Trp Arg Gly Pro
 35 40 45
 Glu Asp Ser Gly Asp Gln Gly Leu Ile Phe Gln Ile Thr Ala Phe
 50 55 60
 Gln Glu Asp Phe Tyr Leu His Leu Thr Pro Asp Ala Gln Phe Leu
 65 70 75
 Ala Pro Ala Phe Ser Thr Glu His Leu Gly Val Pro Leu Gln Gly
 80 85 90
 Leu Thr Gly Gly Ser Ser Asp Leu Arg Arg Cys Phe Tyr Ser Gly
 95 100 105
 Asp Val Asn Ala Glu Pro Asp Ser Phe Ala Ala Val Ser Leu Cys
 110 115 120
 Gly Gly Leu Arg Gly Ala Phe Gly Tyr Arg Gly Ala Glu Tyr Val
 125 130 135
 Ile Ser Pro Leu Pro Asn Ala Ser Ala Pro Ala Ala Gln Arg Asn
 140 145 150
 Ser Gln Gly Ala His Leu Leu Gln Arg Arg Gly Val Pro Gly Gly
 155 160 165
 Pro Ser Gly Asp Pro Thr Ser Arg Cys Gly Val Ala Ser Gly Trp
 170 175 180
 Asn Pro Ala Ile Leu Arg Ala Leu Asp Pro Tyr Lys Pro Arg Arg
 185 190 195
 Ala Gly Phe Gly Glu Ser Arg Ser Arg Arg Arg Ser Gly Arg Ala
 200 205 210
 Lys Arg Phe Val Ser Ile Pro Arg Tyr Val Glu Thr Leu Val Val
 215 220 225
 Ala Asp Glu Ser Met Val Lys Phe His Gly Ala Asp Leu Glu His
 230 235 240
 Tyr Leu Leu Thr Leu Leu Ala Thr Ala Ala Arg Leu Tyr Arg His
 245 250 255
 Pro Ser Ile Leu Asn Pro Ile Asn Ile Val Val Val Lys Val Leu
 260 265 270
 Leu Leu Arg Asp Arg Asp Ser Gly Pro Lys Val Thr Gly Asn Ala
 275 280 285
 Ala Leu Thr Leu Arg Asn Phe Cys Ala Trp Gln Lys Lys Leu Asn
 290 295 300
 Lys Val Ser Asp Lys His Pro Glu Tyr Trp Asp Thr Ala Ile Leu
 305 310 315
 Phe Thr Arg Gln Asp Leu Cys Gly Ala Thr Thr Cys Asp Thr Leu
 320 325 330
 Gly Met Ala Asp Val Gly Thr Met Cys Asp Pro Lys Arg Ser Cys
 335 340 345
 Ser Val Ile Glu Asp Asp Gly Leu Pro Ser Ala Phe Thr Thr Ala
 350 355 360
 His Glu Leu Gly His Val Phe Asn Met Pro His Asp Asn Val Lys
 365 370 375
 Val Cys Glu Glu Val Phe Gly Lys Leu Arg Ala Asn His Met Met
 380 385 390
 Ser Pro Thr Leu Ile Gln Ile Asp Arg Ala Asn Pro Trp Ser Ala
 395 400 405
 Cys Ser Ala Ala Ile Ile Thr Asp Phe Leu Asp Ser Gly His Gly
 410 415 420
 Asp Cys Leu Leu Asp Gln Pro Ser Lys Pro Ile Ser Leu Pro Glu
 425 430 435
 Asp Leu Pro Gly Ala Ser Tyr Thr Leu Ser Gln Gln Cys Glu Leu
 440 445 450
 Ala Phe Gly Val Gly Ser Lys Pro Cys Pro Tyr Met Gln Tyr Cys
 455 460 465
 Thr Lys Leu Trp Cys Thr Gly Lys Ala Lys Gly Gln Met Val Cys
 470 475 480
 Gln Thr Arg His Phe Pro Trp Ala Asp Gly Thr Ser Cys Gly Glu
 485 490 495
 Gly Lys Leu Cys Leu Lys Gly Ala Cys Val Glu Arg His Asn Leu
 500 505 510
 Asn Lys His Arg Val Asp Gly Ser Trp Ala Lys Trp Asp Pro Tyr

WO 01/98468

PCT/US01/19178

Gly Pro Cys Ser Arg Thr Cys Gly Gly Gly Val Gln Leu Ala Arg 515
 530 535 540
 Arg Gln Cys Thr Asn Pro Thr Pro Ala Asn Gly Gly Lys Tyr Cys 545
 550 555
 Glu Gly Val Arg Val Lys Tyr Arg Ser Cys Asn Leu Glu Pro Cys 560
 565 570
 Pro Ser Ser Ala Ser Gly Lys Ser Phe Arg Glu Glu Gln Cys Glu 575
 580 585
 Ala Phe Asn Gly Tyr Asn His Ser Thr Asn Arg Leu Thr Leu Ala 590
 595 600
 Val Ala Trp Val Pro Lys Tyr Ser Gly Val Ser Pro Arg Asp Lys 605
 610 615
 Cys Lys Leu Ile Cys Arg Ala Asn Gly Thr Gly Tyr Phe Tyr Val 620
 625 630
 Leu Ala Pro Lys Val Val Val Asp Gly Thr Leu Cys Ser Pro Asp 635
 640 645
 Ser Thr Ser Val Cys Val Gln Gly Lys Cys Ile Lys Ala Gly Cys 650
 655 660
 Asp Gly Asn Leu Gly Ser Lys Lys Arg Phe Asp Lys Cys Gly Val 665
 670 675
 Cys Gly Gly Asp Asn Lys Ser Cys Lys Lys Val Thr Gly Leu Phe 680
 685 690
 Thr Lys Pro Met His Gly Tyr Asn Phe Val Val Ala Ile Pro Ala 695
 700 705
 Gly Ala Ser Ser Ile Asp Ile Arg Gln Arg Gly Tyr Lys Gly Leu 710
 715 720
 Ile Gly Asp Asp Asn Tyr Leu Ala Leu Lys Asn Ser Gln Gly Lys 725
 730 735
 Tyr Leu Leu Asn Gly His Phe Val Val Ser Ala Val Glu Arg Asp 740
 745 750
 Leu Val Val Lys Gly Ser Leu Leu Arg Tyr Ser Gly Thr Gly Thr 755
 760 765
 Ala Val Glu Ser Leu Gln Ala Ser Arg Pro Ile Leu Glu Pro Leu 770
 775 780
 Thr Val Glu Val Leu Ser Val Gly Lys Met Thr Pro Pro Arg Val 785
 790 795
 Arg Tyr Ser Phe Tyr Leu Pro Lys Glu Pro Arg Glu Asp Lys Ser 800
 805 810
 Ser His Pro Pro His Pro Arg Gly Gly Pro Ser Val Leu His Asn 815
 820 825
 Ser Val Leu Ser Leu Ser Asn Gln Val Glu Gln Pro Asp Asp Arg 830
 835 840
 Pro Pro Ala Arg Trp Val Ala Gly Ser Trp Gly Pro Cys Ser Ala 845
 850 855
 Ser Cys Gly Ser Gly Leu Gln Lys Arg Ala Val Asp Trp Arg Gly 860
 865 870
 Ser Ala Gly Gln Arg Thr Val Pro Ala Cys Asp Ala Ala His Arg 875
 880 885
 Pro Val Glu Thr Gln Ala Cys Gly Glu Pro Cys Pro Thr Trp Glu 890
 895 900
 Leu Ser Ala Trp Ser Pro Cys Ser Lys Ser Cys Gly Arg Gly Phe 905
 910 915
 Gln Arg Arg Ser Leu Lys Cys Val Gly His Gly Gly Arg Leu Leu 920
 925 930
 Ala Arg Asp Glu Cys Asn Leu His Arg Lys Pro Gln Glu Leu Asp 935
 940 945
 Phe Cys Val Leu Arg Pro Cys 950

<210> 12
 <211> 898
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature

WO 01/98468

PCT/US01/19178

<223> Incyte ID No: 7604035CD1

```

<400> 12
Met Glu Asn Trp Thr Gly Arg Pro Trp Leu Tyr Leu Leu Leu Leu
1      5      10      15
Leu Ser Leu Pro Gln Leu Cys Leu Asp Gln Glu Val Leu Ser Gly
20     25     30
His Ser Leu Gln Thr Pro Thr Glu Glu Gly Gln Gly Pro Glu Gly
35     40     45
Val Trp Gly Pro Trp Val Gln Trp Ala Ser Cys Ser Gln Pro Cys
50     55     60
Gly Val Gly Val Gln Arg Arg Ser Arg Thr Cys Gln Leu Pro Thr
65     70     75
Val Gln Leu His Pro Ser Leu Pro Leu Pro Pro Arg Pro Pro Arg
80     85     90
His Pro Glu Ala Leu Leu Pro Arg Gly Gln Gly Pro Arg Pro Gln
95     100    105
Thr Ser Pro Glu Thr Leu Pro Leu Tyr Arg Thr Gln Ser Arg Gly
110    115    120
Arg Gly Gly Pro Leu Arg Gly Pro Ala Ser His Leu Gly Arg Glu
125    130    135
Glu Thr Gln Glu Ile Arg Ala Ala Arg Arg Ser Arg Leu Arg Asp
140    145    150
Pro Ile Lys Pro Gly Met Phe Gly Tyr Gly Arg Val Pro Phe Ala
155    160    165
Leu Pro Leu His Arg Asn Arg Arg His Pro Arg Ser Pro Pro Arg
170    175    180
Ser Glu Leu Ser Arg Ile Ser Ser Arg Gly Glu Glu Pro Ile Pro
185    190    195
Ser Pro Thr Pro Arg Ala Glu Pro Phe Ser Ala Asn Gly Ser Pro
200    205    210
Gln Thr Glu Leu Pro Pro Thr Glu Leu Ser Val His Thr Pro Ser
215    220    225
Pro Gln Ala Glu Pro Leu Ser Pro Glu Thr Ala Gln Thr Glu Val
230    235    240
Ala Pro Arg Thr Arg Pro Ala Pro Leu Arg His His Pro Arg Ala
245    250    255
Gln Ala Ser Gly Thr Glu Pro Pro Ser Pro Thr His Ser Leu Gly
260    265    270
Glu Gly Gly Phe Phe Arg Ala Ser Pro Gln Pro Arg Arg Pro Ser
275    280    285
Ser Gln Gly Trp Ala Ser Pro Gln Val Ala Gly Arg Arg Pro Asp
290    295    300
Pro Phe Pro Ser Val Pro Arg Gly Arg Gly Gln Gln Gly Gln Gly
305    310    315
Pro Trp Gly Thr Gly Gly Thr Pro His Gly Pro Arg Leu Glu Pro
320    325    330
Asp Pro Gln His Pro Gly Ala Trp Leu Pro Leu Leu Ser Asn Gly
335    340    345
Pro His Ala Ser Ser Leu Trp Ser Leu Phe Ala Pro Ser Ser Pro
350    355    360
Ile Pro Arg Cys Ser Gly Glu Ser Glu Gln Leu Arg Ala Cys Ser
365    370    375
Gln Ala Pro Cys Pro Pro Glu Gln Pro Asp Pro Arg Ala Leu Gln
380    385    390
Cys Ala Ala Phe Asn Ser Gln Glu Phe Met Gly Gln Leu Tyr Gln
395    400    405
Trp Glu Pro Phe Thr Glu Val Gln Gly Ser Gln Arg Cys Glu Leu
410    415    420
Asn Cys Arg Pro Arg Gly Phe Arg Phe Tyr Val Arg His Thr Glu
425    430    435
Lys Val Gln Asp Gly Thr Leu Cys Gln Pro Gly Ala Pro Asp Ile
440    445    450
Cys Val Ala Gly Arg Cys Leu Ser Pro Gly Cys Asp Gly Ile Leu
455    460    465
Gly Ser Gly Arg Arg Pro Asp Gly Cys Gly Val Cys Gly Gly Asp
470    475    480

```

WO 01/98468

PCT/US01/19178

```

Asp Ser Thr Cys Arg Leu Val Ser Gly Asn Leu Thr Asp Arg Gly
485 490 495
Gly Pro Leu Gly Tyr Gln Lys Ile Leu Trp Ile Pro Ala Gly Ala
500 505 510
Leu Arg Leu Gln Ile Ala Gln Leu Arg Pro Ser Ser Asn Tyr Leu
515 520 525
Ala Leu Arg Gly Pro Gly Gly Arg Ser Ile Ile Asn Gly Asn Trp
530 535 540
Ala Val Asp Pro Pro Gly Ser Tyr Arg Ala Gly Gly Thr Val Phe
545 550 555
Arg Tyr Asn Arg Pro Pro Arg Glu Glu Gly Lys Gly Glu Ser Leu
560 565 570
Ser Ala Glu Gly Pro Thr Thr Gln Pro Val Asp Val Tyr Met Ile
575 580 585
Phe Gln Glu Glu Asn Pro Gly Val Phe Tyr Gln Tyr Val Ile Ser
590 595 600
Ser Pro Pro Pro Ile Leu Glu Asn Pro Thr Pro Glu Pro Pro Val
605 610 615
Pro Gln Leu Gln Pro Glu Ile Leu Arg Val Glu Pro Pro Leu Ala
620 625 630
Pro Ala Pro Arg Pro Ala Arg Thr Pro Gly Thr Leu Gln Arg Gln
635 640 645
Val Arg Ile Pro Gln Met Pro Ala Pro Pro His Pro Arg Thr Pro
650 655 660
Leu Gly Ser Pro Ala Ala Tyr Trp Lys Arg Val Gly His Ser Ala
665 670 675
Cys Ser Ala Ser Cys Gly Lys Gly Val Trp Arg Pro Ile Phe Leu
680 685 690
Cys Ile Ser Arg Glu Ser Gly Glu Glu Leu Asp Glu Arg Ser Cys
695 700 705
Ala Ala Gly Ala Arg Pro Pro Ala Ser Pro Glu Pro Cys His Gly
710 715 720
Thr Pro Cys Pro Pro Tyr Trp Glu Ala Gly Glu Trp Thr Ser Cys
725 730 735
Ser Arg Ser Cys Gly Pro Gly Thr Gln His Arg Gln Leu Gln Cys
740 745 750
Arg Gln Glu Phe Gly Gly Gly Gly Ser Ser Val Pro Pro Glu Arg
755 760 765
Cys Gly His Leu Pro Arg Pro Asn Ile Thr Gln Ser Cys Gln Leu
770 775 780
Arg Leu Cys Gly His Trp Glu Val Gly Ser Pro Trp Ser Gln Cys
785 790 795
Ser Val Arg Cys Gly Arg Gly Gln Arg Ser Arg Gln Val Arg Cys
800 805 810
Val Gly Asn Asn Gly Asp Glu Val Ser Glu Gln Glu Cys Ala Ser
815 820 825
Gly Pro Pro Gln Pro Pro Ser Arg Glu Ala Cys Asp Met Gly Pro
830 835 840
Cys Thr Thr Ala Trp Phe His Ser Asp Trp Ser Ser Lys Cys Ser
845 850 855
Ala Glu Cys Gly Thr Gly Ile Gln Arg Arg Ser Val Val Cys Leu
860 865 870
Gly Ser Gly Ala Ala Thr Arg Ala Arg Pro Gly Gly Ser Arg Ser
875 880 885
Arg Asn Trp Ala Glu Leu Ser Asn Arg Lys Pro Ala Pro
890 895

```

<210> 13

<211> 631

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3473847CD1

<400> 13

WO 01/98468

PCT/US01/19178

```

Met Phe Leu Leu Ala Trp Gly Gln Asp Pro Trp Arg Leu Pro Gly
1      5      10      15
Thr Tyr Val Val Val Leu Lys Glu Glu Thr His Leu Ser Gln Ser
20      25      30
Glu Arg Thr Ala Arg Arg Leu Gln Ala Gln Ala Ala Arg Arg Gly
35      40      45
Tyr Leu Thr Lys Ile Leu His Val Phe His Gly Leu Leu Pro Gly
50      55      60
Phe Leu Val Lys Met Ser Gly Asp Leu Leu Glu Leu Ala Leu Lys
65      70      75
Leu Pro His Val Asp Tyr Ile Glu Glu Asp Ser Ser Val Phe Ala
80      85      90
Gln Ser Ile Pro Trp Asn Leu Glu Arg Ile Thr Pro Pro Arg Tyr
95      100      105
Arg Ala Asp Glu Tyr Gln Pro Pro Asp Gly Gly Ser Leu Val Glu
110      115      120
Val Tyr Leu Leu Asp Thr Ser Ile Gln Ser Asp His Arg Glu Ile
125      130      135
Glu Gly Arg Val Met Val Thr Asp Phe Glu Asn Val Pro Glu Glu
140      145      150
Asp Gly Thr Arg Phe His Arg Gln Ala Ser Lys Cys Asp Ser His
155      160      165
Gly Thr His Leu Ala Gly Val Val Ser Gly Arg Asp Ala Gly Val
170      175      180
Ala Lys Gly Ala Ser Met Arg Ser Leu Arg Val Leu Asn Cys Gln
185      190      195
Gly Lys Gly Thr Val Ser Gly Thr Leu Ile Gly Leu Glu Phe Ile
200      205      210
Arg Lys Ser Gln Leu Val Gln Pro Val Gly Pro Leu Val Val Leu
215      220      225
Leu Pro Leu Ala Gly Gly Tyr Ser Arg Val Leu Asn Ala Ala Cys
230      235      240
Gln Arg Leu Ala Arg Ala Gly Val Val Leu Val Thr Ala Ala Gly
245      250      255
Asn Phe Arg Asp Asp Ala Cys Leu Tyr Ser Pro Ala Ser Ala Pro
260      265      270
Glu Val Ile Thr Val Gly Ala Thr Asn Ala Gln Asp Gln Pro Val
275      280      285
Thr Leu Gly Thr Leu Gly Thr Asn Phe Gly Arg Cys Val Asp Leu
290      295      300
Phe Ala Pro Gly Glu Asp Ile Ile Gly Ala Ser Ser Asp Cys Ser
305      310      315
Thr Cys Phe Val Ser Gln Ser Gly Thr Ser Gln Ala Ala Ala His
320      325      330
Val Ala Gly Ile Ala Ala Met Met Leu Ser Ala Glu Pro Glu Leu
335      340      345
Thr Leu Ala Glu Leu Arg Gln Arg Leu Ile His Phe Ser Ala Lys
350      355      360
Asp Val Ile Asn Glu Ala Trp Phe Pro Glu Asp Gln Arg Val Leu
365      370      375
Thr Pro Asn Leu Val Ala Ala Leu Pro Pro Ser Thr His Gly Ala
380      385      390
Gly Trp Gln Leu Phe Cys Arg Thr Val Trp Ser Ala His Ser Gly
395      400      405
Pro Thr Arg Met Ala Thr Ala Ile Ala Arg Cys Ala Pro Asp Glu
410      415      420
Glu Leu Leu Ser Cys Ser Ser Phe Ser Arg Ser Gly Lys Arg Arg
425      430      435
Gly Glu Arg Met Glu Ala Gln Gly Gly Lys Leu Val Cys Arg Ala
440      445      450
His Asn Ala Phe Gly Gly Glu Gly Val Tyr Ala Ile Ala Arg Cys
455      460      465
Cys Leu Leu Pro Gln Ala Asn Cys Ser Val His Thr Ala Pro Pro
470      475      480
Ala Glu Ala Ser Met Gly Thr Arg Val His Cys His Gln Gln Gly
485      490      495
His Val Leu Thr Gly Cys Ser Ser His Trp Glu Val Glu Asp Leu

```

WO 01/98468

PCT/US01/19178

Gly Thr His Lys	500	Pro Val Leu Arg	505	Pro Arg Gly Gln Pro	510
Gln Cys Val Gly	515	His Arg Glu Ala Ser	520	Ile His Ala Ser Cys	525
His Ala Pro Gly	530	Leu Glu Cys Lys Val	535	Glu His Gly Ile	540
Ala Pro Gln Glu	545	Gln Val Thr Val Ala	550	Glu Glu Gly Trp	555
Leu Thr Gly Cys	560	Ala Leu Pro Gly	565	Ser His Val Leu	570
Ala Tyr Ala Val	575	Asn Thr Cys Val	580	Arg Ser Arg Asp	585
Ser Thr Thr Gly	590	Ser Thr Ser Glu Glu	595	Ala Val Thr Ala Val	600
Ile Cys Cys Arg	605	Arg His Leu Ala	610	Gln Ala Ser Gln Glu	615
Gln	620		625		630

<210> 14
 <211> 470
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3750004CD1

Met Arg His Arg Thr	1	Asp Leu Gly Gln Asn	10	Leu Leu Leu Phe Leu	15
Trp Ala Leu Leu Asn	5	Cys Gly Leu Gly Val	25	Ser Ala Gln Gly Pro	30
Gly Glu Trp Thr	20	Trp Val Ser Trp Thr	40	Arg Cys Ser Ser	45
Cys Gly Arg Gly Val	35	Ser Val Arg Ser Arg	55	Arg Cys Leu Arg Leu	60
Pro Gly Glu Glu Pro	50	Cys Trp Gly Asp Ser	65	His Glu Tyr Arg Leu	70
Cys Gln Leu Pro Asp	65	Cys Pro Pro Gly Ala	75	Val Pro Phe Arg Asp	80
Leu Gln Cys Ala Leu	80	Tyr Asn Gly Arg Pro	85	Val Leu Gly Thr Gln	90
Lys Thr Tyr Gln Trp	95	Val Pro Phe His Gly	100	Ala Pro Asn Gln Cys	105
Asp Leu Asn Cys Leu	110	Ala Glu Gly His Ala	115	Phe Tyr His Ser Phe	120
Gly Arg Val Leu Asp	125	Gly Thr Ala Cys Ser	130	Pro Gly Ala Gln Gly	135
Val Cys Val Ala Gly	140	Arg Cys Leu Ser Ala	145	Gly Cys Asp Gly Leu	150
Leu Gly Ser Gly Ala	155	Leu Glu Asp Arg Cys	160	Gly Arg Cys Gly Gly	165
Ala Asn Asp Ser Cys	170	Leu Phe Val Gln Arg	175	Val Phe Arg Asp Ala	180
Gly Ala Phe Ala Gly	185	Tyr Trp Asn Val Thr	190	Leu Ile Pro Glu Gly	195
Ala Arg His Ile Arg	200	Val Glu His Arg Ser	205	Arg Asn His Leu Gly	210
Ile Leu Gly Ser Leu	215	Met Gly Gly Asp Gly	220	Arg Tyr Val Leu Asn	225
Gly His Trp Val Val	230	Ser Pro Pro Gly Thr	235	Tyr Glu Ala Ala Gly	240
Thr His Val Val Tyr	245	Thr Arg Asp Thr Gly	250	Pro Gln Glu Thr Leu	255
Gln Ala Ala Gly Pro	260	Thr Ser His Asp Leu	265	Leu Leu Gln Val Leu	270

