

(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2004-521637

(P2004-521637A)

(43) 公表日 平成16年7月22日(2004.7.22)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A 4 B O 2 4
A 6 1 K 38/00	A 6 1 K 39/395	D 4 B O 2 9
A 6 1 K 39/395	A 6 1 K 39/395	N 4 B O 5 0
A 6 1 K 45/00	A 6 1 K 45/00	4 B O 6 3
A 6 1 P 1/04	A 6 1 P 1/04	4 B O 6 4
	審査請求 未請求 予備審査請求 有	(全 255 頁) 最終頁に続く

(21) 出願番号	特願2002-587592 (P2002-587592)	(71) 出願人	301005050 インサイト・ゲノミックス・インコーポレイテッド 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) (22) 出願日	平成14年1月16日 (2002.1.16)	(74) 代理人	100089266 弁理士 大島 陽一
(85) 翻訳文提出日	平成15年7月16日 (2003.7.16)	(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94706・アルバニー・ケインズストリート 624
(86) 国際出願番号	PCT/US2002/001369		
(87) 国際公開番号	W02002/090530		
(87) 国際公開日	平成14年11月14日 (2002.11.14)		
(31) 優先権主張番号	60/263, 083		
(32) 優先日	平成13年1月18日 (2001.1.18)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/271, 205		
(32) 優先日	平成13年2月23日 (2001.2.23)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/271, 117		
(32) 優先日	平成13年2月23日 (2001.2.23)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 キナーゼ及びホスホターゼ

(57) 【要約】

本発明はヒトキナーゼおよびホスファターゼ(KPP)およびKPPを同定し、コードするポリヌクレオチドを提供する。本発明はまた、発現ベクター、宿主細胞、抗体、アゴニストおよびアンタゴニストをも提供する。本発明はまた、KPPの異常発現に関連する疾患を診断、治療または予防する方法をも提供する。

【特許請求の範囲】

【請求項 1】

以下の (a) 乃至 (d) からなる群から選択した単離されたポリペプチド。

- (a) SEQ ID NO:1-8を有する群から選択したアミノ酸配列を含むポリペプチド
- (b) SEQ ID NO:1-8からなる群から選択したアミノ酸配列と少なくとも90%が同一であるような天然のアミノ酸配列を含むポリペプチド
- (c) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片
- (d) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片

10

【請求項 2】

SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有する、請求項 1 に記載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:9-16からなる群から選択したポリヌクレオチド配列を有する、請求項 4 に記載の単離されたポリヌクレオチド。

20

【請求項 6】

請求項 3 に記載のポリヌクレオチドに機能的に連結し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 に記載の組換えポリヌクレオチドを用いて形質転換した細胞。

【請求項 8】

請求項 6 に記載の組換えポリヌクレオチドを含む遺伝形質転換体。

【請求項 9】

請求項 1 に記載のポリペプチドを生産する方法であって、

- (a) 前記ポリペプチドの発現に好適な条件下で、請求項 1 に記載された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連結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過程と、
- (b) そのように発現した前記ポリペプチドを回収する過程とからなる方法。

30

【請求項 10】

前記ポリペプチドが、SEQ ID NO:1-8からなる群から選択したアミノ酸配列を持つことを特徴とする、請求項 9 に記載の方法。

【請求項 11】

請求項 1 に記載のポリペプチドと特異的に結合する単離された抗体。

【請求項 12】

以下の (a) 乃至 (e) からなる群から選択した単離されたポリヌクレオチド。

40

- (a) SEQ ID NO:9-16を有する群から選択したポリヌクレオチド配列を有するポリヌクレオチド
- (b) SEQ ID NO:9-16を有する群から選択したポリヌクレオチド配列と少なくとも90%が同一であるような天然のポリヌクレオチド配列を有するポリヌクレオチド
- (c) (a) のポリヌクレオチドに相補的なポリヌクレオチド
- (d) (b) のポリヌクレオチドに相補的なポリヌクレオチド
- (e) (a) ~ (d) のRNA等価物