WO 01/98468 PCT/US01/19178

Leu	Gln	Glu	Pro	Asn	Pro	Gly	Ile	Glu	Phe	Glu	Phe	Trp	Leu	Pro
Arg	Glu	Arg	Tyr	Ser	Pro	Phe	Gln	Ala	Arg	Val	Gln	Ala	Leu	Gly
Trp	Pro	Leu	Arg	Gln	Pro	Gln	Pro	Arg	Gly	Val	Glu	Pro	Gln	Pro
Pro	Ala	Ala	Pro	Ala	Val	Thr	Pro	Ala	Gln	Thr	Pro	Thr	Leu	Ala
Pro	Asp	Pro	Cys	Pro	Pro	Cys	Pro	Asp	Thr	Arg	Gly	Arg	Ala	His
Arg	Leu	Leu	His	Tyr	Cys	Gly	Ser	Asp	Phe	Val	Phe	Gln	Ala	Arg
Val	Leu	Gly	His	His	His	Gln	Ala	Gln	Glu	Thr	Arg	Tyr	Glu	Val
Arg	Ile	Gln	Leu	Val	Tyr	Lys	Asn	Arg	Ser	Pro	Leu	Arg	Ala	Arg
Glu	Tyr	Val	Trp	Ala	Pro	Gly	His	Cys	Pro	Cys	Pro	Met	Leu	Ala
Pro	His	Arg	Asp	Tyr	Leu	Met	Ala	Val	Gln	Arg	Leu	Val	Ser	Pro
Asp	Gly	Thr	Gln	Asp	Gln	Leu	Leu	Leu	Pro	His	Ala	Gly	Tyr	Ala
Arg	Pro	Trp	Ser	Pro	Ala	Glu	Asp	Ser	Arg	Ile	Arg	Leu	Thr	Ala
Arg	Arg	Cys	Pro	Gly										

275 280 285
 290 295 300
 305 310 315
 320 325 330
 335 340 345
 350 355 360
 365 370 375
 380 385 390
 395 400 405
 410 415 420
 425 430 435
 440 445 450
 455 460 465
 470

<210> 15
 <211> 110
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 4904126CD1
 <400> 15
 Met Ala Asp Lys Val Leu Lys Glu Lys Arg Lys Gln Phe Ile Arg
 1 5 10 15
 Ser Val Gly Glu Gly Thr Ile Asn Gly Leu Leu Gly Glu Leu Leu
 20 25 30
 Glu Thr Arg Val Leu Ser Gln Glu Glu Ile Glu Ile Val Lys Cys
 35 40 45
 Glu Asn Ala Thr Val Met Asp Lys Ala Arg Ala Leu Leu Asp Ser
 50 55 60
 Val Ile Arg Lys Gly Ala Pro Ala Cys Gln Ile Cys Ile Thr Tyr
 65 70 75
 Ile Cys Glu Glu Asp Ser His Leu Ala Gly Thr Leu Gly Leu Ser
 80 85 90
 Ala Gly Pro Thr Ser Gly Asn His Leu Thr Thr Gln Asp Ser Gln
 95 100 105
 Ile Val Leu Pro Ser
 110

<210> 16
 <211> 879
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 71268415CD1
 <400> 16
 Met Ser Leu Phe Ile Phe Cys Arg Gln Leu Phe Ala Pro Ser Tyr
 1 5 10 15

WO 01/98468

PCT/US01/19178

Thr Glu Thr His Tyr Thr Ser Ser Gly Asn Pro Gln Thr Thr Thr
 20 25 30
 Arg Lys Leu Glu Asp His Cys Phe Tyr His Gly Thr Val Arg Glu
 35 40 45
 Thr Glu Leu Ser Ser Val Thr Leu Ser Thr Cys Arg Gly Ile Arg
 50 55 60
 Gly Leu Ile Thr Val Ser Ser Asn Leu Ser Tyr Val Ile Glu Pro
 65 70 75
 Leu Pro Asp Ser Lys Gly Gln His Leu Ile Tyr Arg Ser Glu His
 80 85 90
 Leu Lys Pro Pro Pro Gly Asn Cys Gly Phe Glu His Ser Lys Pro
 95 100 105
 Thr Thr Arg Asp Trp Ala Leu Gln Phe Thr Gln Gln Thr Lys Lys
 110 115 120
 Arg Pro Arg Arg Met Lys Arg Glu Asp Leu Asn Ser Met Lys Tyr
 125 130 135
 Val Glu Leu Tyr Leu Val Ala Asp Tyr Leu Glu Phe Gln Lys Asn
 140 145 150
 Arg Arg Asp Gln Asp Ala Thr Lys His Lys Leu Ile Glu Ile Ala
 155 160 165
 Asn Tyr Val Asp Lys Phe Tyr Arg Ser Leu Asn Ile Arg Ile Ala
 170 175 180
 Leu Val Gly Leu Glu Val Trp Thr His Gly Asn Met Cys Glu Val
 185 190 195
 Ser Glu Asn Pro Tyr Ser Thr Leu Trp Ser Phe Leu Ser Trp Arg
 200 205 210
 Arg Lys Leu Leu Ala Gln Lys Tyr His Asp Asn Ala Gln Leu Ile
 215 220 225
 Thr Gly Met Ser Phe His Gly Thr Thr Ile Gly Leu Ala Pro Leu
 230 235 240
 Met Ala Met Cys Ser Val Tyr Gln Ser Gly Gly Val Asn Met Asp
 245 250 255
 His Ser Glu Asn Ala Ile Gly Val Ala Ala Thr Met Ala His Glu
 260 265 270
 Met Gly His Asn Phe Gly Met Thr His Asp Ser Ala Asp Cys Cys
 275 280 285
 Ser Ala Ser Ala Ala Asp Gly Gly Cys Ile Met Ala Ala Ala Thr
 290 295 300
 Gly His Pro Phe Pro Lys Val Phe Asn Gly Cys Asn Arg Arg Glu
 305 310 315
 Leu Asp Arg Tyr Leu Gln Ser Gly Gly Gly Met Cys Leu Ser Asn
 320 325 330
 Met Pro Asp Thr Arg Met Leu Tyr Gly Gly Arg Arg Cys Gly Asn
 335 340 345
 Gly Tyr Leu Glu Asp Gly Glu Glu Cys Asp Cys Gly Glu Glu Glu
 350 355 360
 Glu Cys Asn Asn Pro Cys Cys Asn Ala Ser Asn Cys Thr Leu Arg
 365 370 375
 Pro Gly Ala Glu Cys Ala His Gly Ser Cys Cys His Gln Cys Lys
 380 385 390
 Leu Leu Ala Pro Gly Thr Leu Cys Arg Glu Gln Ala Arg Gln Cys
 395 400 405
 Asp Leu Pro Glu Phe Cys Thr Gly Lys Ser Pro His Cys Pro Thr
 410 415 420
 Asn Phe Tyr Gln Met Asp Gly Thr Pro Cys Glu Gly Gly Gln Ala
 425 430 435
 Tyr Cys Tyr Asn Gly Met Cys Leu Thr Tyr Gln Glu Gln Cys Gln
 440 445 450
 Gln Leu Trp Gly Pro Gly Ala Arg Pro Ala Pro Asp Leu Cys Phe
 455 460 465
 Glu Lys Val Asn Val Ala Gly Asp Thr Phe Gly Asn Cys Gly Lys
 470 475 480
 Asp Met Asn Gly Glu His Arg Lys Cys Asn Met Arg Asp Ala Lys
 485 490 495
 Cys Gly Lys Ile Gln Cys Gln Ser Ser Glu Ala Arg Pro Leu Glu
 500 505 510
 Ser Asn Ala Val Pro Ile Asp Thr Thr Ile Ile Met Asn Gly Arg

WO 01/98468

PCT/US01/19178

515
 Gln Ile Gln Cys Arg Gly Thr His Val Tyr Arg Gly Pro Glu Glu
 530
 Glu Gly Asp Met Leu Asp Pro Gly Leu Val Met Thr Gly Thr Lys
 545
 Cys Gly Tyr Asn His Ile Cys Phe Glu Gly Gln Cys Arg Asn Thr
 560
 Ser Phe Phe Glu Thr Glu Gly Cys Gly Lys Lys Cys Asn Gly His
 575
 Gly Val Cys Asn Asn Asn Gln Asn Cys His Cys Leu Pro Gly Trp
 590
 Ala Pro Pro Phe Cys Asn Thr Pro Gly His Gly Gly Ser Ile Asp
 605
 Ser Gly Pro Met Pro Pro Glu Ser Val Gly Pro Val Val Ala Gly
 620
 Val Leu Val Ala Ile Leu Val Leu Ala Val Leu Met Leu Met Tyr
 635
 Tyr Cys Cys Arg Gln Asn Asn Lys Leu Gly Gln Leu Lys Pro Ser
 650
 Ala Leu Pro Ser Lys Leu Arg Gln Gln Phe Ser Cys Pro Phe Arg
 665
 Val Ser Gln Asn Ser Gly Thr Gly His Ala Asn Pro Thr Phe Lys
 680
 Leu Gln Thr Pro Gln Gly Lys Arg Lys Val Ile Asn Thr Pro Glu
 695
 Ile Leu Arg Lys Pro Ser Gln Pro Pro Pro Arg Pro Pro Pro Asp
 710
 Tyr Leu Arg Gly Gly Ser Pro Pro Ala Pro Leu Pro Ala His Leu
 725
 Ser Arg Ala Ala Arg Asn Ser Pro Gly Pro Gly Ser Gln Ile Glu
 740
 Arg Thr Glu Ser Ser Arg Arg Pro Pro Pro Ser Arg Pro Ile Pro
 755
 Pro Ala Pro Asn Cys Ile Val Ser Gln Asp Phe Ser Arg Pro Arg
 770
 Pro Pro Gln Lys Ala Leu Pro Ala Asn Pro Val Pro Gly Arg Arg
 785
 Ser Leu Pro Arg Pro Gly Gly Ala Ser Pro Leu Arg Pro Pro Gly
 800
 Ala Gly Pro Gln Gln Ser Arg Pro Leu Ala Ala Leu Ala Pro Lys
 815
 Val Ser Pro Arg Glu Ala Leu Lys Val Lys Ala Gly Thr Arg Gly
 830
 Leu Gln Gly Gly Arg Cys Arg Val Glu Lys Thr Lys Gln Phe Met
 845
 Leu Leu Val Val Trp Thr Glu Leu Pro Glu Gln Lys Pro Arg Ala
 860
 Lys His Ser Cys Phe Leu Val Pro Ala
 875

<210> 17
 <211> 850
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7473301CP1

<400> 17
 Met Asp Lys Glu Asn Ser Asp Val Ser Ala Ala Pro Ala Asp Leu
 1 5 10 15
 Lys Ile Ser Asn Ile Ser Val Gln Val Val Ser Ala Gln Lys Lys
 20 25 30
 Leu Pro Val Arg Arg Pro Pro Leu Pro Gly Arg Arg Leu Pro Leu
 35 40 45
 Pro Gly Arg Arg Pro Pro Gln Arg Pro Ile Gly Lys Ala Lys Pro

WO 01/98468

PCT/US01/19178

Lys	Lys	Gln	Ser	Lys	Lys	Lys	Val	Pro	Phe	Trp	Asn	Val	Gln	Asn
65														
Lys	Ile	Ile	Leu	Phe	Thr	Val	Phe	Leu	Phe	Ile	Leu	Ala	Val	Ile
80														
Ala	Trp	Thr	Leu	Leu	Trp	Leu	Tyr	Ile	Ser	Lys	Thr	Glu	Ser	Lys
95														
Asp	Ala	Phe	Tyr	Phe	Ala	Gly	Met	Phe	Arg	Ile	Thr	Asn	Ile	Glu
110														
Phe	Leu	Pro	Glu	Tyr	Arg	Gln	Lys	Glu	Ser	Arg	Glu	Phe	Leu	Ser
125														
Val	Ser	Arg	Thr	Val	Gln	Gln	Val	Ile	Asn	Leu	Val	Tyr	Thr	Thr
140														
Ser	Ala	Phe	Ser	Lys	Phe	Tyr	Glu	Gln	Ser	Val	Val	Ala	Asp	Val
155														
Ser	Ser	Asn	Asn	Lys	Gly	Gly	Leu	Leu	Val	His	Phe	Trp	Ile	Val
170														
Phe	Val	Met	Pro	Arg	Ala	Lys	Gly	His	Ile	Phe	Cys	Glu	Asp	Cys
185														
Val	Ala	Ala	Ile	Leu	Lys	Asp	Ser	Ile	Gln	Thr	Ser	Ile	Ile	Asn
200														
Arg	Thr	Ser	Val	Gly	Ser	Leu	Gln	Gly	Leu	Ala	Val	Asp	Met	Asp
215														
Ser	Val	Val	Leu	Asn	Gly	Asp	Cys	Trp	Ser	Phe	Leu	Lys	Lys	Lys
230														
Lys	Arg	Lys	Glu	Asn	Gly	Ala	Val	Ser	Thr	Asp	Lys	Gly	Cys	Ser
245														
Gln	Tyr	Phe	Tyr	Ala	Glu	His	Leu	Ser	Leu	His	Tyr	Pro	Leu	Glu
260														
Ile	Ser	Ala	Ala	Ser	Gly	Arg	Leu	Met	Cys	His	Phe	Lys	Leu	Val
275														
Ala	Ile	Val	Gly	Tyr	Leu	Ile	Arg	Leu	Ser	Ile	Lys	Ser	Ile	Gln
290														
Ile	Glu	Ala	Asp	Asn	Cys	Val	Thr	Asp	Ser	Leu	Thr	Ile	Tyr	Asp
305														
Ser	Leu	Leu	Pro	Ile	Arg	Ser	Ser	Ile	Leu	Tyr	Arg	Ile	Cys	Glu
320														
Pro	Thr	Arg	Thr	Leu	Met	Ser	Phe	Val	Ser	Thr	Asn	Asn	Leu	Met
335														
Leu	Val	Thr	Phe	Lys	Ser	Pro	His	Ile	Arg	Arg	Leu	Ser	Gly	Ile
350														
Arg	Ala	Tyr	Phe	Glu	Val	Ile	Pro	Glu	Gln	Lys	Cys	Glu	Asn	Thr
365														
Val	Leu	Val	Lys	Asp	Ile	Thr	Gly	Phe	Glu	Gly	Lys	Ile	Ser	Ser
380														
Pro	Tyr	Tyr	Pro	Ser	Tyr	Tyr	Pro	Pro	Lys	Cys	Lys	Cys	Thr	Trp
395														
Lys	Phe	Gln	Thr	Ser	Leu	Ser	Thr	Leu	Gly	Ile	Ala	Leu	Lys	Phe
410														
Tyr	Asn	Tyr	Ser	Ile	Thr	Lys	Lys	Ser	Met	Lys	Gly	Cys	Glu	His
425														
Gly	Trp	Trp	Glu	Ile	Tyr	Glu	His	Met	Tyr	Cys	Gly	Ser	Tyr	Met
440														
Asp	His	Gln	Thr	Ile	Phe	Arg	Val	Pro	Ser	Pro	Leu	Val	His	Ile
455														
Gln	Leu	Gln	Cys	Ser	Ser	Arg	Leu	Ser	Gly	Lys	Pro	Leu	Leu	Ala
470														
Glu	Tyr	Gly	Ser	Tyr	Asn	Ile	Ser	Gln	Pro	Cys	Pro	Val	Gly	Ser
485														
Phe	Arg	Cys	Ser	Ser	Gly	Leu	Cys	Val	Pro	Gln	Ala	Gln	Arg	Gly
500														
Asp	Gly	Val	Asn	Asp	Cys	Phe	Asp	Glu	Ser	Asp	Glu	Leu	Phe	Cys
515														
Val	Ser	Pro	Gln	Pro	Ala	Cys	Asn	Thr	Ser	Ser	Phe	Arg	Gln	His
530														
Gly	Pro	Leu	Ile	Cys	Asp	Gly	Phe	Arg	Asp	Cys	Glu	Asn	Gly	Arg
545														

WO 01/98468

PCT/US01/19178

```

Asp Glu Gln Asn Cys Thr Gln Ser Ile Pro Cys Asn Asn Arg Thr
560 565 570
Phe Lys Cys Gly Asn Asp Ile Cys Phe Arg Lys Gln Asn Ala Lys
575 580 585
Cys Asp Gly Thr Val Asp Cys Pro Asp Gly Ser Asp Glu Glu Gly
590 595 600
Cys Thr Cys Ser Arg Ser Ser Ser Ala Leu His Arg Ile Ile Gly
605 610 615
Gly Thr Asp Thr Leu Glu Gly Gly Trp Pro Trp Gln Val Ser Leu
620 625 630
His Phe Val Gly Ser Ala Tyr Cys Gly Ala Ser Val Ile Ser Arg
635 640 645
Glu Trp Leu Leu Ser Ala Ala His Cys Phe His Gly Asn Arg Leu
650 655 660
Ser Asp Pro Thr Pro Trp Thr Ala His Leu Gly Met Tyr Val Gln
665 670 675
Gly Asn Ala Lys Phe Val Ser Pro Val Arg Arg Ile Val Val His
680 685 690
Glu Tyr Tyr Asn Ser Gln Thr Phe Asp Tyr Asp Ile Ala Leu Leu
695 700 705
Gln Leu Ser Ile Ala Trp Pro Glu Thr Leu Lys Gln Leu Ile Gln
710 715 720
Pro Ile Cys Ile Pro Pro Thr Gly Gln Arg Val Arg Ser Gly Glu
725 730 735
Lys Cys Trp Val Thr Gly Trp Gly Arg Arg His Glu Ala Asp Asn
740 745 750
Lys Gly Ser Leu Val Leu Gln Gln Ala Glu Val Glu Leu Ile Asp
755 760 765
Gln Thr Leu Cys Val Ser Thr Tyr Gly Ile Ile Thr Ser Arg Met
770 775 780
Leu Cys Ala Gly Ile Met Ser Gly Lys Arg Asp Ala Cys Lys Gly
785 790 795
Asp Ser Gly Gly Pro Leu Ser Cys Arg Arg Lys Ser Asp Gly Lys
800 805 810
Trp Ile Leu Thr Gly Ile Val Ser Trp His Gly Cys Gly Arg
815 820 825
Pro Asn Phe Pro Gly Val Tyr Thr Arg Val Ser Asn Phe Val Pro
830 835 840
Trp Ile His Lys Tyr Val Pro Ser Leu Leu
845 850

```

<210> 18

<211> 254

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473308CD1

<400> 18

```

Met Gln Asp His Arg Lys Gly Lys Ala Ala Val Gly Val Ser Phe
1 5 10 15
Asp Asp Asp Asp Lys Ile Val Gly Gly Tyr Asn Cys Glu Glu Asn
20 25 30
Ser Val Pro Tyr Gln Val Ser Leu Asn Ser Gly Tyr His Phe Cys
35 40 45
Val Gly Ser Leu Asn Arg Glu Tyr Cys Ile Gln Val Arg Leu Gly
50 55 60
Glu His Asn Ile Glu Val Leu Glu Gly Asn Glu Gln Phe Ile Tyr
65 70 75
Ala Val Lys Ile Ile Arg His Pro Lys Tyr Asn Ser Trp Thr Leu
80 85 90
Asp Asn Asp Ile Leu Leu Ile Lys Leu Ser Thr Pro Ala Ile Ile
95 100 105
Asn Ala His Val Ser Thr Ile Ser Leu Pro Thr Thr Pro Pro Ala
110 115 120

```

WO 01/98468

PCT/US01/19178

Ala Gly Thr Glu Cys Leu Ile Ser Gly Trp Gly Asn Thr Leu Ser
 125 130 135
 Ser Gly Ala Asp Tyr Pro Asp Glu Leu Gln Cys Leu Asp Ala Pro
 140 145 150
 Val Leu Ser Gln Ala Glu Tyr Glu Ala Ser Tyr Pro Gly Lys Ile
 155 160 165
 Thr Asn Asn Val Phe Cys Val Gly Phe Leu Glu Gly Gly Lys Asp
 170 175 180
 Ser Cys Gln Ile Ile Pro Ile Lys Val Gln Gln Leu Val Thr Ser
 185 190 195
 Ser Gln Glu Thr Asp Ile Arg Ile Pro Met Ala Leu Gln Thr Ala
 200 205 210
 Ala Ser Thr Ser Tyr Leu Gly Pro Leu Asp Ser Leu His Arg Lys
 215 220 225
 Val Ser His Pro Thr Glu Lys Arg Cys Gln Gln Lys Gln Gly Met
 230 235 240
 Lys Ile Thr Asp Asn His Gly Ile Thr Ser Lys Trp Ser Val
 245 250

<210> 19
 <211> 568
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7478021CD1

<400> 19
 Met Leu Ala Ala Ser Ile Phe Arg Pro Thr Leu Leu Leu Cys Trp
 1 5 10 15
 Leu Ala Ala Pro Trp Pro Thr Gln Pro Glu Ser Leu Phe His Ser
 20 25 30
 Arg Asp Arg Ser Asp Leu Glu Pro Ser Pro Leu Arg Gln Ala Lys
 35 40 45
 Pro Ile Ala Asp Leu His Ala Ala Gln Arg Phe Leu Ser Arg Tyr
 50 55 60
 Gly Trp Ser Gly Val Trp Ala Ala Trp Gly Pro Ser Pro Glu Gly
 65 70 75
 Pro Pro Glu Thr Pro Lys Gly Ala Ala Leu Ala Glu Ala Val Arg
 80 85 90
 Arg Phe Gln Arg Ala Asn Ala Leu Pro Ala Ser Gly Glu Leu Asp
 95 100 105
 Ala Ala Thr Leu Ala Ala Met Asn Arg Pro Arg Cys Gly Val Pro
 110 115 120
 Asp Met Arg Pro Pro Pro Pro Ser Ala Pro Pro Ser Pro Pro Gly
 125 130 135
 Pro Pro Pro Arg Ala Arg Ser Arg Arg Ser Pro Arg Ala Pro Leu
 140 145 150
 Ser Leu Ser Arg Arg Gly Trp Gln Pro Arg Gly Tyr Pro Asp Gly
 155 160 165
 Gly Ala Ala Gln Ala Phe Ser Lys Arg Thr Leu Ser Trp Arg Leu
 170 175 180
 Leu Gly Glu Ala Leu Ser Ser Gln Leu Ser Val Ala Asp Gln Arg
 185 190 195
 Arg Ile Glu Ala Leu Ala Phe Arg Met Trp Ser Glu Val Thr Pro
 200 205 210
 Leu Asp Phe Arg Glu Asp Leu Ala Ala Pro Gly Ala Ala Val Asp
 215 220 225
 Ile Lys Leu Gly Phe Gly Arg Arg His Leu Gly Cys Pro Arg Ala
 230 235 240
 Phe Asp Gly Ser Gly Gln Glu Phe Ala His Ala Trp Arg Leu Gly
 245 250 255
 Asp Ile His Phe Asp Asp Asp Glu His Phe Thr Pro Pro Thr Ser
 260 265 270
 Asp Thr Gly Ile Ser Leu Leu Lys Val Ala Val His Glu Ile Gly
 275 280 285

WO 01/98468

PCT/US01/19178

His Val Leu Gly Leu Pro His Thr Tyr Arg Thr Gly Ser Ile Met
 290 295 300
 Gln Pro Asn Tyr Ile Pro Gln Glu Pro Ala Phe Glu Leu Asp Trp
 305 310 315
 Ser Asp Arg Lys Ala Ile Gln Lys Leu Tyr Gly Ser Cys Glu Gly
 320 325 330
 Ser Phe Asp Thr Ala Phe Asp Trp Ile Arg Lys Glu Arg Asn Gln
 335 340 345
 Tyr Gly Glu Val Met Val Arg Phe Ser Thr Tyr Phe Phe Arg Asn
 350 355 360
 Ser Trp Tyr Trp Leu Tyr Glu Asn Arg Asn Asn Arg Thr Arg Tyr
 365 370 375
 Gly Asp Pro Ile Gln Ile Leu Thr Gly Trp Pro Gly Ile Pro Thr
 380 385 390
 His Asn Ile Asp Ala Phe Val His Ile Trp Thr Trp Lys Arg Asp
 395 400 405
 Glu Arg Tyr Phe Phe Gln Gly Asn Gln Tyr Trp Arg Tyr Asp Ser
 410 415 420
 Asp Lys Asp Gln Ala Leu Thr Glu Asp Glu Gln Gly Lys Ser Tyr
 425 430 435
 Pro Lys Leu Ile Ser Glu Gly Phe Pro Gly Ile Pro Ser Pro Leu
 440 445 450
 Asp Thr Ala Phe Tyr Asp Arg Arg Gln Lys Leu Ile Tyr Phe Phe
 455 460 465
 Lys Glu Ser Leu Val Phe Ala Phe Asp Val Asn Arg Asn Arg Val
 470 475 480
 Leu Asn Ser Tyr Pro Lys Arg Ile Thr Glu Val Phe Pro Ala Val
 485 490 495
 Ile Pro Gln Asn His Pro Phe Arg Asn Ile Asp Ser Ala Tyr Tyr
 500 505 510
 Ser Tyr Ala Tyr Asn Ser Ile Phe Phe Phe Lys Gly Asn Ala Tyr
 515 520 525
 Trp Lys Val Val Asn Asp Lys Asp Lys Gln Gln Asn Ser Trp Leu
 530 535 540
 Pro Ala Asn Gly Leu Phe Pro Lys Lys Phe Ile Ser Glu Lys Trp
 545 550 555
 Phe Asp Val Cys Asp Val His Ile Ser Thr Leu Asn Met
 560 565

<210> 20
 <211> 306
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4333459CD1

<400> 20
 Met Ser Leu Lys Met Leu Ile Ser Arg Asn Lys Leu Ile Leu Leu
 1 5 10 15
 Leu Gly Ile Val Phe Phe Glu Arg Gly Lys Ser Ala Thr Leu Ser
 20 25 30
 Leu Pro Lys Ala Pro Ser Cys Gly Gln Ser Leu Val Lys Val Gln
 35 40 45
 Pro Trp Asn Tyr Phe Asn Ile Phe Ser Arg Ile Leu Gly Gly Ser
 50 55 60
 Gln Val Glu Lys Gly Ser Tyr Pro Trp Gln Val Ser Leu Lys Gln
 65 70 75
 Arg Gln Lys His Ile Cys Gly Gly Ser Ile Val Ser Pro Gln Trp
 80 85 90
 Val Ile Thr Ala Ala His Cys Ile Ala Asn Arg Asn Ile Val Ser
 95 100 105
 Thr Leu Asn Val Thr Ala Gly Glu Tyr Asp Leu Ser Gln Thr Asp
 110 115 120
 Pro Gly Glu Gln Thr Leu Thr Ile Glu Thr Val Ile Ile His Pro
 125 130 135

WO 01/98468

PCT/US01/19178

His Phe Ser Thr Lys Lys Pro Met Asp Tyr Asp Ile Ala Leu Leu
 140 145 150
 Lys Met Ala Gly Ala Phe Gln Phe Gly His Phe Val Gly Pro Ile
 155 160 165
 Cys Leu Pro Glu Leu Arg Glu Gln Phe Glu Ala Gly Phe Ile Cys
 170 175 180
 Thr Thr Ala Gly Trp Gly Arg Leu Thr Glu Gly Gly Val Leu Ser
 185 190 195
 Gln Val Leu Gln Glu Val Asn Leu Pro Ile Leu Thr Trp Glu Glu
 200 205 210
 Cys Val Ala Ala Leu Leu Thr Leu Lys Arg Pro Ile Ser Gly Lys
 215 220 225
 Thr Phe Leu Cys Thr Gly Phe Pro Asp Gly Gly Arg Asp Ala Cys
 230 235 240
 Gln Gly Asp Ser Gly Gly Ser Leu Met Cys Arg Asn Lys Lys Gly
 245 250 255
 Ala Trp Thr Leu Ala Gly Val Thr Ser Trp Gly Leu Gly Cys Gly
 260 265 270
 Arg Gly Trp Arg Asn Asn Val Arg Lys Ser Asp Gln Gly Ser Pro
 275 280 285
 Gly Ile Phe Thr Asp Ile Ser Lys Val Leu Ser Trp Ile His Glu
 290 295 300
 His Ile Gln Thr Gly Asn
 305

<210> 21

<211> 953

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6817347CD1

<400> 21

Met Thr Leu Leu Ala Pro Trp Tyr Thr Gly Pro Met Ile Pro Met
 1 5 10 15
 Asp Val Asn Glu Pro Ser Ser Val Thr Thr Ala Pro Thr Leu Ser
 20 25 30
 Ser Ser Leu Gln His Ile Ser Ser Phe Leu Ala Thr Gly Lys Lys
 35 40 45
 Leu Ser Leu His Phe Gly His Pro Arg Glu Cys Glu Val Thr Arg
 50 55 60
 Ile Asp Asp Lys Asn Arg Arg Gly Leu Glu Asp Ser Glu Pro Gly
 65 70 75
 Ala Lys Leu Phe Asn Asn Asp Gly Val Cys Cys Cys Leu Gln Lys
 80 85 90
 Arg Gly Pro Val Asn Ile Thr Ser Val Cys Val Ser Pro Arg Thr
 95 100 105
 Leu Gln Ile Ser Val Phe Val Leu Ser Glu Lys Tyr Glu Gly Ile
 110 115 120
 Val Lys Phe Glu Ser Asp Glu Leu Pro Phe Gly Val Ile Gly Ser
 125 130 135
 Asn Ile Gly Asp Ala His Phe Gln Glu Phe Arg Ala Gly Ile Ser
 140 145 150
 Trp Lys Pro Val Val Asp Pro Asp Asp Pro Ile Pro Gln Phe Pro
 155 160 165
 Asp Cys Cys Ser Ser Ser Ser Arg Ile Pro Ser Val Ser Val
 170 175 180
 Leu Val Ala Val Pro Leu Val Ala Gly His Lys Gly Gln Ala Phe
 185 190 195
 Ile Glu Arg Met Leu Gly Cys Phe Lys Glu Leu Lys Gln Glu Leu
 200 205 210
 Thr Gln Glu Gly Pro Gly Gly Gly His Pro Arg Ser Ala Trp Pro
 215 220 225
 Pro Arg Arg His Ala Gln Trp Pro Pro Glu Pro Cys Glu Gln Gly
 230 235 240

WO 01/98468

PCT/US01/19178

Glu Glu Pro Pro Pro Val Glu Ala Glu Glu Val Glu Glu Ala Glu
 245 250 255
 Thr Ala Glu Lys Ala Glu Arg Lys Val Glu Ala Glu Ala Lys Val
 260 265 270
 Glu Gly Lys Ala Glu Ala Ala Gly Lys Ala Glu Ala Ala Gly Lys
 275 280 285
 Val Asp Ala Thr Glu Lys Val Glu Thr Ala Gly Lys Val Asp Ala
 290 295 300
 Ala Gly Lys Val Glu Thr Ala Glu Gly Pro Gly Arg Arg Ala Glu
 305 310 315
 Leu Lys Leu Glu Pro Glu Pro Glu Pro Val Arg Glu Ala Glu Gln
 320 325 330
 Glu Pro Lys Gln Glu Leu Glu Asp Glu Asn Pro Ala Arg Ser Gly
 335 340 345
 Gly Gly Gly Asn Ser Asp Glu Val Pro Pro Pro Thr Leu Pro Ser
 350 355 360
 Asp Pro Pro Arg Pro Pro Asp Pro Ser Pro Arg Arg Ser Arg Ala
 365 370 375
 Pro Arg Arg Arg Pro Arg Pro Arg Pro Gln Thr Arg Leu Arg Thr
 380 385 390
 Pro Pro Gln Pro Arg Pro Arg Pro Pro Pro Arg Pro Arg Pro Arg
 395 400 405
 Arg Gly Pro Gly Gly Gly Cys Leu Asp Val Asp Phe Ala Val Gly
 410 415 420
 Pro Pro Gly Cys Ser His Val Asn Ser Phe Lys Val Gly Glu Asn
 425 430 435
 Trp Arg Gln Glu Leu Arg Val Ile Tyr Gln Cys Phe Val Trp Cys
 440 445 450
 Gly Thr Pro Glu Thr Arg Lys Ser Lys Ala Lys Ser Cys Ile Cys
 455 460 465
 His Val Cys Gly Thr His Leu Asn Arg Leu His Ser Cys Leu Ser
 470 475 480
 Cys Val Phe Phe Gly Cys Phe Thr Glu Lys His Ile His Glu His
 485 490 495
 Ala Glu Thr Lys Gln His Asn Leu Ala Val Asp Leu Tyr Tyr Gly
 500 505 510
 Gly Ile Tyr Cys Phe Met Cys Lys Asp Tyr Val Tyr Asp Lys Asp
 515 520 525
 Ile Glu Gln Ile Ala Lys Glu Glu Gln Gly Glu Ala Leu Lys Leu
 530 535 540
 Gln Ala Ser Thr Ser Thr Glu Val Ser His Gln Gln Cys Ser Val
 545 550 555
 Pro Gly Leu Gly Glu Lys Phe Pro Thr Trp Glu Thr Thr Lys Pro
 560 565 570
 Glu Leu Glu Leu Leu Gly His Asn Pro Arg Arg Arg Arg Ile Thr
 575 580 585
 Ser Ser Phe Thr Ile Gly Leu Arg Gly Leu Ile Asn Leu Gly Asn
 590 595 600
 Thr Cys Phe Met Asn Cys Ile Val Gln Ala Leu Thr His Thr Pro
 605 610 615
 Ile Leu Arg Asp Phe Phe Leu Ser Asp Arg His Arg Cys Glu Met
 620 625 630
 Pro Ser Pro Glu Leu Cys Leu Val Cys Glu Met Ser Ser Leu Phe
 635 640 645
 Arg Glu Leu Tyr Ser Gly Asn Pro Ser Pro His Val Pro Tyr Lys
 650 655 660
 Leu Leu His Leu Val Trp Ile His Ala Arg His Leu Ala Gly Tyr
 665 670 675
 Arg Gln Gln Asp Ala His Glu Phe Leu Ile Ala Ala Leu Asp Val
 680 685 690
 Leu His Arg His Cys Lys Gly Asp Asp Val Gly Lys Ala Ala Asn
 695 700 705
 Asn Pro Asn His Cys Asn Cys Ile Ile Asp Gln Ile Phe Thr Gly
 710 715 720
 Gly Leu Gln Ser Asp Val Thr Cys Gln Ala Cys His Gly Val Ser
 725 730 735
 Thr Thr Ile Asp Pro Cys Trp Asp Ile Ser Leu Asp Leu Pro Gly

WO 01/98468

PCT/US01/19178

```

740
Ser Cys Thr Ser Phe Trp Pro Met Ser Pro Gly Arg Glu Ser Ser
755
Val Asn Gly Glu Ser His Ile Pro Gly Ile Thr Thr Leu Thr Asp
770
Cys Leu Arg Arg Phe Thr Arg Pro Glu His Leu Gly Ser Ser Ala
785
Lys Ile Lys Cys Gly Ser Cys Gln Ser Tyr Gln Glu Ser Thr Lys
800
Gln Leu Thr Met Asn Lys Leu Pro Val Val Ala Cys Phe His Phe
815
Lys Arg Phe Glu His Ser Ala Lys Gln Arg Arg Lys Ile Thr Thr
830
Tyr Ile Ser Phe Pro Leu Glu Leu Asp Met Thr Pro Phe Met Ala
845
Ser Ser Lys Glu Ser Arg Met Asn Gly Gln Leu Gln Leu Pro Thr
860
Asn Ser Gly Asn Asn Glu Asn Lys Tyr Ser Leu Phe Ala Val Val
875
Asn His Gln Gly Thr Leu Glu Ser Gly His Tyr Thr Ser Phe Ile
890
Arg His His Lys Asp Gln Trp Phe Lys Cys Asp Asp Ala Val Ile
905
Thr Lys Ala Ser Ile Lys Asp Val Leu Asp Ser Glu Gly Tyr Leu
920
Leu Phe Tyr His Lys Gln Val Leu Glu His Glu Ser Glu Lys Val
935
Lys Glu Met Asn Thr Gln Ala Tyr
950

```

<210> 22

<211> 2204

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 275791CB1

<400> 22

```

atatgccaat agacotgaca agtctgaatt ggaaactcag attgacagaa tgacgaagaa 60
gagcttttagc agctgtcttg gagataagta agagagatgc ttccacalc tcyagtcacg 120
aagatgatga taagccaact agcagcccag ataccggatt tgcagaagat gatattcaag 180
aaatgccgga aaatccagac actatggaaa ctgagaagc caaaacaac acagagctgg 240
atccctccag ttttactgag ataactsaag actgtgatga gaataaagaa aacaaactc 300
cagaagatc tcaggggagaa gttgattggc tccagcagta tgatatggag cgtgaaagg 360
aagaagcaaga gcttcagcag gcaactggctc agagccttca agagcaagag gcttgggaac 420
agaaagaaga tgatgacctc naaagagcta ccgagtttaag tcttcaagag tttacaact 480
cctttgtgga tgcattgggt tctgatgagg actctggaaa tgaggatgtt tttgatagg 540
agtacacaga agctgaagct gagggaactga aaagaaatgc tgagacagga aatctgctc 600
atctgtaccg gctatcaggt gttgtcagtc acattgttag caattctct tcaagtcat 660
acattagtga tgtatatgac attaagaagc aagcgtggtt tacttaaat gacctggagg 720
tatcaaaat ccaaggagct gccgtgcaga gtgctcagaa tggagtggtc tacatctct 780
tttatatgca caaggagatc tttgatgagc tgcctggaaac agaaagaac tctcagtcac 840
ttagcagcga agtggggaag actaccgctc aggcctcgtg aggaacaaac tctgggttg 900
gcagcatgca ctgcatattt gttactgctg ccacactcac ctctcctcty ctgaaggaga 960
atttggaatt ctacttgatg cgggagcaac aaacagctca gggccaaacc aaaagacaaa 1020
aatggagta acgtagaatg ctccatgcta tttatggaa actttggtct cacatccgta 1080
gttgattac ctcttttct cctatgagtg gcaattctt tgccttagga ataatgttg 1140
taaatatata ctgtgtgatg tgtgtataca cacacacaga cacacacaga cacacaggg 1200
atgaatggag ccttaaaagag ttaggatgag ccaccagaat atgcctgctc aaaatlaata 1260
gcacagcagt ttgagagaaga aatgaagtg tcacagagtc cattcacctg agaaatgtgt 1320
gaagacatac ttatcagttg gotttttagct tttatgttcc ttgagtagtt caactaagt 1380
ctgtaacctt ttgtgtttcc ttattagtaa aattcaactg aaagccagct cttcatgta 1440
cactaatgac agtttgttct ctttgcaaga gaggggcatt actgtcacct gacttgagga 1500
cgtgttttgt tgttgttgtt gctgcacaa ttcatgaatt tgtgatgtct ttgctgtta 1560
catgcagttc caagaaatgg attgttgggt ctttggaata tgttacagtc ccacattga 1620

```


WO 01/98468 PCT/US01/19178

```

tatttttat atactttgtt ttctctaagg agattttctt acacagtat ttcatacat 1680
atcatcatca ttatttatgtt ggttaagata gaattctttt tttttttttg tcatctggc 1740
catggagcag cattacccta atggatlgca accaaaactt taacaagta gaaagataat 1800
attttccaa ttgggacttc ccaggagaa tacttaggga taaggagaa tcttagcalt 1860
tctgtctctc aaacataggg aggtatagaa gagtgttctt ctggttaagc taaaattctg 1920
gaacctgaa gctaaaagcc ctattgcaag tatgaatta agtacttgag ctataggaca 1980
aaccttgggc atttaaccat ttactgtctg gcttttgcct taaatagggt ttgcaattaa 2040
aatgtgattg gcttaggtta tccaaaaaac taacaaataa caaaggtgca taatttattt 2100
atctactttt taggtgtctt gagttgagc aaagtagagc ggcaacatta agtgctatgc 2160
tagtcaacta gctgaagtaa ccagctttgg taagcagctt atga 2204

<210> 23
<211> 2036
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1389845CB1

<400> 23
ccgatggggg ttaggctcca gggcttctgt cgagaccaag gatgcccaa tatctgggtg 60
gtgggtgctg catacctggg ccttgggcag aacgaagggt atacagctg ggcaccagg 120
ataagtccag aaccaccag gagctgagga cagacagaag gaccacggag ggggtgacgg 180
gctgggtgtga ggattgtgtc ccttgggcca ggactctctc ctcttctccc tgcgtggctcc 240
agaccagagt ccaagcccta ggcagtgcga cccctaccga gccagcctt gaagacagaa 300
tgagaggggt ttctgtcttc caggtctcgc tcttctgtgt gctgggagct gctgggactc 360
aggtaaagaa gtctgcagcc ttggggcagc ccgcaggtc cagtggatc gttgggggac 420
gggtggcccg ggacggagag tggccgtggc aggcagcat ccagcatctg ggggcaacg 480
tgtgcggggg gtgcctcatc gcccccaggt ggtgtgctgac agcggcgcaac gcttctccca 540
ggagggcact gccagctgag tacgcgtgct gcttgggggc gctgcgtctg ggtccacct 600
cgccccgcac gctctcggtg ccgtgctgac ggtgtgctgt gcccccggac tactccgagg 660
acggggcccg cggcgacctg gcactgctgc agctgcgtcg cccggtgccc ctgagcgctc 720
ggctcccaacc cgtctgcctg ccgtgcccgc gcgcccgcgc gccgcccgc acaccatgcc 780
gggtccacgg cbggggcagc ctccgcagc gactgcctc ccagagtg ggcacgtac 840
aaggagtaag ggtgcgctg ctggactcgc gcaactgcga cggcctctac cagtgggcg 900
cggaactgac ccaggtgtgag cgcattgtgc tgcctgggag tctgtgtgac ggtaccccc 960
agggccacaa ggaagcctgc caggtgtgca ccagcctccg gactccctc gactccctc 1020
cctgtgccca gcaacccccc tccctgaact ccagaccca ggacatccca actcaggctc 1080
aggaatcctg cctccaaact agaggcacca cgcaggggt ctggnaacct gagaactgaa 1140
gtcctgggag ggtcgggact taggtctctc ttctctctgc aggggtgattc tgggggacct 1200
ctgacctgcc tgcagctctg gactgggtgc ctgggtggcg tgggtgagct gggcaagggt 1260
tgtgcctgcg ccaacccctc aggggtctac accagtgtg ccacatagc cccctggatt 1320
caggctcgcg ctacgttcta atgtagccg gtgaggtgca cctggagcca gctgctggg 1380
tccctcagcc tcttggttca tcaaggcacc tgcctatacc ccacatccct tctgctcga 1440
ggccaagatg ctaaaaaag ctaaaagcca cccaccccc caccacccc ctctgctc 1500
ctctctcttt tgggatcac cagctctgac tccaccaacc ctcatccagg aatctgccat 1560
gagtccagg gactcacact cccactctcc ttcttgctt gtatttactt ttcttgccc 1620
tggccagggc tgggcgcaag gcaagcagtg atgggcaaac caattgctgc ccatctggcc 1680
tgtgtgccca tctttctctg gagaagtca gattcacagc atgacagaga ttgacacca 1740
ggagatctct ccatagctg cttagagga accgggacca cagcaatgag cggctctaa 1800
gagctgagag acagccggca ggaatcgga accctcagac ccacagccc aaggcactg 1860
attctggcag caacctgaag gagctgggaa tgaagtctt cccagcctc cagataagag 1920
cccgcggc caatcccttc atttcaact aaagagacc taagcagaga acctagctga 1980
gcaactcctg acctacaaag ttgtgactta ataatgtgt gctttaagct gctcca 2036

<210> 24
<211> 2185
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1726609CB1

<400> 24
gcaatgcttc ctgcccacgg ccaccagcaa gctgtgggc gcagtggagc agtggctgag 60

```

WO 01/98468

PCT/US01/19178

```

tgcagctgag cggtctgtat ggcctacat gtggggcagg tacgacatty tcttctgcc 120
acctctcttc cccatctgtg ccatggagaa cccctgcctc accttcata tctctccat 180
cctggagagc gatgagttcc tggatcatga tgtcatccac gagggtggcc acagttggtt 240
cggcaacgct gtacacaacg ccacgtggga agagatgtgg ctgagcgagg gcttggccac 300
ctatggccag cgcgtatca ccccgagac ctacgtgtgc gcttctacct gctggagac 360
tgcttccgc ctggacgccc tggacccgca gatgaagctt ctgggagagg acagcccggt 420
cagcaaacctg caggtcaagc tggagccagg agtgaatccc agccacctga tgaacctgtt 480
cacctacgag aagggtact gcttctgtga ctacctgtcc cagctctgcg gagaccaca 540
gcgctttgat gactttctcc gagcctatgt ggagaagtac aagttcacca gcgtggtggc 600
ccaggacctg ctggactect tctgagctt ctcccgagg ctgaaggagc agagcgtgga 660
ctgcggggca gggctggaat tcgagcctg gctcaatgcc acaggcccg cgtggctga 720
gcccagacctg tctcagggat ccagctgac ccggccctgt gaggctctt tcagctgtg 780
gaccgcagaa cctctggac aggcagctgc ctggcgagc gccattgaca tctcaagtg 840
gaggaccttc cagacagcac tcttctgtga ccggtctctg gatgggtccc cgtgcgcga 900
ggaggtgttg atgagcctgt ccaagtgtga ctctccctg ctggactga tgaacctga 960
gatccgcatc cgtggtctgc agattgtgtt ccgaacgac tactatctg acctccacg 1020
ggctcggcgc ttcctggaga gccagatgtc acgcatgtac accatccgc tgtacagga 1080
cctctgcacc ggtgcctca agtctctgc gctggaggtc tctaccaga cgcggggcg 1140
gctgcacccc aacctggcca gagccatcca gcagatcctg tcccaggcc tgggtccag 1200
cacagagccc gctcagagc ccagcacgga gctgggcag gctgaagcag acacagactc 1260
ggacgcacag cccctgtctc ttgggagca ggcgccagc agtgcctct ctctcaggga 1320
cgtaaatgtg tctgctagc cctgttggcg ggtgaacct cgaactcca gacaccaca 1380
tttgtccctc tgtggccag gcttgcacat actgctctc gctctggcc atgagctctg 1440
cccagggcca caagccctc cctgggctc tcccaggcag ggagaatggg gagagggacc 1500
tcttgtgttc tggcagagac ctgtggact ggctcccca ctccagctc tcttgcactg 1560
cagccctctg ggcagcccg ccacacacct gctctctgtc tcaacactga cagctgtgct 1620
tagcccgga tgcagcacc tgcaggtgc cgcgccggg caaggggccc agcagcccta 1680
tgggtaccgc caccctgtgc cttaagtgc gcgggggccc caggtgtgct tctctctga 1740
gcacgctctc ttgcagggat ctgagcaacc ctcccgcac agcctgcac cccgcccta 1800
gggttggcag cctcagttgg cccctggcag aggaacaagg acacagacat tccctcagt 1860
tggggggcag ggaacacagg gagaggtgg ttgtccctg ggagggcct ctggcccag 1920
gcaaccttag cccctcagaa caggagttcc caggccccag ggagagtgtg gggacaggac 1980
agcctgtctc ttgtagcttc ctgggttggg aggcacaggg gcaaaagcaat accccaggga 2040
aagtgggagg tgggtgtgtt gctctctcca ggccacacat gctgggagag gcggccagag 2100
ctgggggctt ccagcctggg actgctgtga tggggtatca cgggtgatgtt cccattaac 2160
ttccctctg caaaaaaaaa aaaaaa

```

<210> 25
 <211> 3486
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4503846CB1

```

<400> 25
ctgtcttaaa aaaagaggga gggaagatta ctgatttaatt ttataaagga gaattattat 60
agctccaaca cctgacttta ttatctata tggtttaant acacaaacaa ttcagtyttt 120
gaattatata aatttcatta aaactatgta attatgcag aaaaatagga aatcacgggg 180
cacttagttt tgtgcatatg tgttaccctg agagattatg ctgtttttt taaaagggtt 240
cttttaatt taattattaa ttttataatg cacattcata tgttgacttt ggaccaacag 300
aaatccctaa tctctattct ttttctgatt ctttttagag ttggtgtggt caggatttta 360
ctcagaatga cgttaggaag agaagtgtat tctctcttc aggcaatgtc tctctact 420
gtggctggca gaaatgtttt aagatgggat ctttcaccag agcaaatata aacaagaact 480
gaggagctca tltgacagac caaacagggt tacgatgctg ttggaatgct cgttatagag 540
gaagtaactt acgagaactg tctgcaggca ctggcagatg tagaagtaaa gtatatagt 600
gaaaggacca tgcagactt ccccagcat gtatcctctg acaagaagat acgagcagca 660
agtaacagaa gagaacaag actttctctg ttgtatattg agatgagcat gagaggagat 720
atatttgaga gaattgttca ttacaggaaa acctgtgac tggggaagat aaaacctgag 780
gccagacgat acttggaaaa gtcaattaaa atgggggaaa gaaatggct ccatcttct 840
gaacagctac agaattgaaat caaatcaatg aagaaagaa tgaatgagct atgtattgat 900
tttaacaaaa acctcaatga ggaatgatac ttctctgtat ttccaaggc tgaacttggt 960
gcttctctg atgatttcat tgacagttta gaaaagacag atgatgacaa gtataaaatt 1020
accttaaaat atccacacta tttccctgtc atgaagaaat gttgtatccc tgaaccaga 1080
agaaggtatg aatggcttt taatacaagg tgcagaagg aaaaacccat aattttgcag 1140
cagctactcc cactgcgaa caaggtggcc aaactactg gttatagcac acatgctgac 1200

```

WO 01/98468

PCT/US01/19178

```

ttcgtcccttg aaatgaacac tgcaagagc acaagccgcg taacagccctt tctagatgat 1260
ttaagccaga agttaaaacc ctggggtgaa gcagacgcag agtttatctt gaatttgaag 1320
aaaaaggaat gcaaaagacag ggyttttgaa tatgatggga aaatcaatgc ctgggatcta 1380
tattactaaca tgactcagac agaggaaatc aaglatctca tagaccaga gtctctcaag 1440
gaataacttcc caattgaggt ggtcactgaa ggccttgctga aacctacca ggagttgttg 1500
ggactttcat ttgaacaaat gacagatgct catgtttgga acaagagtgt tacactttat 1560
actgtgaagg ataaagctac aggaagaagta ttgggacagt tctatttgga cctctatcca 1620
agggaaggaa aatacaatca tgcggcctgc ttgggtctcc agcctggctg cctctgtcct 1680
gatggaaagg ggaatgatgc agtggctgoc ctggtgtgga actctccaca gccagtgga 1740
ggtcgtccct ctctcctgag acacgacgag gtgaggactt actttcatga gtttggctac 1800
gtgatgcac agattttgtgc acagactgat ttgacagat tttagcgaac aaatgtggaa 1860
actgactcttg tagaggtgac atcgcaaatg cttyaaaatt ggggtgtggg cgtcgatcc 1920
ctcgaaagat tgcataaaga ctataaagat ggaagcccta ttgcagcaga tctggttgaa 1980
aaactgtgttg cttctaggtc ggtcaacaca ggtctcttga cctgcgcca gatgttttg 2040
agcaaaagtg atcagctctc tcataccaac acatcgctgg atgctgcaag tgaatatgcc 2100
aaatactgct cagaaatatt aggagttgca gctactccag gcacaatat gccagctacc 2160
tttggacatt tggcaggggg atacgatggc caataattat gatattcttg gagtgaagta 2220
ttttccatgg atatgtttta cagctgtttt aaaaaagaag ggataatgaa tccggaggtt 2280
ggaaatgaaat acagaaacct aatcctgaaa cctgggggat ctctggacgg catggacatg 2340
ctccacaatt ccttgaaagg tgagcacaac caaaaagct tctaatgg tagagagctg 2400
catgctccgt gaactggggg tctttggtag ccgtcaatgt ctggaggaca agtcgacatc 2460
accaatgttt actggcctgg aaactgaagg gagttttga agtgaaatt tagatttcta 2520
ttgacatcct tttgttttct aattttaaaa attataaaga tgaataatga attataaata 2580
ctgtgacctc agaaaagacc cactagaagg taattgtact ataaaattc ataaaactgg 2640
atttgatttc tttttatgaa agtttcaatg gaatgtact tgatttttta ctattataat 2700
ctagataaata tgaataaaga gggctaaaga tttttaaat gaatcatata tatgatataa 2760
tttgatccct ctgtgatctt gaagttttgt acttgggact ttgggactga taatgaatc 2820
atccatctct tctgtgaaat atttcttggt agctctgtgt caacttggat cctttgtctc 2880
ccaggaaggt gtgacctctc ctttgccctg atacctcaag gccaggggaa tatgctcag 2940
tgatgcatct atctttgtat atcaggccgc atgatccca actttctgcc acacttaaat 3000
taagtctctc catttcagtt ttgtcttttc tgtctaaagt tcagtcaaag agtatcaaaa 3060
aattatgttt cagctagact ggtgtaatt ataatgttt gtatcttgta tttagagatt 3120
tcgtatgctt tattagagcc tcatttccac ctacgacatc aagatctgta gttctttggc 3180
atgtgtgccca attagaatac taaagcaagt ccaagcacat ttttctcttc tcagtttctc 3240
aataagtggt agggactttg cctcttttac taccaggtc cccaaaagg tcaggtagac 3300
atgtcacaaa tggctctgtg gagagccatg ggaagagaga ggaggtggat gtggaacata 3360
aagggttcag aaactccaga agaggagtgg gttttggata gaagcatttg aggcacagct 3420
ctccaaagcc ttatgtgtat gatgaacctt aaccacgggg aagagactct tcagttagct 3480
gttctg 3486

```

<210> 26
 <211> 2847
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5544089CB1

```

<400> 26
caatgacgct tggacgagga tttatttcta caagotaatt gaatccagga gcagctttaa 60
ttatatacac taacggaaga gaaaaggagt atttccaagg gctcaaatgg aagctgtact 120
cagtcgggtg gaggcagggg gaggtaaagt ttctcacact caagctgctt tcatagttaa 180
ctgtcctttt ccaaacaaaa gctaataacg ccatacgcac ccacacactc cctcctggat 240
gaacctaaat ctgctcccca ctgtccacco aaggccagtt atcaaaaaat gttcctcttc 300
tgccctcaaa gaactgaagcc gcaggccctg ttctgcctct gctcaggaaat ctgattgtct 360
ttaaagtgtc attacaagat tccgtcgatg ttgtctcctc ctgtcttgtc atcaggacta 420
agtggtggag catcaaaagg tagaaaagtg gaacttatcc agctaaaggc gccaatctca 480
cagtaaaatt ctctttgtca tgaattgaat actttgttcc aagtcaatgt gtcgtggaa 540
tggggtgttg tctcaccatt tgctatgcta cactcaatgt ggagactcat tctgtccttt 600
cgtgtgttacg cccaacaaga cgtccagga tttctttgtg aacttttaga taaaatacaa 660
cgtgaattag agcaactcgg taccaagtta ccagctctta tcccacttcc tcaaaagaaa 720
ctcatcaaac aagtcttgaa tgttgttaaat aacatttttc atggacaact tcttagtcag 780
gttacatgtc ttgcatgtga caacaaatca aataccatag aacctttctg ggacttgtca 840
ttggagtttc cagaaaagta tcaatgcagt ggaaaagata ttgcttccca gccatgtctg 900
gttactgaaa tgttggccaa atttacagaa acttgaagctt tagaaggaaa aatctactga 960
tgtgaccagt gtaactcaaa cgttagaagg ttctctcca aaccagttgt actcacagaa 1020

```

WO 01/98468

PCT/US01/19178

```

gccagaaac aacttatgat atgccaccta cctcaggttc tcagactgca cctcaaacga 1080
ttcaggtggt caggacgtta taaccgagag aagattggtg ttcattgttg ctttgaggaa 1140
atcttaacaa tggagcccta ttgctgcagg gagacctgaa atccctcag accagaatgc 1200
tttatctaat acttgtccgc ggtggtgatg caccatggga aaggatttgg ctccaggcac 1260
tacactgcct actgctataa ttctgaagga ggtttctggg tacactgcaa tgattccaaa 1320
caagcaatgt gcactatgga tgaagtatgc aaggctcaag cttatatctt gttttatacc 1380
caacgagtta ctgagaatgg acattotaaa cttttgcctc cagagctcct gttgggggag 1440
caacatccca atgaagacgc tgatacctcg tctaataaaa tctttagctg atccaaagac 1500
aatgggtttt tcttcctgtg atttataat atacttttta aaagactgat gtaccatttt 1560
aaaactcatt ttttctgttg aatcagtgta tactacattt atacatttta tatctaaaca 1620
tttttttttt tacaagatat aaalgatat atcaactgaa ggtaactact ttthctatat 1680
ttggagtttt aaacttttgg tgtttacctc agactgaagt taactctttt atatttttat 1740
gtcttaattg gctoggatga tgaacttgtg caatcttcta ccaacnaagt tcaagtggca 1800
tcattttata tacatgtatc tttttcaggt attttctata caaatcttta atagatggaa 1860
aattagactc tactttggtc actaatagtc tttcatttgt atattgaagt taacttgccc 1920
cttggagtta ttgaagtgc atgtcaaggt atcacctaaa tattcttcag tcacactcac 1980
tggattttct gaggctttgt gtgttaacag gctttgtaat tgacattatt ttggttaaty 2040
taaccccaaa attgctttag caatgtctct ttggcatagt caaactataa atgaaaatgg 2100
cagctttaca aatagatat ttaagtgaa cctgggaacta tggcatgaa aaaaatgatg 2160
gctgggattt atgatttttg tctggcagca aacaggtttg tcagagaagt taataattaa 2220
gcagtcataa aaagtctgaa tttagtaaac cagtgtatga tgttattcaa atagttttac 2280
ttgggtatga gtctatttta taatgtctga tgacattaga tctcttaaaa ctttatgtat 2340
tttttttagt tcaaggaat agagtcttga agagaaaaaa ttatagggca gaaaagataa 2400
tggttcaaaa ttggcaactg gactattatt atgtctagca tctcattcta aataactaaa 2460
gcttgattta ctcttgctag gattatgtga ctactaggtg ggagcctctt aaaacactgg 2520
ccctgagcat taacaaaaaa aaaaaaaact aaagctatc tacttaact tgcacaaaaa 2580
aaactccggt ggggtccac cttttccttc ttctgaaat ctacagggt ttctttaag 2640
ccctgttgct gcaaacctta tcttttttgg ggggggtaga atcaccat ctctgtagac 2700
cagctatgtt tctaagctct gttaaccacg gggagatctg gtacccttt tttaaaagg 2760
ggtttattgt cgggttgagg tcttagtgaa aagtagtccc ctggagaatg cgytccacc 2820
ctgggggcca tctgttaggt aaaaactt 2847

```

<210> 27

<211> 890

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7474081CB1

<400> 27

```

gagggcaaga attcggcagc aggcacttac tccctgagct aagggggaag agctggatca 60
ccatgaataa tgtctcttat ttgggtgtcc tgcgtgggac atttttcttt gctgactcat 120
ctgttcagaa agaagaccct gctccctatt tgggttacct caagtctcac ttcaaccct 180
gtgtggcgct cctcatcaaa ccagctggg tgcgtggccc agctcactgc tatttaccaa 240
atctgaaagt gatgctggga aatttcaaga gcagagtcag agacgggtact gaacagacaa 300
ttaaaccctat tcagatcgtc cgtactggga actacagtc tagcgcccca caggatgacc 360
tcagtctcat caagctggct aagcctgcca tgcctcaat caaagtcag cctcttaacc 420
tcgcccaccac caalgctcagg ccaggcactg tctgtctact ctcaggtttg gactggagcc 480
aagaaaaacag ttgcccagac cctgacttgc ggcagaacct ggagccccc ggtgatgtctg 540
atcgagaatg ccaaaaaaca gaacaaggaa aaagccacag gaattcctta tgtgtgaaat 600
ttgtgaaagt attcagccga atttttgggg aggtggccgt tgcctactgc atctgcaaac 660
acaagctcca gggantcag gtggggcact tcatgggagg ggaagtcgga atctacacca 720
atgtttacaa atagtatccc tggattgaga acactgctaa ggacaagtga gacccacttt 780
ctccctctgc attccactgg ctctgccatg gactatacaa gcagataatt tccctctat 840
tcaaaataaa atctccaaat gaaaatttgg gaattgagca aaaaaaaaaa 890

```

<210> 28

<211> 1577

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5281209CB1

WO 01/98468

PCT/US01/19178

```
<400> 28
atgcagccca cgggcgcgga ggggtccgcg gcgcacagcc ggccgtatct ggcgcgtctg 60
ctgctcctgc tactgctgct gctgctgcgg cagcccgtaa cccgcgggga gaccacgcgg 120
gggcgcgcga gacccctctc cagcctgggg ccccccagcc tcttcaccac gccgggtgtc 180
cccagggccc tcactacccc aggcctcact accgcaggca ccccacaaac cctggacctt 240
cgggggtcgcg cgcagccctt gatcggaggt tcccaactcg tggacggcca caatgacctg 300
ccccaggtcc tgagacagcg ttacaagaat gtgcttcagg atgttaacct gcgaaatttc 360
agccatggtc agaccagcct ggacagcctt agagacggcc tctgtgggtg ccagttcttg 420
tcagcctcgg tctcatgcca gtcccaggac cagactcgcc tgcgcctcgc cctggagcag 480
attgacctca ttacccgcat gtgtgcctcc tactctgaac tcgagcttgt gacctcagct 540
gaagggtctga acagctctca aaagctgggc tgcctcatgt gcgtggaggg tggctactca 600
ctggaaagca gctctctgt gctgcgcagt tctctatgtg tgggggtgcg ctactcgaca 660
cttaccttca cctgcagtac accatgggca gagagtcca ccaagttcag acaccacatg 720
tacaccaacg tcacgagatt gaacaagcttt ggtgagaaag tagtagagga gttgaaccgc 780
ctgggcagtg tgatagattt gtccatgca tcggacacct tgataagaag ggtcctggaa 840
gtgtctcagg ctctctgtat cttctccac tcagctgcca gagctgtgtg tgacaatttg 900
ttgaatgttc ccgatgatat cctgcagctt ctgaagaaga acggtggcat cgtgatgggt 960
acactgtcca tgggggtgct gcagtgcaac ctgcttgcta acgtgtccac tgtggcagat 1020
cactttgacc acatcagggc agtcatgtga tctyagltca tccggatttg tggaaattat 1080
gaagggaatg gccggtctcc tcagggtgtg gagatgtgt cccatatacc agtcttgata 1140
gaggagtgtc tgagtcttag ctggagcgag gaagagcttc aagtggtcct tctgtgaaac 1200
ctgctcgggg tcttcagaca agtggaanaag gtgagagag agagcagggc gcagagcccc 1260
ctggaggtgt agtttcacata tgggcaactg agcacatcct gccactccca cctcgtgctt 1320
cagaatggac accaggtcac tcaatctggag gtgaccaagc agccaaacca tccgggtccc 1380
tggaggtcct caaatgcctc cccatacctt gttccaggcc ttgtggctgc tgcaccatcc 1440
ccaatcttca cccagtggtc ctgctgacac agtcgggtcc ccagagggtc actgtggcaa 1500
agctcaaaaa cctccctcct cctagtctat tcacaagcat atgctgagaa taacatgttt 1560
acacatggaa aaaaaaaa 1577

<210> 29
<211> 1958
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2256251CB1

<400> 29
aagcggctga gctcggcatt cattgtaacg gcgcacatgt ctggaaaggt cgtgtgggtt 60
ctgctcgcac ctctcgggtg agtggggctg gtcggggtgt gctgcagggc tgtctcccc 120
aocaccactg tagtcaagtc gtacctgtgg agatgctctg aggcacatga acaacctgtc 180
cctctcgaaa ctctctcccc acagcgtccc caccgtggcc ctggaaecag ctgggggctt 240
tgccgtctgg gagagccggg gcccccagcc acaccgcctg cctatctata aacatctctt 300
tctgtctaca tcccagcttc ccttccattc agcccagtg gcacactcca tcaccagcac 360
aatctacag ctacggcaga cccaggtgtg gccagtgggt ctccagctga ctctctctta 420
atttctcttt aacttactga ggtgaagttt acagaagata aagttacatt atcaagcgtc 480
caatctagag gccccagaca ccttgagcag tgcgtctgc cccccgcgc ccgaatttta 540
tcagcttca aatatctcca ccccccctga aggaaaaag gggccacta gggagcacc 600
cctagcaccc cgggccttct cgggggctcc acggttctga gccctttct agccgcctag 660
gggccctctg cagcctttcc acgcctccg ggagcccttg ttagtttgt gagcatggt 720
cttagaaaag acgacctgag cccagcgcc gctcactgct cctgagacgt cattctgt 780
gcacccacga tgcctctggg gagagtctgg cagacgagag agctgaagag caaagtcccc 840
aagaaggcag ggaaggtgtg tcagggaagg ctctatggag gaagtgcagt gggctctctg 900
ggatccccac caggcacccc ttctctcttc gacttagggt gtcggccggc gcaggtttcg 960
gatcaggcgg gccggatcgt ggggggtcac gctgcccgg cccggcctg gccatggcag 1020
gcacagctcc gctcgggag ggtgcaagtg tgcggcggt cactgctcag ccccagtg 1080
gtgtctacag ctgcccactg ctctctcggg tccctgaact catccgacta ccagtgacac 1140
ctgggggaac tggagatcac tctgtctccc cactctcca cgtgaggga cateactgt 1200
cactccagcc cctcaggaca gccggggacc agcggggaca tgcctctgt ggagctcagt 1260
gtccccgtga cctctctcag ccggtatctg cccgtctgcc tcccggaggg ctcagatgac 1320
ttctgcctct ggaatccgtg ctgggtgacc ggctggggct atacggggga gggagagcct 1380
ctgcaccccc agtacagcct gcggggagtg aaagtctcc tggtagaac agagacctgc 1440
cgcggggact atcccggccc cgggggcagg atccctcag ccgaatgtg gtgtgccgg 1500
ggccccggg atgctctcca ggcagactcc gggggcctc tggctctgca ggtgaacgg 1560
gctcgggtgc agcctggcat tgtgagctgg ggtgagggct gccggccccc caacagccg 1620
ggagtctaca ctggtgtccc tgcctacgtg aactggatcc gcccccacat cacagcatca 1680
```

WO 01/98468

PCT/US01/19178

```

ggggggctcag agtctgggta cccacggctc cccctctctg ctggcttatt cctccccggc 1740
ctctctcttc tcttagtctc ctgtgtctct ctggccaagt gctgctgca cccatctgcg 1800
gatggtactc cctccccgco cctgtaactg tggcaggaat ccaagtgcac ttcttaata 1860
agttactatt tctccgctc cgccccctcc ctctctcttg agaagctgag tctctgcat 1920
cagattattg caacatttaa cctgaattta acgacacc 1958

```

```

<210> 30
<211> 3106
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7160544CB1

```

```

<400> 30
gctccgaggg caaggccgct gctactgccc ccgctgcttc ttagtgccgc gttccgccc 60
tgggttgtca ccggccgccc ccctgaggaa gccactgcaa ccaggacggg agtggagggc 120
gcgcagccatg aagccggcga ggcccgcctc atagcgacg ccgggacggg ccgggcccgg 180
ccggggggaa ggaataatga acatggcagc agcaatggaa ccggaacagg tgggtttgga 240
gatatttgaa actgcggact gtgaggagaa tattgaatca caggatcgcc ctaaatggga 300
gcctttttat gttgagcggg attcctggag tcagcttaaa aagctgcttg ccgataccag 360
aaaatatcat ggctacatga tggctaaggg acccatgat ttcattgttt tgaagaggaa 420
tgatccagat ggaacctcatt cagacagaat ctattaccct gccatgtctg gtgagaacag 480
agaaaaataca ctgttttatt ctgaaattcc caaaactatc aatagagcag cagtcttaat 540
gctctcttgg aagcctcttt tggatctttt tcaggcaaca ctggactatg gaatgtatcc 600
tcgagaagaa gaactatata gagaagaana acgcatgga ccagtggaa ttgctctta 660
cgatatacac caaggaaagt gaacatitct gtttcaagcc ggtagtggaa ttatacagt 720
aaaagatgga gggccacaag gatttacgca acaaccttta aggcccaatc tagtggaaac 780
tagttgtccc aacatacgga tggatccaaa attatgcctt gctgatccag actggattgc 840
tttatcacat agcaacgata ttggatatac taacatcgta accagagaag aaaggagact 900
cacttatgtg caaatgagc tagccaacat ggaagaagat gccagatcag ctggagtccg 960
tacctttgtt cccaagaag aatttgatag atattctggc tatttgggtg gtccaaaagc 1020
tgaacaactt ccagtggttg gtaaaattct tagaattcta tatgagaasa atgalgaatc 1080
tgaggtggaa attatctatg ttacatcccc taigtggaa acsaaggagg cagattcatt 1140
ccgttatcct aaaaacaggta cagcaaatcc taaagtcact ttttaagatg cagaataaat 1200
gattgatgct gaaggaagga tcatagatgt catagataag gaactaatc aaacttttga 1260
gattctattt gaaggagtgt aatatattgc cagagctgga tggactcctg agggaaaata 1320
tgcttggctc atctactag atcgctccca gactcgcta cagatagtgt tgatctcacc 1380
tgaattattt atcccagtag aagatgatgt tatggaaagg cagagactca ttgagtcatg 1440
gcctgatctg gtgacgcac taattatata tgaagaaaca acagacatct ggataaatat 1500
ccatgacatc ttctcagtgt ttccccaaag tcaagagag gaaatttgag ttaatttttg 1560
ctctgaatgc aaaaacagggt tccgtcattt atacaaaatt acatctattt taagggaag 1620
caaatataaa cgatccagtg gtggcgctgc tgcctcaagt gatttcaagt gtctatcaa 1680
agaggagata gcaattacca gtggtgaatg ggaagttctt ggccggcatg gatctaata 1740
ccaagtgtat gaagtcagaa ggctggtata ttttgaagcc accaaagact cccctttaga 1800
gcatcacctg tacgtagtca gttacgtaaa tccctggagag gtgacaagcc tgactgaccg 1860
tggctactca catctctgct gcatcagta gcaactgtgac ttctttataa glaagtatag 1920
taacaggaag atccacact gtgtgtcctt ttcaagcta tcaagtcttg aagatgacc 1980
aacctgcaaa acaaggaat ttggggccac cattttggat tcagcaggtc ctctctga 2040
ctatacctct ccagaaatct tctcttttga aagtactact ggattacat tgtatgggag 2100
gctctacaag cctcatgac taccgctgg aagaaatat cctactgtgc tgttcatata 2160
tgggtgtcct caggtgcagt tggtsaataa teggtttaa ggaagcaagt atttccgctt 2220
gaatacccta gccctctctag gttatgtggt tgtagtata gacaacaggg gatcctgtca 2280
ccgaggcgctt aaatttgaa ggccttttaa atataaaatg ggtcaaatag aaattgacga 2340
tcagggtgaa ggaactcaat atctagcttc tcatatgat tcatctgact tagatctgt 2400
gggcatccao gctcggtcct atggagata cctctccctg atggcatata tgaagaggtc 2460
agatatcttc aggttgcta tgcctggggc ccagtcact ctgtggatct tctatgatac 2520
agatacacg gaacgttata tgggtcacc tgaccagaat gaacaggct attactagg 2580
atctgtggcc atgcaagcag aaaagttccc cctgaacca aatcgtttac tgccttaca 2640
tgggttctct gatgagaatg tccattttgc acataccagt atattactga gttttttagt 2700
gaggcgtgga aagccatag atttacagat ctatcctcag gagagacaca gcataagagt 2760
tctgaaatcg ggagaacatt atgaaactgca tcttttgac taccttcaag aaaaccttgg 2820
atccagatatt gctgtctaa aagtgtatata attttgacct gtgtagaact cctggatac 2880
ccctggctat ttaaccasat gaggaggttt atccacaga aaacacagaa tctgatac 2940
cattttgata cctgcaatgt aacatctact cctgaaaata aatgtggtgc catgcagggg 3000
tctacgggtt gtggtagtta totaatacct taaccccaca tgcctcaaat caaatgatac 3060

```

WO 01/98468 PCT/US01/19178
 atatctcctga gagacccagc aataccataa gaattactaa aaaaaa 3106

<210> 31
 <211> 3567
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7477386CB1

<400> 31
 atggctccac tccgcgcgct gctgtctctac ctgctgctt tgcactgtgc gctctggccc 60
 gccgcgggca gccgcgaccc agagctgcac ctctctggaa agctcagtga ctatgtgtgt 120
 acagtgcctt gcagcacaga ctttcgggga cgtctctctt cccactggtt gtctggccca 180
 gcagcagcct ctgcaggagag catggttagtg gacacgccac cccactacc acgacactcc 240
 agtcacctcc aggtggctcg cagccctctg caccaggag ggacctgtg gctctggcag 300
 gtggggcgcc actccctcta cttcaatgtc actgttttcg ggaaggaaact gcaactgtgc 360
 ctggcgccca atcgaggtt ggtagtgcga ggtctctcag tggagtggaa ggagatctt 420
 cgggagctgt tccggcagcc ctacggcag gagtgtgtg acactggag tctactggaa 480
 atgcctgggg cagctgttgc catcagcaac tctgacggat tggcgggctt catccgcaca 540
 gacagcaccc acttcttcat ttagcctctg gacggggccc agcaggagaa ggaggccagc 600
 gggaggacac atgtgtgtta ccgcgggag gccgtccagc aggagtgggc agaacctgac 660
 ggggacctgc acaatgaagc ctttggcctg ggagaccttc ccaacctgct gggcctgggt 720
 ggggacacag tgggcgcacac agagcggaaag cggcggcatg ccaaggcagg cagctacagc 780
 atcgaggtgc tgcgtgtgtt ggaagactcg gtgttctgct tccatggcaa ggagcatgtg 840
 cagcaatatg tctctacctt catgaatat gtgttagagt agatttaca cgtatgagtc 900
 ctgggggttc atataaatat tgcctctgtc cgtttgata tggttggcta ccgacagcag 960
 tccctgagcc tgatcgagcg cgggaacccc tcacgcagcc tggagcaggt gtgtcgtgtg 1020
 gccactctcc agcagcgcca ggaacccagc cagctgagc accatgacca cgttgtgttc 1080
 ctcaaccggc aggaactttg gccctcaggt gggatatcac ccgtcactgg catgtgtcac 1140
 cccctgagga cgtgtgcctt caacctgag gatggcttct cctcagcctt cgttatagct 1200
 catgagacgg gccacgtgtc cggcatggag catgacggtc aggggaatgg ctgtgcagat 1260
 gtagccagcg tgggcagcgt catggcgccc ctgtgtcagg ctgtcttcca cgtcttccat 1320
 tgttcccgct gcagcaagct ggaagctcag cgttaacct cgtctacaga ctgctctctc 1380
 gatgacctct ttgatcctgc ctggcccccag cccccagagc tgcctgggat caactaotca 1440
 atggtatgagc agtgcgcctt tgactttggc agtggctaac agacctgctt ggcattcagg 1500
 acctttgagc cctgcaagca gctgtgtgtc agccatctct acnaccgta cttctgcaag 1560
 accaagaagg ggcctccgct ggaatgggact gagtgtgcac ccggcaagtg gtgctcaaa 1620
 ggtcactgca tctggaagtc ccggagcagc acatatggcc aggatggagg ctggagctcc 1680
 tggaccaggt ttgggtcatg ttccggttca tttggggggc gggatgcgac ccgcagcggg 1740
 agctgcacaa accctcgccc agcatatgta ggcgcctgt gcttagggac catgttcagg 1800
 taacaggtct gcaacagcga ggaagtccct gggacctacg aggacltccg ggcccagcag 1860
 tgtgccaaagc gcaactccta ctatgtgcac cagaatgcca agcacagctg ggtgccctac 1920
 gagcctgacg atgacgcccc gaagtgtgag ctgatctgcc agtggcgga caggggggac 1980
 gtgtgttcca tgaaccaggt ggttcaagat gggacacgct gcagctaccg ggaaccatac 2040
 agcgtctgtg ccgctggcga gtgtgtgctt gtcggctgtg acaaggaggt ggggtccatg 2100
 aagggcagtg acaagtgttg agtctgcggg ggtgacaact cccactgcag gactgtgaag 2160
 gggagcgttg gcaaggcttc caagcaggca ggtgtcttca agtgtgtgca gatccagaga 2220
 ggtgcagggc acatccagat tgaagcactg gagaagtccc cccaccgcat tgtgggtgaag 2280
 aaccaggcca ccggcagcctt catctccaac cccaagggca aggaaggcca aagccggacc 2340
 ttaccgcca tgggcctgga gtgggaggat gccgtggagg atgccaagga aagcctcaag 2400
 accagcgggc cctgtcctga agccattgcc atcctgtgct tcccccaac tggagggtgg 2460
 ccccgacgca gcttgcccta caagtaagtc atccatgagg acctgtgccc ccttatcggy 2520
 agcaacaaty tgcctctgga ggagatggac cccatagagt gggcgctcaa gactggggcc 2580
 ccccgacgca aggcctgttg aggaggtatc cagttcacca aalacygctg ccggcgagaa 2640
 cgagacacac acatggtgca gcgacacct tctgacacaa agaaagggc caagccatc 2700
 ccgcgcgctt gcaacagcga cccgtgctct cagcctgtgt gggtagcgga gggatgggt 2760
 ccctgcagcc ggaagctgtg gaagctgggg gtgcagacac gggggataca gtgctgctg 2820
 cccctctcca atggaaccca caagctcatg ccggcacaag cctgcgcggg ggaccgctc 2880
 gaggcccgac ggccctgtct ccgagtgccc tgcccagccc agtggaggot gggagcctgg 2940
 tccagtgctc ctgcacacct tggagagggc atccagcagc ggcaggtggt gtgcagagac 3000
 aagcccaaca gcctcgggca ttgcgagggg gataggccag acactgtcca ggtctgcagc 3060
 ctgcccgcct gtggagcggg gccctgcacg ggagacaggt ctgtcttctg ccagatggaa 3120
 gtgctcgatc gctactgctc catctccggt taccacggc tctgtgtgtg gtcctgctc 3180
 aagaaggcct cgggccccaa cccctggcca gacctggccc caacctcaet gcccccttc 3240
 tcaactctgt gaagccccct accaggaccc caggacctgt cagatgtgc agagcctct 3300

WO 01/98468 PCT/US01/19178

ggaagagccaa cgggatoaga ggaccatcag catggccgag ccacacagct cccaggagct 3360
 ctgggatacaa gctccccagg gacccagcat ccccttggcc ctgagacacc aatccctgga 3420
 gcatcctgga gcatcctccc taccaccccc ggggggctgc cttggggctg gaactcagaca 3480
 cctacggcag tccctgagga caaaggcaca cctggagaag acctgagaca tccgggacc 3540
 agctccctg ctgctcccc ggtgaca 3567

<210> 32
 <211> 2930
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7473089CB1

<400> 32
 cagcgagacc cggcgagcgg ccgagagccc gggccagccc cttccacag cggcgcttg 60
 cgtgcccggg cggcatgctt ctgctgggca tcttaacct ggctttcgcc gggcgaaacc 120
 ctggaggctc tggagccagag cgggaggtag tegtcccat cggactggac cgggacalta 180
 acggcgccgg ctactactgg cgggttccc aggactccg ggaacaggga ctcattttc 240
 agatccacag atttcaggag gacttttacc taacactgca cggcgatgct cagttctctg 300
 ctcccgcttt ctccactgag catctgggcy tccccteca ggggtcacc ggggctctt 360
 cagacctcgg acgtgcttc tattctgggg acgtgaacc cggagccgac tctgtctgct 420
 ctgtgagcct gtgcgggggg ctcccgagg cctttggcta ccgaggcgcc gagtatgtca 480
 ttggccgctt gcccaatgct agcgcccggg cggcgacgcy caacagccag ggcgcacacc 540
 ttctccagcg cgggggtgtt ctcggcgggc cttccggaga ccccactct cgtgcgggg 600
 tggccctggg ctgggacccc gccatcctac gggccctgga ccttacaag ccgggggggg 660
 cgggctctgg gggagctggt agcgcgcca ggtctgggcy cgcgaagct tctgttcta 720
 tcccgggcta cgtggagagc ctggtggteg cggacgagtc aatggtcaag ttccagcgc 780
 cggacctgga acattatctg ctgacgtgcy tggcaacggo ggcgcgactc taccgcatc 840
 ccagcatcct caaccccacc aacatcgtt tggtaaggt gctgcttct agagatcgtg 900
 actccgggcc caaggtcacc ggcaatcgg cctgacgct cgcgaactc tgtgctggc 960
 agaagaagct gaacaaagt agtgaacagg accccgagta ctgggacact gccatcctc 1020
 tcaccaggca ggaactgtgt ggaagccacca cctgtgacac cctgggcatg gctgatgtg 1080
 gtacattgtg tgaacccagg agaagctgct ctgtcattga ggaagatgg cttccatcag 1140
 ccttccaccac tggccacgag ctgggccaog tgttcaacat ccccatgac aatgtgaag 1200
 tctgtgagga ggtgtttggg aagctccgag ccaaccacat gatgtcccg accctcatc 1260
 agatgcagcc tggcaacccc tggtaagcct gcaagtgtgc catoatcac gaactcctg 1320
 acagcgggca cggtagctgc ctctggacc aacccagcaa gccatctcc ctgcccagg 1380
 atctgcccgg gcgcagctac accctgagcc agcagtgcy gctggctttt ggcgtgggt 1440
 ccaagccctg tctttacatg cagtactgca ccaagctgtg gtgcaccggg aaggccaaag 1500
 gacagatggt gtgccaagac cgcacttcc cttggggcga tggcaccagc tgtggcgagg 1560
 gcaagctctg cctcaaggg cctcgtgtgg agagacacaa cctcaacag caaagggtg 1620
 atggttcttg ggcacaatgg gatccatg gccctgtct cgcacacatg ggtggggcg 1680
 tgcagctggc caggaggcag tgcaccaacc ccacccctgc caacgggggc aagtagctg 1740
 agggagtgag ggtgaaatc cgtacctgca atctggagcc ctgcccagc tcaactcctg 1800
 gaaagagctt cgggagggag cagtgtgagg ctctcaacgg ctacaaccac agcaccacc 1860
 ggctactctt cgcctgggca tgggtgccca agtactccgg cgtgtctcc cgggcaagt 1920
 gcaagctcat ctgcccagcc aatggcactg gctacttcta tgtgctggca cccaaggtg 1980
 tgggtggagc cagctgtgcy tctctgact ccaactccgt ctgtgtcaca ggcaggtgca 2040
 tcaaggctgg ctgtgtggg aaactgggt ccaagaagag attcgacac tgtgggtgt 2100
 gtgggggaga caataagagc tgcagaaggg tgactggact cttaaccaag cccatgcatg 2160
 gctacaaatt cgtgtgtgccc atcccccgag ggcctcgaag catcgacac cgcgacgcy 2220
 gtacaaagg cctgatcggg gatgacaact acctggctct gaagaacagc caaggcaagt 2280
 acctgctcaa cgggcatttc gtgtgttcgg cgtgggagcy ggaactgtgt gtaagggca 2340
 gtctgctgcy gtacagcgcc cggggcacag cgtgggagag cctgagggct tccgggcca 2400
 tctctggagcc gctgacgty gaggctctct cgtggggaa gatgaaccc ccccggtcc 2460
 gctactcctt ctacttgccc aaagagccc gggaggacaa gtccctcat ccccggaac 2520
 cccggggagg accctctgac ttgcaaacac ggttccacg cctctcaac cagggtggagc 2580
 agcggagaca caggccccct gaacgtgtgg tggctggcag ctgggggccc tgcctcgca 2640
 cgtcccgag tggcctgag aagcggcgcy tggactggcg gggctccgc gggcagcgca 2700
 cggctccctg ctgtgatgca gccatcgcc cgtggagac acaagcctgc ggggagccct 2760
 gccccacctg ggaactcagc gctgggtcac cctgtccaa gagctgcgcy cggggtattc 2820
 agaggcgctc actcaagtgt gtgggccaag gaggccgctc gctggccgcy gaccagtgca 2880
 acttgacccy caagcccacg gactggact tctgctctc gaggcgctg 2930

<210> 33

WO 01/98468

PCT/US01/19178

<211> 4230

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7604035CB1

<400> 33

```

agcgagggttg cctggagaga ggcgctggggc gcagaagggt taacggggcca ccggggggctc 60
gcagagcagg aggggtgctct cggacagtggt gtcccccact gcaactctga acttggaggga 120
cagggttcgct gcgagggagc cagagagagac cctccacgco cagatgcctg cgttagttttt 180
gtgaccagtc cgtcctgccc tcccctgggg gcagtagagg gggagcgatg gagaactgga 240
ctggcaggcc ctggctgtat ctgctgtgct ttctgtccct cctcagctc tgcttggatc 300
aggagggttt gtccggacac tctcttcaga cactacaga ggaggggcag ggcccgaag 360
gtgtctgggg accttgggtc cagtgggcct ctgtctccca gccctgcggt gtgggggtgc 420
agcgaggagc cgggacatgt cagctcccta cagtgcagct ccacccgagt ctgcccctcc 480
ctccccggcc ccaagacat cagagagccc tctcccccg gggccagggt ccagagcccc 540
agactctctc agaaacccct cctctgtaca ggcacagtc tcggggaggt ggtggccac 600
ttcgaggtec cgtctccac ctaggggagag aggagacca gggalttga gcgccagga 660
ggtcocggct tcgagacccc atcaagccag gaatgttcgg ttatgggaga gtgcccctt 720
catttgcact gcacgggaac cgcagggacc ctgggagccc accagatct gagctgtccc 780
tgatctcttc tagaggggaa gaggctatct cgtccctac tccagagca gagccattct 840
cgcgaacggg cagcccccaa actgagctcc ctccacaga actgtctgtc cacaccccat 900
cccccaagc agaacctcta agccctgaaa ctgctcagac agaggtggcc ccagaaacca 960
ggctgtgccc cctacggcat caccccagag cccaggctc tggcacagag ccccctcac 1020
ccacgactc cttagagaa ggtgctctct tctgtcatc cctcagcca cgaaggccaa 1080
gttcccaagg tggggcagtt ccccaggtag caggggagcg ccttgatcct ttctcttcgg 1140
tccctggggg ccgagggcag caggggcgaag ggccttgggg aacggggggg actcctcac 1200
ggccccgctt ggagcctgac cctcagcacc cgggcgctg gctgcccctg ctgagcaacg 1260
gccccatgac cagctccctc tggagcctct ttgtctccag tagccctatt ccaagatgtt 1320
ctggggagag tgaacagcta agagcctgca gccaaagccc ctgccccctt gagcagccag 1380
acccccgggc cctgcagtc gcagccttta actcccagga attcatgggc cagctgtatc 1440
agtggagcc cttaactgaa gtccagggt ttgaaatgaa tgcgggccc tgcggggagg 1500
gtggcttcgg ctctatgtc cgtcacactg aaaaggtcca ggaatgggac ctgtgtcagc 1560
ctggagcccc tgaactctgt gtggctggac gctgtctgag ccccgctgt gatgggacc 1620
ttggctctgg caggctctct gatggctgtg gagtctgtgg ggggtgatg ttactcgtc 1680
gccttgtttc ggggaacctc actgaccgag ggggccccct gggctatcag aagatcttgt 1740
ggattccagc gggagccttg cggctccaga ttgcccagct cgggcctagc tccaaatacc 1800
tggcacttct tggcccttgg ggcgggtcca tcatcaatgg gaacttgggt gtggatcccc 1860
ctgggtctta caggccgggc gggacgctct ttgatabaa ccgtctctcc agggaggagg 1920
gcgaaggggg gagtctgtcg gctgaaggcc ccacaccca gcctgtggat gtctatatga 1980
tctttcagga ggaaaaccca ggcgtttttt atcagtatgt catctcttca cctcctccaa 2040
tctttagaaa cccaccccca ggcgccctgt tcccagctt tcagccggag attctgaggg 2100
tggagcccc acttgcctcg gcaccccgcc cagcccgagc ccaggcacc ctccagctc 2160
aggtgcggat ccccagatg cccgccccgc ccatccagc gacacccctg ggtctccag 2220
ctggctactg gaaacgagtg ggaacactct catgctcagc gtctgtcggg aaaggtgtct 2280
ggcgcccat ttctctctgc atctccgttg agtcgggaga ggaactggat gaaacgagct 2340
gtgcggcggg tgccaggccc ccagcctccc ctgaacctgt ccccggaacc ccatgcccct 2400
catactggga ggcctggcag tggacatcct gcagccgctc ctgtggccc ggcaaccagc 2460
accgacagct cagctgcogg caggaaattt ggggggttgg cctctcggtg ccccgggagc 2520
gctgtggaca tctccccgg ccaacatca ccagctcttg ccagctgcgc ctctgtggcc 2580
atttgggaagt tggctctcct tggagcagb gctcgtgtcg gtgcggccgg ggccagagaa 2640
gccggcaggt tgcgtgttt gggaaacaag gtgatgaagt gacgagcag gagtgtcgt 2700
caggccctccc acagccccc agcagagag cctgtgacat ggggcccctgt actactgcct 2760
ggttccacag cgaatgggag tccagtgct cagcggatgt tgggacgggt atccagcgc 2820
gctctgtgtt ctgcttggg agtggggcag ccactgggg caggccagg ggaagcagga 2880
ccaggaaact ggcagagctg tccaaacagga agccggcccc ctgacatgcg cgcctgcagg 2940
ctggggccct gtgagagaa tgggcgctgg tacaagggc cctgggggtg gtgctctccc 3000
gaatgtggct ctggcacaca gcttagagac atcatctgt tatccaaact ggggacggag 3060
ttcaactgta cttctccgag caactgttct caectccca ggcctccgct cctgcagccc 3120
tgtcaagggc aggcctgcca ggcagatgg ttttccagc cctggagccc atgttctcgc 3180
tctgtccaa ggggaacgca gacaaggag gtccagtgcc tgagcaacca ccagacccc 3240
agaccccgat gccctccca actgcggccc tccaggaaag gccctgtta cagcaacccc 3300
tgcagccagc gccctgatga tcaatgcaag gacagctctc cacattgcc cctgtgtgta 3360
acggcccggc tctgctcta cccctactac acagccaact gttgcgcctc ttgcgcacat 3420
gtctctggagc ggtctccca ggaatccctc tgaagggggt ccggggcacc ttacggttt 3480

```

WO 01/98468

PCT/US01/19178

```

tctgtgccac catcggtcac ccattgatcg gccactctg aacccctgg ctctccagcc 3540
tgtccagtc  toagcaggga  tgcctccag  gtgacagagg  gtggcaagg  gactgcacaa 3600
aagtgacttt  cagggctgtg  gtcaggccca  tgggtgggt  tgatgggtg  gtgcacatat 3660
gectcagtg  tgccttggg  actgaatgga  tatgtgtgt  ctcaaatgt  taicacttt 3720
caaaaggagg  ttaacagagc  tgagaaggac  aagactgtt  tccttgagac  ttctcagg 3780
ggaaaggaaa  gcaagttgc  agttccttgc  taactctgag  taactagagt  gtggtctccc 3840
caccactcc  agttttgtgc  cctaagctc  atttctcat  ttccagactc  acatctctta 3900
agccgcctg  tgtctctgac  cccttctcat  ttgcttagta  ttctgcccc  tgctcccta 3960
attagctagg  gctggggta  gccactgcca  atcctgctt  actcaggag  gcaggaggaa 4020
agagactgcc  tctccagagc  aaggccoaag  tgggcagagg  gtgaaaaaga  gaaatgtgag 4080
catcgctcc  cccaccacc  cyccagctc  ctgcccac  tccctgctc  ctgaaatgt 4140
tccccccag  aactaattta  tttttatcta  aagatggta  tgcacaaatg  aaaaaaaa 4200
aaaaaataa  aaaaacaaa  aaaaaaata

```

<210> 34

<211> 3699

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3473847CB1

<400> 34

```

cgagtggtgc  tggcaaaagt  tgactttccc  agcaggccca  tgtcataggt  actgtggtct  60
ctacaatata  gcagaggatc  ctgagggtcc  gagaggttga  gtgacttgct  catggctgca  120
caacaaagtaa  atattggagc  tgggaattcag  gtccacgtt  tcttggtccc  aaagccatg  180
attttttccc  tcaatttatt  ctgactgggg  catgggggag  ggggtggcct  ttgggcagg  240
ccaccaggag  cgaccaggcc  cgtagagagc  tgggtggaag  tacagaggaa  aacctgttgt  300
cgagtgtggc  cgttagttcc  catttttgcc  tgaatggcac  atttgaagt  gttatataac  360
catgtgaata  ataattagtt  gcttatatga  gttttttaa  ttgctttttg  gtccgcatt  420
ggtaacttct  ttatcateta  ctatactctg  ttgtgtctct  ttgtgtgtaa  ttgttaagta  480
ggggttagat  aaagtacacc  tagggtttgc  tgggtttctt  coattgcatc  atgttctccc  540
ttgcaatggg  ctatgatacc  ttgaggttgc  ctggacact  cgttggtgtg  ctgaaggagg  600
agaccacact  ctgcaagtc  gagcgaactg  ccgcgcctc  gcagcccaag  gctgcacccc  660
ggggtacact  caccagatc  ctgcatgtct  tccatggcct  tcttctggc  ttctgtgtga  720
agatgagtgg  cgacctgtg  gagctggcct  tgaagtgtcc  coattgcgac  tacactgagg  780
aggactctcc  tgtctttgcc  cagagcatcc  cgtggaacct  ggagcgga  accctccac  840
ggtacccggc  ggaatgaatc  cagccccccg  acggagggag  cctggtggag  gtgtatctcc  900
tagacaccag  catcacaggt  gaccaccggg  aaatcgagg  caggggtcatg  gtcaccgact  960
tcagaaagt  gcccgaggag  gacgggaccc  gcttccacag  accggccagc  aagttgaca  1020
gtcatggcac  ccactggca  ggggtgttca  gccggccgga  tgcggcgtg  gccaaagggt  1080
ccagcatgag  cagcctggc  gtgctcaact  gccaaaggaa  ggccacggtt  agcggcacc  1140
tcataggcct  ggaatttatt  cggaaaagcc  agctgggtca  gctgtgggg  ccaactgtgt  1200
tgcctgtgcc  cctggcgggt  gggatcacgc  cgtctctcaa  cgcgcctgc  cagcgcctgg  1260
cgagggtcgg  ggtcgtgctg  gtcaccgctg  ccggcaact  ccgggacgat  gcctgcctct  1320
actccccagc  ctcaactccc  gagggtcatca  cagltggggc  caccatgccc  caggaccagc  1380
cgggtgacct  ggggactttg  gggaccaact  ttggccgtg  tgggacctc  ttgtcccag  1440
gggaggacat  cattgtgtcc  tccagagct  gcagcaactg  ctgtgttca  caagttggga  1500
catcacaggc  tgcgccac  gtggctggca  ttgcagccat  gatgtgtct  gccagaccg  1560
agctcaccct  ggccgagttg  aggcagagac  tgatccact  ctctgccaaa  gatgtcatca  1620
atgaggcctg  gttccctgag  gaccagcggg  tactgacccc  naacctgggt  gccgcctgc  1680
ccccagcac  ccattgggca  ggttggcagc  tgttttgag  gactgtgtg  tcagcacact  1740
cggggcctac  ccggatggcc  acagccatcg  ccgcctgccc  ccagatgag  gagctgctga  1800
gctgtccag  ttctccagg  agtgggaagc  ggcggggcga  ggcgatggag  gcccaagggg  1860
gcagactggt  ctgcggggcc  cacaacgtt  ttgggggtga  ggggtctac  gccattgca  1920
ggtgtgctct  gctaccccag  gccactgca  ggttccaac  agctccaac  gctgaggcca  1980
gcatggggac  cgtgtccac  tgccaccac  agggccacgt  cctcacagg  tgcagctccc  2040
actggcaggt  ggaggaactt  ggcacccaca  agccgcctgt  gctgaggcca  cgaggtcaga  2100
ccaaccagt  cgtggggcac  agggagggca  gcattccacg  ttctgtctg  catgcccag  2160
gtctggaatg  caaagtcaag  gagcatggaa  tccggcccc  taaggagcag  gtgacctgg  2220
cctgcaggga  gggctggacc  ctgactggct  gcagtgcct  cctggggacc  tcccactcc  2280
tgggggccta  gcgcgtagac  aacactgtgt  tagtcaggag  cggggacgtc  agcactacag  2340
gcagaccag  cgaagaggcc  gtgacagcgt  ttgcaactgt  ctgcggagc  cggcaactgt  2400
cgcaggcctc  ccaggagctc  cagtgcagc  ccatccacag  gatgggtgtc  tgggggggt  2460
caagggtgg  ggtgagctt  taaaatggtt  ccgaactgtc  cctctctcag  cctccatgg  2520
cctggcacga  ggggatgggt  atgcttccgc  ctctccggg  ctgctggcct  ggcccttgag  2580

```

WO 01/98468

PCT/US01/19178

```

tggggcagcc  tecttgccgt  gaactcactc  actctgggtg  cctctcccc  aggtggaggt  2640
gccagggaagc  tccctccctc  actgtggggc  atttcacat  tcaaacaggt  cagctgtgc  2700
tcgggtgctg  ccagctgctc  ccaatgtgco  gatgtccgtg  ggcagaatga  cttttattga  2760
cctgtgttcc  cgtgccaggc  atcaaatcct  caggtctcca  ccaaggaggc  aggatcttc  2820
ccatggatag  gggagggggc  ggttaggggt  gcaggagcaa  acatcgttgg  ggggtgagtg  2880
tgaagggtgc  tgaatggcct  catctccagc  taactgtgga  gaagccctgt  ggggtcctc  2940
gattaatgga  ggccttagctt  totggatggc  atctagccag  aggtctggga  caggtgtgco  3000
cctgttggtc  acaggctgtg  ccttggttto  ctgagccacc  ttactctgc  tctatgccag  3060
cgtgtgctag  caacacccaa  aggtggcctg  cggggagcca  tcactagga  ctgactcggc  3120
agtgtcagtt  ggtgcattga  ctgtctcagc  caaccgctc  cactaccgg  cagggtacac  3180
attcgcaacc  ctacttcaca  gaggaagaaa  cctggaacca  gaggggggct  gctgtccaa  3240
ctccacacgc  aggaactgag  ccagaaacgc  agaltgggtc  ggccttgaag  ccagctctc  3300
cttacttcaa  ccggctggg  ctctcattt  ttacgggtaa  cagttaggct  ggggaaggga  3360
acacagacca  ggaagctcgg  tgaatgatgg  cagaacgatg  cctgcaggca  tggaaacttt  3420
tcgtttatca  ccagcctgtg  attcaactgg  ctggcggaga  tgcttctaag  gcattgctgg  3480
gggagagggc  caacaactgt  cctccttga  gcaccagccc  caccacagca  agcagacatt  3540
tatcttttgg  gtctgtcttc  tctgttgctt  ttttacagcc  aacttttcta  gacctgtttt  3600
gcttttgttaa  ctgaagata  ttattctggt  gttttgtagc  atttttatta  atattgtgac  3660
ttttttaaatt  aaaaacaaac  aacgtygttc  ctaaaaaaa  3699

```

<210> 35

<211> 2410

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3750004CB1

<400> 35

```

cctcagcagt  ggccccttcc  ctccacgggc  tggccgggag  ctcagtccca  cccctccgc  60
cgatgagcgc  atgagggacc  gaacggacct  gggccagaa  cctcctctct  tctgtgggc  120
cctcctgaac  tgtggttttg  gggctcagtg  tcagggtccg  ggcagagtga  cccgtgggt  180
gtcctggaac  cgtgtctcca  gctcctggcg  gctggcgtc  tctgtgcgca  gccggcgctg  240
cctccggctt  cctggggaa  aaccgtgctg  gggagaccc  catgagtaac  gcccttgcca  300
gttgcagagc  tggcccccag  gggctgtgct  ctccagagac  ctcaagtgtg  cctgttcaa  360
tggccgctct  gtctggggca  cccagaagac  ctaccagtgg  gtgcccctcc  atggggcgc  420
caaccagtgc  gacctcaact  gcctgcctga  ggggcaagcc  ttctaccaca  gcttggcgc  480
cgtcctggac  ggcacccct  gcagcccggg  tggccagggg  gtctgctgg  ctggcgcgt  540
ccttagcgcc  ggcctgtgat  ggttgttggt  ctgggtgccc  ctgaggacc  gctgtggcc  600
ctgcggaggg  gccaacgact  cgtgcctttt  cgtgcagcgc  gtgtttcgtg  acgcccgtg  660
cttctgtggg  tactggaaag  tgacctgat  ccccgagggc  gccagacaca  tccgcttga  720
acacagagcc  cgaacccacc  tgggtatcct  aggatcactg  atggggggcg  atggcgcta  780
cgtgcttaat  ggcactggg  tggtcagccc  accagggacc  tacgagcgg  ccggcacgca  840
tgtgtcttac  acccagaca  caggccccc  ggagacattg  caagcagccg  gcccacctc  900
ccatgacctg  ctctacaggg  tctctctgca  ggagcccaac  cctggcatcg  agtttgatt  960
ctggctcctt  cgggagcgct  acagcccctt  ccaggctcgt  gtgcaggccc  tgggttgccc  1020
cctgaggcag  cctcagcccc  ggggggttga  gctcagccc  cccgcagccc  ctgctgtcac  1080
ccttgacagc  accccaagc  tggcccccaga  cccctgcaca  cccctgcctg  acaccccg  1140
ccgcgcacac  cgaactactcc  actattggcg  cagtgacttt  gtgttccagg  cccgagtgct  1200
gggcacacac  caccagccc  aggaagaccg  ctatgaggtg  cgtatccagg  tctgttaca  1260
gaacccgtct  ccactggcg  caccgagta  cgtgtggcg  ccaggccact  gccctgcc  1320
gatgttgcca  ccccacggg  actacctgat  ggtgttcag  cgtttgtca  gccccagcg  1380
cacacaggac  cagctgtctg  tggccacgc  cggctacgac  cggcccttga  gccctggga  1440
ggacagccgc  atacgctga  ctgcccggcg  ctgtcctgct  tgagccctg  caggagcccc  1500
ggccacacac  agcaagaaag  atacatctga  ccagcctcaa  cgtcaacgta  ttccctct  1560
caccctggct  tccaggcagc  totgaaatac  gtcccacctg  tgcagctatg  tgactccctc  1620
ccacacagc  ttaagacacc  ctgtcatgca  gtcaaaagcc  ctgtcacag  ccggcaggca  1680
ctggtagga  ggcactaagg  agacbtgac  ttattatgg  cctctctct  tggctgcag  1740
gaagctcata  gctatttata  ctcaaaagt  ttaacgtgct  tttctttct  ttggcgcg  1800
tcacacttgc  ttggagacac  tgtcatgaac  gacatgaca  cctctgtgco  ctgggtaccc  1860
agaagatcat  ctgtttactt  cccagacact  gtgtgtctc  tgctctctg  tactcacaca  1920
acccctcatg  tgtgaaggcg  agagacactg  tcacaaacag  gcatgcccc  tagaagacat  1980
gcctaaccag  gcaactgaac  gtaccaactg  accaatttcc  ccttttcccc  tggctaccag  2040
gaaactcgga  gacaactctt  tcaagctcag  cattctgccc  tggatttcca  ccatcaaca  2100
cgtgcttggc  cctccttttt  tttttttctg  aggtagacct  tggctgtgca  cctaggtcga  2160
agtgcgggtg  tgaactcatg  gctcactgca  gctcaaatc  cctgggtcca  agcgtacctc  2220

```

WO 01/98468 PCT/US01/19178

```

ccactcagct ccacagcagt ggaactcacg tgtgatcaca tgcgggctaa ttttaatttt 2280
gttagagatgg gcttgtacgt gccaaatgtc tcaactatggc tcaacatctc tgcctgggtcc 2340
aagactgaat aggatgacat gatgggtgtac cccttatcct tattttcagct ttaaaaaatc 2400
taaaaaaaa 2410

<210> 36
<211> 549
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4904126CB1

<400> 36
gggaggagag aaaaaccatg gccgacaagg tctgaagga gaagagaaag cagtttatcc 60
gttcagtggg cgaaggtaca ataaatggct tactgggtga attattggag acaagggtgc 120
tgagccagga agagatagag atagtaaaat gtgaaaatgc tacagttaat gataaggccc 180
gagcttttgt tgaactctgt attcggaaag gggctccagc atgccaatgt tgcatacat 240
acatttgtga agaagacagt cactggcag ggaagctggg acctccagca ggtccacat 300
ctggaaatca ccttaactaca caagatctc aatagtacl tccctccag gtaacgtgt 360
ttttaaagaa agagcattct ttgaaccgtg gcttcccggt acattaatgt ttaggatga 420
accacagtta aaggggctat gaagaattcc catagagtga tcatacaatt ttcttttgt 480
aatctattct gctttgttag caactgtcaa aacagcttca ctatctatgt ctacattaaa 540
atttggat 549

<210> 37
<211> 2755
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71268415CB1

<400> 37
ttgctaggag ggtggagttc atccacttat gatataaatg tctcttttta ttttttgag 60
gcaacttttt gctccttctc acacagaaac ccattatact tcaagtggta accctcaaac 120
caccacaggg aaattggagg atcaactgctt ttaccacggc acggtgaggg agacagaaat 180
gtccagctgc acgtcagca cttgcgaggg aattagagga ctgattacgg tgagcagcaa 240
cctcagctac gtcatcgagc cctcctctga cagcaagggg caacacctta ttacagatc 300
tgaacatctc aagccgcccc cgggaaactg tgggttcgag cactccaagc ccaccacag 360
ggactgggtc cttcagttta cacaacagac caagaagcga cctgcagga tgaagaggga 420
agatttaaac tcaatgaagt atgtggagct ttacctcgtg gctgattatt tagagttaa 480
gaagaatcga cgagaccagg acgccaacca acacaagctc atagagatcg ccaactatgt 540
tgacaagtgt tacgatctc tgaacatccg gattgctctc gtgggcttgg aagtgtggac 600
ccacgggaac atgtgtgaag ttccagagaa tccattctc accctctggg cctttctcag 660
ttggaggcgc aagctgcttg cccagaagta ccatgacaac gcccaattaa tcacgggcat 720
gtccttcac ggcaccacca tggccttggc cccctcatg gccatgtgt cgtgtacca 780
gtctggagga gtcaacatgg accactccga gaatgccatt ggcgtggctg ccaccatggc 840
ccacagatg ggcacaaact ttggcatgac ccatgattct gcagattgt gctcggccag 900
tgcggctgat ggtgggtgca tcatggcagc tgccactggg cacccttcc ccaagtggt 960
caatggatgc aacaggaggg agctggacag gtatctgcag tcaggtgggt gaatgtgtct 1020
ctccacatg ccagacacca ggaatgttga tggaggccgg aggtgtggga acgggtatct 1080
ggaagatggg gaagagtgtg actgtggaga agaagaggaa tgaacaacc cctgctgcaa 1140
tgctctaat tgaacctga ggcggggggc ggaatgtgtc caggctctc cgtgccacca 1200
gtgtaagctg ttggctcctg ggaacctgtg ccgagagcag gccaggcagt gtgacctccc 1260
ggagtctctg acgggcaagt cctcccactg cctaccaac ttctaccaga tggatggtag 1320
cccctgtgag ggcggccagg cctactctga caacggcatg tgcctcaact accaggagca 1380
gtgccagcag ctgtggggac ccggagcccg acctgccctt gacctctgtc tcgagaaggt 1440
gaatgtggca ggaacacct ttggaaactg tggaaaggac atgaatgggt aacacaggaa 1500
gtgcaacatg agagatcgca agtgtgggaa gatccagtgt cagagctctg aggccgggac 1560
cctggagtc cacycggtgc ccaattgacac cactatcat atgaatggga ggcagatcca 1620
gtgcgggggc acccagctct accgaggtcc tgaggaggag ggtgacatgc tggaccagg 1680
gctgggtgat actggaacca agtgtggcta caacctatt tgcctttgag ggcagtgcag 1740
gaacacctcc tctttgaaa ctgaaggctg tgggaagaa tgcaatggcc atggggtctg 1800
taacaacac cagaactgcc actgcctgcc gggctggggc ccgccttct gcaacacacc 1860

```

WO 01/98468

PCT/US01/19178

```

gggccaacggg ggcagtatcg acagtggggc tatgccccct gagagtgtgg gtctgtgtgt 1920
agctggagtg ttggtggcca tcttggtgct ggcgttcctc atgctgatgt actactgctg 1980
cagacagaaac aacaaactag gccaaactcaa gccctcagct ctcccttcca agctgaggca 2040
acgttcagtg tgtcccttca ggggtttctca gaacagcggg actgttcagt ccaacccaac 2100
tttcaagtgt cagagcgccc agggcaagcg aaaggtggtc aaactccggg aaatccctgg 2160
gaagccctcc cagcctcctc ccgggccccc tccagattat ctgctgtgtg ggtcccccac 2220
tgacccactg ccagctccacc tgagcagggc tgcctaggaa cccccagggc ccgggtctca 2280
aatagagagg acggagtgct ccaggaggcc tccccaagc cggccaatcc ccccccaccc 2340
aaattgcctc gtttcccagg acttctccag gcctcggcgg ccccagaagg cactcccgcc 2400
aaacccagtg ccaggccgca ggagcctccc caggccagga ggtgcacccc cactgcggcc 2460
ccctggtgct gggccctcagc agtcccgccc tctggcagca cttgccccaa aggtgagctc 2520
acggggaagc ctcaaggtga aagctgggtc caggggggca ggtgagagt 2580
tgagaaaaaa angcaattca tgttctttgt ggtctggact gaacttccag aacaaaagcc 2640
aaggccaata catctcatgt tcttggtgct cgcttgactg tggagttttt gcttcagtgt 2700
aaaggtgatt cttagaatcc tgagctgtgg tggcttcagt cctgccccct cacct 2755

```

<210> 38

<211> 2553

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473301CB1

<400> 38

```

atgggcaaaag aaaacagcga tgtttcagcc gcacctgctg acctgaaatc atccaatc 60
tcagtccaaag tgggtcagtg ccaaaagaag ctgccaagtg gacgacaccc gttgcgaagg 120
agacagactac catbgccagg aagacgacca ccacaaagac caattggcaa agccaaaccc 180
aagaagcaat ccaagaaaaa agttcccttt tggaaatgtac aaaaataaat cattctctc 240
acagtatttt tattcatcct agcagtcata gctggaccac ttctgtggtc gtatatcagt 300
aagacagaaa gcaaaagatgc tttttacttt gctgggatgt ttgcctacac caacatcgag 360
tttcttcccg aataccgaca aaaggagttc aggggaattc ttccagtgtc acggactgtg 420
cagcaagtgca taaacctggt ttatcaacaa tctgctctct ccaattttta tggagcagtc 480
gttgtgtgagc algtcagcag caacacaaa ggggacctcc ttgtccact ttggatgtgt 540
tttgtcatgc ccgtgccaag aggcacacac ttctgtgaag actgtgtttg cgcacatctg 600
aaggactoca tccagacaag catcataaac cggacctctg tggggagctt gcagggactg 660
gctgtgaaga tggactctgt ggtactaaat ggtgattgtt ggtcattctc aaaaaaaag 720
aaaagaaggg aaaaatggtg tgtctccaca gacaaaggct gctctcagta cttctatgca 780
gagcatctgt ctctccacta ccgcctggag atttctgcag cctcagggag gctgatgtgt 840
cacttcaagc tgggtggccat agtgggctac ctgattctgc tctcaatcaa gtccatcaa 900
atcgaagccg caaactgtgt cactgaactc ctgacatttt acgactccct ttggccatc 960
cggagcagca tctgtcaag aatttgtgaa cccacaagaa catnaatgtc atttgtttc 1020
acaaataatc tcatgttgtt gacatttaag tctctcata tacggaggct ctcaagaaat 1080
cggcctatgt ttgaagtgaa aatttcaagc caataatacc cgactacta tctccaaaa 1200
atcactggct ttgaagtgaa cctggaaatt teagacttct ctatcaactc ttggcatagc actgaaatc 1260
tgcaagtgtc cctggaaatt teagacttct ctatcaactc ttggcatagc actgaaatc 1320
tataactatt caataaccaa gaagagtatg aaaggctgtg agcatggatg gttggaaatt 1380
tatgagcaca tgtactgtgt ctcctacatg gatcatcaga caatttttct agtgcacagc 1440
cctctggttc acatcagct ccagtgagct tcaaggcttt caggcaagcc acttttggca 1500
gaatatggca gttacaacat cagtcaaccc tgcctgtgtg gatcttttag atgtctctc 1560
ggtttatgtg tccctcagcc ccagcgtgtg gatggagtaa atgactgctt tgaatgaagt 1620
gatgaactgt tttcgtgtg ccctcaacct gcttgcataa ccagctcctt caggcagcat 1680
ggcctctcca tctgtgtgtg cttcagggac tgtgagaatg gccgggatga gcaaaactgc 1740
actcaagata ttccatgcaa caacagaact ttttaagtgt gcaatgalat ttgcttttag 1800
aaacaaaatg caaaatgtga tgggacagtg gattgtccag atggaagtga tgaagaaggc 1860
tgacacctgca gcaagagttc ctcggccctt caccgcatca tgggaggcac agacacctg 1920
gaggggggtt gggcgtggca ggtcagcctc caatttgttg gatctgcta ctgaggtgct 1980
tcagtcaact ccaggagtg gttcttttct gcagcccaat gttttcagtg aaacccctg 2040
tcagatccca caccatggac tgcaacacct gggatgtatg ttcaggggaa tgccaagtgt 2100
gtctcccggt tgcagaagaat tgtgtgtccc gagtactata acagtacagc ttttgattat 2160
gatattgctt tgcacagct cagtattgct tggcctgaga cctgaaaca gctcattcag 2220
ccaatgtgca ttctcccaac tggtcagaga gtctgcagtg gggagaagtg ctgggtaact 2280
ggctgggggc gaagacacga agcagataat aaagctccc tctgtctgca gcaaggagag 2340
gtagagctca tgcataaac gctctgtgtt tccacotacg ggatcatcac ttctcgatg 2400
ctctgtgagc gcataatgtc agucaagaga gatgcctgca aaggagatcc ggttggacct 2460
tlatctgttc gaagaaaaag tgatggaaaa tggattttga ctggatgtgt tagctgggga

```

WO 01/98468 PCT/US01/19178

catggatgtg gacgacaaaa ctttctgggt gtttacacaa ggggtgtcaaa ctttgttccc 2520
 tggattcata aatatgtccc tttctttttg taa 2553

<210> 39
 <211> 1041
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7473308CB1

<400> 39
 atgttcaagcg gcaacacagg aaaaacccat attatcaatg ctcaaaaacc tggccacctc 60
 aggccttagcc agttattcgt gacgagagag gtgtgtcatc tacatggcag tcatggcctg 120
 gatgggtctg gaactgtggc aagaatcctt ccaggaaaca gccgtctccc ctctctgctc 180
 tcagaaggca agtttccotta tcaactgtct gctctcagaa ggcaagtttc cttatcacct 240
 gtgaatcaca aaccacaga gtggccaaac atactgatgc aagaccatag gaaggggaaa 300
 gctgcagttg gtgtctcctt tgaatgatgat gacaagaltg ttggggggcta caactgtgag 360
 gagaattctg tcccclacaa ggtgtccctg aattctggct accacttctg tgttggctcc 420
 ctaacacagg aatactgcct ccagggtaga ctgggagagc acaacatcga agtcttagag 480
 gggaatgaac agttcatcta tgcggtcaag atcatccgcc accccaata caacagctgg 540
 actctggaca atgacatcct gctgatcaag ctctccacac ctgccatcat caatgcccat 600
 gtgtccacca tctctctgcc caccacccct ccagctgctg gcactgagtg cctcatctct 660
 ggctggggca acactctgag ttctggcgcc gactaccgag acgagctgca gtgcctggat 720
 gctcctgtgc tgaacacagg gcctcctaac ctggaagatg taccacaac 780
 gtgttttctg tgggtttcct tgaaggagag aaggattcct gccagattat tccatcaaaa 840
 gtgcagcagc tgggttacctc aagcacaag acagacataa ggttccctat ggcttgcag 900
 acagctgctt ccacctccta cctgggcccc tttagactct tacacaggaa agtgagtcac 960
 cccactgaga agcgttgcca gcagaaacag ggcattgaaa tcacagataa ccatggattt 1020
 acttccaagt ggtcagtata a 1041

<210> 40
 <211> 1707
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7478021CB1

<400> 40
 atgtctgccc cctccatctt ccgtccgaca ctgctgctct gctggctgga tgctccctgg 60
 cccaccacgc ccgagagtct ctccacacgc cgggaccgct cggacctgga gccgtcccca 120
 ctgcgccagg ccaagcccat tgcgacacct caccgtgctc agcggttctc gtccagatcc 180
 ggctggtcag ggggtgtggc gccctggggg ccagctcccg aggggcccgc ggaagccccc 240
 aaggcgccgc ccctggccga ggcggtgcgc aggttccagc gggcgaaacg gctgcgggcc 300
 agcggggagc tggacgoggc caccctagcg gccatgaacc ggccgcgctg cggggtcccc 360
 gacatggccc caccgcccc ctcgccccgc ccttcgcccc cgggcccgcg ccccagagcc 420
 cgtctcaggg cgtcccgccg ggcgcgcgtg tcttgtctcc ggcggggttg gcaagcccg 480
 gctaccccc agcgcgagcg tgcacaggcc tctccaaaga ggaagctgag ctggcggtg 540
 ctggcgaggg ccctgagcag ccaactgtcc gtggccgacc agcgcgcat agaggcgctg 600
 gcoctcagga tgtggagcga ggtgacgcgc ctggacttcc gcgagacctt ggcgcgcgcc 660
 ggggcgcggg tgcacatcaa gctggggttt gggagacggc acctgggctg tccgcgggcc 720
 ttcgatggga gcgggcagga gtttgccacac gcctggcgcc taggtgacat tcaactttgac 780
 gacgacgagc acttcacacc tcccaccagt gacacgggca tcagccttct caaggtggcc 840
 gtccatgaaa ttggccatgt cctggggttg cctcacacct acaggacggg atccataatg 900
 caacaaatt acattcccaa ggcgcgtgct ttgtagtgg actggtcaga caggaaagca 960
 attcaaaagc tgtatggttc ctgtgagggg taatttgata ctgcgttga ctggattcgc 1020
 aaagagagaa accaatatgg agaggtgatg gtgagattta gcacatattt ctcccgtaac 1080
 agotggtact ggcctttatga aatccgaac aataggacac gctatgggga cccatccaa 1140
 atctcactg gctggcctgg aatcccaaca cacaacatag atgcttttgt tccatctggt 1200
 acatggaaaa gagatgaacg ttattttttt caaggaaatc aatactggag atatgacagt 1260
 gacaaggatc aggcctcac agaagatgaa caaggaaaaa gctatcccaa attgatttca 1320
 gaaggatttc ctggcatccc aagtccccca gacacggcgt tttatgacgg aagacagaag 1380
 ttaatttact tcttcaagga gtcccttcta ttgcatttg atgtcaacag aaatcagata 1440
 cttaatcttt atccaaagag gattactgaa gtttttccag cagtaatacc acaaatcat 1500

WO 01/98468 PCT/US01/19178

```

cctttcagaa ataatagatt cgtttattac tctatgcac acaactccat tttcttttcc 1560
aaaggccaatg cactactggaa ggtagtttaac gacaaggaca aacaacagaa ttcctggcct 1620
cctgctaattg gcttatttcc aaaaaagttt atttcagaga agtgggttga tgtttgtgac 1680
gtccatattct ccacactgaa catgttaa 1707

<210> 41
<211> 1262
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4333459CB1

<400> 41
aaaagatctt tgcaaaacac tacattcaga aacatcagat ggacatgctt gattcaccac 60
gtcttgggta atgaataaac ttgttttaaa ttggcttatt gctggtctct caaggcttcc 120
tatctttggt ttgttttagtc tctctaaaaa ttccagggaat aactatgagt ctcaaaatgc 180
ttataagcag gaacaagctg attttactac taggaattagt ctttttllaa cgaaggtaaat 240
ctgcacactct ttgcgtcccc aaagctcccc gtgttgggca gagtctggtt aaggtaacagc 300
cttggaaatta tttaacatt ttcagtcgaa ttcttggagg aagccaagtg gagaagggtt 360
cctatccctg cgagggtatct ctgaaacaaa ggcagaagca tatttltgga ggaagctatg 420
tctacccaca gtgggtgato acggcggctc actgcattgc aaacagaaac attgtgtcta 480
ctttgaattgt tactgtctga gagtatgaat taagccagac agaccagga gagcaaac 540
tcactattga aactgtcatc atacatccac atttctccac caagaaccca atggactatg 600
atatgtgccc ttgaaagatg gctggagcct tccaatttgg ccaatttgtg gggcccatat 660
gtcttccaga gctgcgggag caatttgagg ctggttttat ttgtacaact gcaaggctggg 720
gctgttaaac tgaaggttgg gtctctccac aagtcttgg ggaagtgaat ctgctattt 780
tgacctggga agagtgtgtg gcagctctgt taacactaaa gaggccatc agtgggaaga 840
cctttctttg cacaggtttt cctgatggag ggagagacgc atgtcaggga gattcaggag 900
gttccactcat gtgcccgaat aagaaaagggt cctggactct ggctggtgtg acttctctgg 960
gtttgggctg tggctcggagc tggagaaaca atgtgaggaa aagtgatcaa ggaatccctg 1020
ggatcttcac agacattagt aaagtgcctt cctggatcca cgaacacatc caaactggta 1080
actaagccat cacacaaggt taagaagctg ccattctgct agggccagag acagcatcag 1140
cagagtcctg gcaaatcaga gcacctgaac caacaggtct taactctgtt ctacgttag 1200
cacacaagga ttgtgaggtt taccagttct aaataaaca agagttaaat atggtaaaaa 1260
aa

```

<210> 42
<211> 3067
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6817347CB1

```

<400> 42
gcactgtgaa cgttgggtgc atccaaatct gaattttgtc tgggaccagg gtcaggggacc 60
agaatcaccc agagctgagg gccagcccta cctgagaacc atcaacaaac ttaccccaca 120
tccattataa cctctcaact cctgcagcc ttctagcttc ccaatctcc caccctcaat 180
gtcacctggg gctctgggtc accagatgac actaactgtt cctgttaca caggccccat 240
gatccccatg gatgttaaat agcccagctc cgtgaccacg gctctctacc tcagctctag 300
ctgcacgat atctcctcat tcttggccac tggtaagaaa ctttccctcc attttgttca 360
tccacgtgag tgtgaatca ccaggattga tgacaaaaat agaaggagat tggaaagacag 420
tgagccaggt gccaaactct tcaataatga tggagtctgt tgttgctgc aaaaacgggg 480
gccagtgaac attacatcag tgtgtgtgag tcccaggacc ttacaaatat cagtttttgt 540
gttatcagag aaatacagag gtattgttaa atttgaatcg gatgaattac cttttgtgtt 600
aattggttct aatattgggt atgcacattt tcaagaattc agggctggaa tctctggaa 660
gctgtggta gatctgatg acccaattcc tcagttccct gatgtgca gacagcagag 720
cagcaggatt ccttcagtga gtgtgctagt tgcagttcct ctggttgcag gccacaaagg 780
gcagcatttt attgaaagga tgcgtgggtg cttaagagaa ttgaagcaag agctgactca 840
ggaaagggcc ggccgggggac accccagggt tgcgtggccc ccgcgccgcc acgcccagtg 900
gccgcccgag cctgcgagc agggggagga gccgccgcca gtggaggcgg agggagtaga 960
ggaggcggag acggcgagga aggcggagag gaaggtggag gggaggcgga aggtggaggg 1020
gaaggcggag gccgcgggga aggcggaggg gccggggagg gtggagccca ccgagaaggt 1080
ggagacggcg gggaaggtgg acgcgcgtgg gaaggtggag acggcgaggg gtcggggccg 1140

```

WO 01/98468

PCT/US01/19178

```

ccgggctgag ctcaagctgg agcccgaacc cgagccgggtc cgggaggcgg agcaggagcc 1200
gaagcaggag ctggaggatg agaaccagac ggggagcggc ggtggggcca acagcgacga 1260
ggttcctccc cccacccttc cctccgaccc accgcgcccc cccgacccct ctcccgctcg 1320
cagtcgttgcg ccgcgcgcgc gaaccgggcc ccggcccccag acccggtccc gtaccgggcc 1380
cgagcctagg ccccgccccc ccgcccggcc ccggcccccgg ccgggcctcg gggcgcgatg 1440
cctggatgtg gattttgcg tggggccacc aggtctgttc caagtgaaca gctttaaggt 1500
gggagagAAC tggaggcagg aactgcgggt tatctaccag tgcctcgtgt ggtgtggaac 1560
cccagagacc aggaaaagca aggcAAAgtc ctgcactctg catgtgtgtg gccccatct 1620
gaacagactc cactcttgcc ttctctgtgt ctctcttggc tgcctcagg agaacacat 1680
tcacgagcac gcagagacga aacaacacaa cttagcagta gacctgtatt acggaggtat 1740
atctgtcttt atgttlaagg actatgtata tgacaaagac atlgagcaaa ttgcaaaaa 1800
agagcaagga gaagctttga aattacaagc ctccacctcc acagaggttt ctaccagca 1860
gtgttcagtg ccaggccttg gtgagaaatt cccaacctgg gaaacaccca aaccagatt 1920
agaactgctg ggcacacacc cgaggagaag aagaatcacc tccagcttta cgatcggttt 1980
aagaggactc atcaatcttg gcaacacgtg cttlatgaac tgcattgtcc aggcctcac 2040
ccacacgcgc atactgagag attctcttct ctctgacagg caccgatgtg agatgccgag 2100
tcccagttg tgtctggtct gtgagatgtc gtctctgtt cgggagttgt attctggaaa 2160
ccgctctct catgtgcctc ataagttact gcaactgtgt tggatcacat ccgctcattt 2220
agcagggac aggcacacag atgcccacga gtctctcatt gcaaggttag atgtcttga 2280
caggtcactgc aaagggtatg atgtcgggaa ggcggccaac aatcccaacc actgtactg 2340
ctcatagac caaatcttca caggtggcct gcagtctgat gtcactgtc aagcctgcca 2400
tggtcctctc accacgatag acccatgctg ggacattagt ttggacttgc ctggctcttg 2460
caccctcttc tggcccatga gccacgggag ggagagcagt gtgaacgggg aaagccacat 2520
accaggaac accaccccca cggactgctt cgggaggttt acgaggccag agcacttagg 2580
aagcagtgcc aaaaacaaat gtggtagtgy ccaagcctac caggaaatcta ccaaacagct 2640
cacaatgaat aaattacctg tcttgcctgy tttcatttc aaacgggttg aacattcagc 2700
gaacacagag ccgaagatca ctacatacat ttcttctct ctggagctgg atatgaagcc 2760
gtttatggcc tcaagtaag agagcagaat gaatggacaa ttgcagctgc caaccatag 2820
tggaacacac gaaaataagt attcctgttt tctgtgtggt aatcccaag gaaccttga 2880
gagtgccac tataccagct tcatccggca ccacaaggac cagtggttca agtgtgatga 2940
tgccgtcatc actaaggcca gtattaaagga cgtactggac agtgaagggt atttactgtt 3000
ctatcacaaa caggtgctag aacatgagtc agaaaaagt aaagaaatga acacacaagc 3060
ctactga
3067

```


【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
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/098468 A3(51) International Patent Classification: C12N 15/57,
9/64, 5/10, A01K 67/027, C07K 16/40, C12Q 1/68, A61K
38/43, G01N 33/68, A61K 39/395, C12Q 1/37

(21) International Application Number: PCT/US01/19178

(22) International Filing Date: 13 June 2001 (13.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/212,336 16 June 2000 (16.06.2000) US
60/213,955 22 June 2000 (22.06.2000) US
60/215,396 29 June 2000 (29.06.2000) US
60/216,821 7 July 2000 (07.07.2000) US
60/218,946 14 July 2000 (14.07.2000) US(71) Applicant (for all designated States except US): INCYTE
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): YUE, Henry
[US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).
ELLIOTT, Vicki, S. [US/US]; 3770 Pulton Place Way, San
Jose, CA 95121 (US). GANDHI, Ameena, R. [US/US];
837 Roble Avenue, #1, Menlo Park, CA 94025 (US).
LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA
95056 (US). AU-YOUNG, Janice [US/US]; 233 Golden
Eagle Lane, Brisbane, CA 94005 (US). TRIBOULEY,
Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San
Francisco, CA 94107 (US). DELEGEANE, Angelo, M.
[US/US]; 594 Angus Drive, Milpitas, CA 95035 (US).
BAUGHN, Mariyah, R. [US/US]; 14244 Santiago Road,
San Leandro, CA 94577 (US). NGUYEN, Dannie, B.
[US/US]; 1403 Ridgewood Drive, San Jose, CA 95118
(US). LEE, Ernestine, A. [US/US]; 624 Kains Street,
Albany, CA 94706 (US). HAFALIA, April [US/US]; 2227
Calle de Primavera, Santa Clara, CA 95054 (US). KHAN,
Farrah, A. [IN/US]; 3617 Central Road #102, Glenview,
IL 60025 (US). WALIA, Narinder, K. [US/US]; 890
Davis Street #205, San Leandro, CA 94577 (US). YAO,
Monique, G. [US/US]; 111 Frederick Court, Mountain
View, CA 94043 (US). LU, Dyung, Aina, M. [US/US];233 Coy Drive, San Jose, CA 95123 (US). PATTERSON,
Chandra [US/US]; 490 Sherwood Way #1, Menlo Park,
CA 94025 (US). TANG, Y., Tom [US/US]; 4230 Ranwick
Court, San Jose, CA 95118 (US). WALSH, Roderick, T.
[US/US]; 8 Boundary Court, St. Lawrence Road, Canter-
bury, Kent CT1 3EZ (GB). AZIMZAI, Yalda [US/US];
5518 Boulder Canyon Drive, Castro Valley, CA 94552
(US). LI, Yun [CN/US]; 3885 Corina Way, Palo Alto,
CA 94303 (US). RAMKUMAR, Jayalaxmi [IN/US];
34359 Maybird Circle, Fremont, CA 94555 (US). XU,
Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA
94040 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley
Avenue #3, Sunnyvale, CA 94086 (US). DAS, Debopriya
[IN/US]; 1179 Bonita Avenue, Apt. 3, Mountain View, CA
94040 (US). KEARNEY, Liam [IE/US]; 50 Woodside
Avenue, San Francisco, CA 94127 (US). KALLICK,
Deborah, A. [US/US]; 900 Olive Street, Menlo Park, CA
94025 (US).(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,
Inc., 3160 Porter 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, CO, CR, CU,
CZ, DE, DK, DM, DZ, ES, FI, GB, GD, GH, GI, GM,
HN, HU, ID, IL, IN, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(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, LU, 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:
3 January 2003For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: PROTEASES

(57) Abstract: The invention provides human proteases (PRIS) and polynucleotides which identify and encode PRIS. The inven-
tion 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 PRIS.

WO 01/098468 A3

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01/19178										
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/57 C12N9/64 C12N5/10 A01K67/027 C07K16/40 C12Q1/68 A61K38/43 G01N33/68 A61K39/395 C12Q1/37										
According to International Patent Classification (IPC) or to both national classification and IPC										
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N A61K C12Q 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) EPD-Internal, EMBL, BIOSIS, SEQUENCE SEARCH, WPI Data, PAJ										
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category *</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>DATABASE EMBL [Online] 1 January 1998 (1998-01-01) B. LAUER ET AL: "Hypothetical 16.6 KDA protein (fragment)" Database accession no. AAG10401 XP002200355 abstract</td> <td>1</td> </tr> <tr> <td>A</td> <td>FUJIWARA TSUTOMU ET AL: "Identification and chromosomal assignment of USPL, a novel gene encoding a human ubiquitin-specific protease" GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 54, no. 1, 15 November 1998 (1998-11-15), pages 155-158, XP002153254 ISSN: 0888-7543 the whole document</td> <td>1-45,66</td> </tr> </tbody> </table>		Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	DATABASE EMBL [Online] 1 January 1998 (1998-01-01) B. LAUER ET AL: "Hypothetical 16.6 KDA protein (fragment)" Database accession no. AAG10401 XP002200355 abstract	1	A	FUJIWARA TSUTOMU ET AL: "Identification and chromosomal assignment of USPL, a novel gene encoding a human ubiquitin-specific protease" GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 54, no. 1, 15 November 1998 (1998-11-15), pages 155-158, XP002153254 ISSN: 0888-7543 the whole document	1-45,66
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.								
X	DATABASE EMBL [Online] 1 January 1998 (1998-01-01) B. LAUER ET AL: "Hypothetical 16.6 KDA protein (fragment)" Database accession no. AAG10401 XP002200355 abstract	1								
A	FUJIWARA TSUTOMU ET AL: "Identification and chromosomal assignment of USPL, a novel gene encoding a human ubiquitin-specific protease" GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 54, no. 1, 15 November 1998 (1998-11-15), pages 155-158, XP002153254 ISSN: 0888-7543 the whole document	1-45,66								
☒ Further documents are listed in the continuation of box C. ☒ Patent family members are listed in annex.										
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document published on or after the international filing date "L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document relating to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family										
Date of the actual completion of the international search 28 May 2002	Date of mailing of the international search report 25. 09. 2002									
Name and mailing address of the ISA European Patent Office, P.O. Box 5818 Patentlaan 2 NL - 2200 LV Rijswijk Tel: (+31-70) 346-5040, Tx: 31 651 epo nl, Fax: (+31-70) 340-5018	Authorized officer LE CORNEC N.D.R.									

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

INTERNATIONAL SEARCH REPORT

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		International Application No PCT/US 01/19178
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	D'ANDREA A ET AL: "DEUBIQUITINATING ENZYMES A NEW CLASS OF BIOLOGICAL REGULATORS" CRITICAL REVIEWS IN BIOCHEMISTRY AND MOLECULAR BIOLOGY, CRC PRESS, BOCA RATON, FL, US, vol. 33, no. 5, October 1998 (1998-10), pages 337-352, XP000943127 ISSN: 1040-9238 the whole document	1-45,66
A	YASUO KAJIMOTO ET AL: "cDNA Cloning and tissue distribution of a rat Ubiquitin carboxyl-terminal hydrolase PGP9.5" JOURNAL OF BIOCHEMISTRY., vol. 112, 1992, pages 28-32, XP002200351 JAPANESE BIOCHEMICAL SOCIETY, TOKYO., JP ISSN: 0021-924X the whole document	1-45,66
A	SLOPER-MOULD KATHERINE E ET AL: "Characterization and chromosomal localization of USP3, a novel human ubiquitin-specific protease." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 274, no. 38, pages 26878-26884, XP002200352 ISSN: 0021-9258 cited in the application the whole document	1-45,66
A	DATABASE EMBL [Online] 27 August 1995 (1995-08-27) T. FUJIIWARA ET AL: "Human fetal brain cDNA Cdna 5'-END gen-184b02" Database accession no. D61366 XP002200356 abstract & UNPUBLISHED,	11-15
A	DATABASE EMBL [Online] 21 June 1999 (1999-06-21) S. ZHAO ET AL: "CITBI-E1-2646J3:TF CITBI-E1 Homo sapiens genomic clone 2646J3, genomic survey sequence" Database accession no. A0626556 XP002200357 abstract & UNPUBLISHED,	11-15
A	WO 99 25373 A (MILLENNIUM PHARM INC) 27 May 1999 (1999-05-27) the whole document	1-45,66
2	---	---

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

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 01/19178

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE EMBL [Online] 11 September 2000 (2000-09-11) O. OHARA ET AL: "Homo sapiens mRNA for KIAA1594 protein, partial cds" Database accession no. AB046814 XP002200358 abstract -& T. NAGASE ET AL: "Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro" DNA RESEARCH, vol. 7, 31 August 2000 (2000-08-31), pages 273-281, XP002200353 ---	1-7, 11-15
T	DATABASE EMBL [Online] 6 September 2000 (2000-09-06) J. KIM ET AL: "ubiquitin-specific processing protease (Homo sapiens)" Database accession no. AA610401 XP002200359 cited in the application abstract -& J. KIM ET AL: "Discovery of a novel, paternally expressed Ubiquitin-specific processing protease gene through comparative analysis of an imprinted region of mouse chromosome 7 and human chromosome 19q13.4" GENOME RESEARCH, vol. 10, no. 8, August 2000 (2000-08), pages 1138-1147, XP002200354 ---	
P,X	WO 00 57903 A (HUMAN GENOME SCIENCES INC ;ROSEN CRAIG A (US); RUBEN STEVEN M (US)) 5 October 2000 (2000-10-05) sequence ID no.18 and 114 claims ---	1
P,A	---	2-45,66
E	WO 02 00860 A (SUGEN INC ;WHYTE DAVID (US); CAENEPEEL SEAN (US); CHARYDCZAK GLEN) 3 January 2002 (2002-01-03) figures 2C,16-H claims -----	1-45,66

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		International application No. PCT/US 01/19178
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)		
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1. ☒	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely: see FURTHER INFORMATION sheet PCT/ISA/210	
2. ☒	Claims Nos.: 20-21, 23-24 all partially because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically: see FURTHER INFORMATION sheet PCT/ISA/210	
3. ☐	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:		
1. ☐	As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.	
2. ☐	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.	
3. ☒	As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.: Claims 1-44, all partially and 45, 66 completely.	
4. ☐	No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:	
Remark on Protest	☐ The additional search fees were accompanied by the applicant's protest. ☐ No protest accompanied the payment of additional search fees.	

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01/19178

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 18, 24 are directed to a method of treatment of the human/animal body (rule 39.1 IV PCT), the search has been carried out and based on the alleged effects of the compound/composition.
Although claim 32, 34 are directed to a diagnostic method practised on the human/animal body (rule 39.1 IV PCT), the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.2

Claims Nos.: 20-21, 23-24 all partially

Claim 20 refers to a composition comprising an agonist of a polypeptide, claim 21 refers to a method of treatment using said composition, claim 23 refers to a composition comprising an antagonist of the polypeptide and claim 24 to a method of treatment using the composition comprising said antagonist without a true technical characterization. Moreover, no such 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.

No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved. But a partial search has been carried out as far as the agonist and/or the antagonist is an antibody, an antisense nucleic acid or a ribozyme.

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

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01/19178

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-44 all partially and 45, 66 totally

Polypeptide represented by the amino acid sequence ID no.1 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.22. Antibodies. Therapeutic and diagnostic uses.

2. Claims: 1-44 all partially and 46, 67 totally

Polypeptide represented by the amino acid sequence ID no.2 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.23. Antibodies. Therapeutic and diagnostic uses.

3. Claims: 1-44 all partially and 47, 68 totally

Polypeptide represented by the amino acid sequence ID no.3 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.24. Antibodies. Therapeutic and diagnostic uses.

4. Claims: 1-44 all partially and 48, 69 totally

Polypeptide represented by the amino acid sequence ID no.4 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.25. Antibodies. Therapeutic and diagnostic uses.

5. Claims: 1-44 all partially and 49, 70 totally

Polypeptide represented by the amino acid sequence ID no.5 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.26. Antibodies. Therapeutic and diagnostic uses.

6. Claims: 1-44 all partially and 50, 71 totally

Polypeptide represented by the amino acid sequence ID no.6 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.27. Antibodies. Therapeutic and diagnostic uses.

7. Claims: 1-44 all partially and 51, 72 totally

Polypeptide represented by the amino acid sequence ID no.7 and its corresponding polynucleotide represented by the

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01/19178

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

nucleic acid sequence ID no.28. Antibodies. Therapeutic and diagnostic uses.

8. Claims: 1-44 all partially and 52, 73 totally

Polypeptide represented by the amino acid sequence ID no.8 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.29. Antibodies. Therapeutic and diagnostic uses.

9. Claims: 1-44 all partially and 53, 74 totally

Polypeptide represented by the amino acid sequence ID no.9 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.30. Antibodies. Therapeutic and diagnostic uses.

10. Claims: 1-44 all partially and 54, 75 totally

Polypeptide represented by the amino acid sequence ID no.10 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.31. Antibodies. Therapeutic and diagnostic uses.

11. Claims: 1-44 all partially and 55, 76 totally

Polypeptide represented by the amino acid sequence ID no.11 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.32. Antibodies. Therapeutic and diagnostic uses.

12. Claims: 1-44 all partially and 56, 77 totally

Polypeptide represented by the amino acid sequence ID no.12 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.33. Antibodies. Therapeutic and diagnostic uses.

13. Claims: 1-44 all partially and 57, 78 totally

Polypeptide represented by the amino acid sequence ID no.13 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.34. Antibodies. Therapeutic and diagnostic uses.

14. Claims: 1-44 all partially and 58, 79 totally

Polypeptide represented by the amino acid sequence ID no.14

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01/19178

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

and its corresponding polynucleotide represented by the nucleic acid sequence ID no.35. Antibodies. Therapeutic and diagnostic uses.

15. Claims: 1-44 all partially and 59, 80 totally

Polypeptide represented by the amino acid sequence ID no.15 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.36. Antibodies. Therapeutic and diagnostic uses.

16. Claims: 1-44 all partially and 60, 81 totally.

Polypeptide represented by the amino acid sequence ID no.16 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.37. Antibodies. Therapeutic and diagnostic uses.

17. Claims: 1-44 all partially and 61, 82 totally

Polypeptide represented by the amino acid sequence ID no.17 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.38. Antibodies. Therapeutic and diagnostic uses.

18. Claims: 1-44 all partially and 62, 83 totally

Polypeptide represented by the amino acid sequence ID no.18 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.39. Antibodies. Therapeutic and diagnostic uses.

19. Claims: 1-44 all partially and 63, 84 totally

Polypeptide represented by the amino acid sequence ID no.19 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.40. Antibodies. Therapeutic and diagnostic uses.

20. Claims: 1-44 all partially and 64, 85 totally

Polypeptide represented by the amino acid sequence ID no.20 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.41. Antibodies. Therapeutic and diagnostic uses.

21. Claims: 1-44 all partially and 65, 86 totally

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 01/19178

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Polypeptide represented by the amino acid sequence ID no.21 and its corresponding polynucleotide represented by the nucleic acid sequence ID no.42. Antibodies. Therapeutic and diagnostic uses.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.
PCT/US 01/19178

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9925373	A	27-05-1999	US 5932422 A 03-08-1999 AU 1409199 A 07-06-1999 WO 9925373 A1 27-05-1999 US 6322982 B1 27-11-2001
WO 0057903	A	05-10-2000	AU 3766800 A 16-10-2000 EP 1165123 A1 02-01-2002 WO 0057903 A2 05-10-2000
WO 0200860	A	03-01-2002	AU 7142701 A 08-01-2002 WO 0200860 A2 03-01-2002 US 2002064856 A1 30-05-2002

Form PCT/ISA/210 (patent family annex) (July 1999)

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 29/00	A 6 1 P 35/00	4 B 0 6 5
A 6 1 P 35/00	C 0 7 K 16/40	4 C 0 8 4
C 0 7 K 16/40	C 1 2 N 1/15	4 C 0 8 5
C 1 2 N 1/15	C 1 2 N 1/19	4 H 0 4 5
C 1 2 N 1/19	C 1 2 N 1/21	
C 1 2 N 1/21	C 1 2 N 9/48	
C 1 2 N 5/10	C 1 2 P 21/08	
C 1 2 N 9/48	C 1 2 Q 1/37	
C 1 2 N 15/02	C 1 2 Q 1/68	A
C 1 2 P 21/08	G 0 1 N 33/15	Z
C 1 2 Q 1/37	G 0 1 N 33/50	Z
C 1 2 Q 1/68	G 0 1 N 33/53	D
G 0 1 N 33/15	G 0 1 N 33/53	M
G 0 1 N 33/50	G 0 1 N 33/566	
G 0 1 N 33/53	C 1 2 N 5/00	A
G 0 1 N 33/566	C 1 2 N 15/00	C
	A 6 1 K 37/02	

(31)優先権主張番号 60/216,821

(32)優先日 平成12年7月7日(2000.7.7)

(33)優先権主張国 米国(US)

(31)優先権主張番号 60/218,946

(32)優先日 平成12年7月14日(2000.7.14)

(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,CO,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,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 エリオット、ビッキー・エス

アメリカ合衆国カリフォルニア州9 5 1 2 1・サンノゼ・ポルトンプレイスウェイ 3 7 7 0

(72)発明者 ガンディー、アミーナ・アール

アメリカ合衆国カリフォルニア州9 4 0 2 5・メンロパーク・# 1・ローブルアベニュー 8 3 7

(72)発明者 ラル、プリーティ

アメリカ合衆国カリフォルニア州9 5 0 5 6・サンタクララ・ピーオーボックス 5 1 4 2

(72)発明者 オウーヤング、ジャニス

アメリカ合衆国カリフォルニア州9 4 0 0 5・ブリスベン・ゴールデンイーグルレーン 2 3 3

(72)発明者 トリボレー、キャサリン・エム

アメリカ合衆国カリフォルニア州9 4 1 0 7・サンフランシスコ・# 5・テネシーストリート 1 1 2 1

(72)発明者 デレジーン、アンジェロ・エム

アメリカ合衆国カリフォルニア州9 5 0 3 5・ミルピタス・アングスドライブ 5 9 4

(72)発明者 ボーグン、マライア・アール

アメリカ合衆国カリフォルニア州9 4 5 7 7・サンレアンドロ・サンティアゴロード 1 4 2 4 4

(72)発明者 ニュエン、ダニエル・ビー

- アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・サンノゼ・リッジウッドドライブ 1 4 0 3
(72)発明者 リー、アーンステーション・エイ
- アメリカ合衆国カリフォルニア州 9 4 7 0 6 ・アルバニー・ケインズストリート 6 2 4
(72)発明者 ハファリア、エープリル
- アメリカ合衆国カリフォルニア州 9 5 0 5 4 ・サンタクララ・コーレデブリマベール 2 2 2 7
(72)発明者 カーン、ファラ・エイ
- アメリカ合衆国イリノイ州 6 0 0 2 5 ・グレンビュー・# 1 0 2 ・セントラルロード 3 6 1 7
(72)発明者 チョーラ、ナリンダー・ケイ
- アメリカ合衆国カリフォルニア州 9 4 5 8 7 ・ユニオンシティ・# 7 1 2 ・ユニオンスクエア 3
3
(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 9 4 0 4 3 ・マウンテンビュー・フレデリックコート 1 1 1
(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 9 5 1 2 3 ・サンノゼ・コイドライブ 2 3 3
(72)発明者 アービズ、チャンドラ・エス
- アメリカ合衆国カリフォルニア州 9 5 1 2 5 ・サンノゼ・モロッコドライブ 1 7 0 6
(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 9 5 1 1 8 ・サンノゼ・ランウィックコート 4 2 3 0
(72)発明者 ウォルシュ、ロドリック・ティー
- イギリス国ケント州・シーティー 1 3 イーजीー・カンタベリー・セントローレンスロード・バ
ウンダリーコート 8
(72)発明者 アジムザイ、ヤルダ
- アメリカ合衆国カリフォルニア州 9 4 5 5 2 ・カストロバレー・ボールダーキャニオンドライブ
5 5 1 8
(72)発明者 リュ、ヤン
- アメリカ合衆国カリフォルニア州 9 4 3 0 3 ・パロアルト・コリーナウェイ 3 8 8 5
(72)発明者 ランクマール、ジャヤラクシミ
- アメリカ合衆国カリフォルニア州 9 4 5 5 5 ・フレモント・メイバードサークル 3 4 3 5 9
(72)発明者 スー、ユーミング
- アメリカ合衆国カリフォルニア州 9 4 0 4 0 ・マウンテンビュー・ウォルナットドライブ 1 7 3
9
(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 9 4 0 8 6 ・サニーバイル・# 3 ・ウェストマッキンレーアベニ
ュー 1 2 3 3
(72)発明者 ダス、デボプリバ
- アメリカ合衆国カリフォルニア州 9 4 0 4 0 ・マウンテンビュー・アパートメント 3 ・ボニータ
アベニュー 1 1 7 9
(72)発明者 キーニー、ライアム
- アメリカ合衆国カリフォルニア州 9 4 1 2 7 ・サンフランシスコ・ウッドサイドアベニュー 5 0
(72)発明者 カリック、デボラー・エイ
- アメリカ合衆国カリフォルニア州 9 4 0 2 5 ・メンロパーク・オリーブストリート 9 0 0

F ターム(参考) 2G045 AA34 AA35 AA40 BA11 BB50 DA12 DA13 DA14 DA36 FB02

FB03

4B024 AA01 BA14 CA04 DA06 EA04 GA11 HA01 HA14 HA15

4B050 CC03 DD11 LL01

4B063 QA01 QA18 QQ36 QQ43 QQ79 QR16 QR32 QR55 QS02 QS33

QS34

4B064 AG27 CA02 CA19 CC24 DA01

4B065 AA26X AA99Y AB01 AC14 BA02 CA33 CA44

4C084 AA01 AA16 BA02 BA08 BA22 CA18 DC03 NA14 ZA662 ZB112

ZB262

4C085 KA04

4H045 AA11 CA40 DA89 EA50 FA74

专利名称(译)	蛋白酶		
公开(公告)号	JP2004515217A	公开(公告)日	2004-05-27
申请号	JP2002504617	申请日	2001-06-13
[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	ユエヘンリー エリオットビッキーエス ガンディーアミーナアール ラルプリーティ オウヤングジャニス トリボレーキャサリーンエム デレジーンアンジェロエム ボーグンマライアアール ニュエンダニエルビー リーアーンステイーンエイ ハファリアエープリル カーンファラエイ チョーラナリンダーケイ ヤオモニークジー リュデユングアイナエム アービズチャンドラエス タングワイトム ウォルシュロドリックティー アジムザイヤルダ リュヤン ランクマールジャヤラクシミ スーユーミング レディルーパ ダスデボプリバ キーニーライアム カリックデボラーエイ		
发明人	ユエ、ヘンリー エリオット、ビッキー・エス ガンディー、アミーナ・アール ラル、プリーティ オウ・ヤング、ジャニス トリボレー、キャサリーン・エム デレジーン、アンジェロ・エム ボーグン、マライア・アール ニュエン、ダニエル・ビー リー、アーンステイーン・エイ ハファリア、エープリル カーン、ファラ・エイ チョーラ、ナリンダー・ケイ ヤオ、モニーク・ジー リュ、デユング・アイナ・エム アービズ、チャンドラ・エス タング、ワイ・トム ウォルシュ、ロドリック・ティー		

アジムザイ、ヤルダ
リュ、ヤン
ランクマール、ジャヤラクシミ
スー、ユーミング
レディ、ルーパ
ダス、デボプリバ
キーニー、ライアム
カリック、デボラー・エイ

IPC分类号	G01N33/50 A61K38/00 A61K45/00 A61K49/00 A61P1/04 A61P29/00 A61P35/00 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/48 C12N9/64 C12N15/02 C12N15/09 C12N15/57 C12P21/08 C12Q1/37 C12Q1/68 G01N33/15 G01N33/53 G01N33/566
CPC分类号	A61P1/04 A61P29/00 A61P35/00 C12N9/6421
FI分类号	C12N15/00.ZNA.A A61K45/00 A61K49/00.A A61P1/04 A61P29/00 A61P35/00 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N9/48 C12P21/08 C12Q1/37 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 C12N5/00.A C12N15/00.C A61K37/02
F-TERM分类号	2G045/AA34 2G045/AA35 2G045/AA40 2G045/BA11 2G045/BB50 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/FB02 2G045/FB03 4B024/AA01 4B024/BA14 4B024/CA04 4B024/DA06 4B024/EA04 4B024/GA11 4B024/HA01 4B024/HA14 4B024/HA15 4B050/CC03 4B050/DD11 4B050/LL01 4B063/QA01 4B063/QA18 4B063/QQ36 4B063/QQ43 4B063/QQ79 4B063/QR16 4B063/QR32 4B063/QR55 4B063/QS02 4B063/QS33 4B063/QS34 4B064/AG27 4B064/CA02 4B064/CA19 4B064/CC24 4B064/DA01 4B065/AA26X 4B065/AA99Y 4B065/AB01 4B065/AC14 4B065/BA02 4B065/CA33 4B065/CA44 4C084/AA01 4C084/AA16 4C084/BA02 4C084/BA08 4C084/BA22 4C084/CA18 4C084/DC03 4C084/NA14 4C084/ZA662 4C084/ZB112 4C084/ZB262 4C085/KA04 4H045/AA11 4H045/CA40 4H045/DA89 4H045/EA50 4H045/FA74
優先権	60/212336 2000-06-16 US 60/213955 2000-06-22 US 60/215396 2000-06-29 US 60/216821 2000-07-07 US 60/218946 2000-07-14 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。 本发明还提供了用于诊断，治疗或预防与PRTS异常表达有关的疾病的方法。

インサート プロセストID	インサート ID NO.	インサート ID NO.	インサート ID NO.	インサート ID NO.	インサート ID NO.
275791	1	275791C01	22	275791C01	22
1389445	2	1389445C01	23	1389445C01	23
1726609	3	1726609C01	24	1726609C01	24
4503846	4	4503846C01	25	4503846C01	25
5544089	5	5544089C01	26	5544089C01	26
7474081	6	7474081C01	27	7474081C01	27
5281209	7	5281209C01	28	5281209C01	28
2256251	8	2256251C01	29	2256251C01	29
7160544	9	7160544C01	30	7160544C01	30
7477386	10	7477386C01	31	7477386C01	31
7473309	11	7473309C01	32	7473309C01	32
7684035	12	7684035C01	33	7684035C01	33
3473847	13	3473847C01	34	3473847C01	34
3751004	14	3751004C01	35	3751004C01	35
4904126	15	4904126C01	36	4904126C01	36
71268415	16	71268415C01	37	71268415C01	37
7473301	17	7473301C01	38	7473301C01	38
7473308	18	7473308C01	39	7473308C01	39
7478021	19	7478021C01	40	7478021C01	40
4334659	20	4334659C01	41	4334659C01	41
6817347	21	6817347C01	42	6817347C01	42