【請求項 13】

請求項 12 に記載のポリヌクレオチドの少なくとも60の連続したヌクレオチドを含む単離されたポリヌクレオチド。

50

【請求項14】

請求項12に記載のポリヌクレオチドの配列を有する標的ポリヌクレオチドをサンプル中から検出する方法であって、

(a)前記サンプル中の前記標的ポリヌクレオチドに相補的な配列を持つ少なくとも20の連続したヌクレオチドを持つプローブを用いて前記サンプルをハイブリダイズする過程と、前記プローブと前記標的ポリヌクレオチドあるい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、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ことを特徴とする過程と、

(b)前記ハイブリダイゼーション複合体の有無を検出し、該複合体が存在する場合にはオプションでその量を検出する過程とを含む方法。

10

【請求項15】

前記プローブが少なくとも60の連続したヌクレオチドを含むことを特徴とする請求項14に記載の方法。

【請求項16】

請求項12に記載のポリヌクレオチドの配列を有する標的ポリヌクレオチドをサンプル中から検出する方法であって、

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

(b)前記の増幅した標的ポリヌクレオチドまたはその断片の存在・不存在を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程を含むことを特徴とする方法。

20

【請求項17】

請求項1に記載のポリペプチドと、薬剤として許容できる賦形剤とを含む組成物。

【請求項18】

前記ポリペプチドが、SEQ ID NO:1-8からなる群から選択されたアミノ酸配列を持つことを特徴とする、請求項17の組成物。

【請求項19】

機能的なKPP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17の組成物を投与することを特徴とする治療方法。

【請求項20】

請求項1に記載のポリペプチドのアゴニストとして有効性を確認するために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すステップと、

(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。

30

【請求項21】

請求項20に記載の方法によって同定したアゴニスト化合物と、薬剤として許容できる賦形剤とを含むことを特徴とする組成物。

【請求項22】

機能性KPPの発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して請求項21に記載の組成物を投与する過程を含むことを特徴とする方法。

40

【請求項23】

請求項1に記載のポリペプチドのアンタゴニストとして有効性を確認するために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すステップと、

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

【請求項24】

請求項23に記載の方法によって同定したアンタゴニスト化合物と、薬剤として許容できる

50

賦形剤とを含む組成物。

【請求項 25】

機能性KPPの過剰発現に関連する疾患又は病状の治療方法であって、そのような治療を必要とする患者に対して請求項24に記載の組成物を投与する過程を含むことを特徴とする方法。

【請求項 26】

請求項1に記載のポリペプチドに特異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のポリペプチドを適切な条件下で少なくとも1つの試験化合物に混合させる過程と、

(b)請求項1のポリペプチドの試験化合物との結合を検出し、それによって請求項1のポリペプチドに特異結合する化合物を同定する過程を含む方法。

10

【請求項 27】

請求項1に記載の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であって、

(a)請求項1のポリペプチドの活性が許容される条件下で、請求項1に記載のポリペプチドを少なくとも1つの試験化合物に混合させる過程と、

(b)請求項1に記載のポリペプチドの活性を試験化合物の存在下で算定する過程と、

(c)試験化合物の存在下での請求項1に記載のポリペプチドの活性を、試験化合物の不存在下での請求項1に記載のポリペプチドの活性と比較する過程を含み、試験化合物の存在下での請求項1に記載のポリペプチドの活性の変化が、請求項1に記載のポリペプチドの活性を調節する化合物を標示することを特徴とする方法。

20

【請求項 28】

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

(a)前記標的ポリヌクレオチドの発現に好適な条件下で、該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過程と、

(b)前記標的ポリヌクレオチドの発現改変を検出する過程と、

(c)可変量の前記化合物の存在下と前記化合物の不存在下で、前記標的ポリヌクレオチドの発現を比較する過程とを含むことを特徴とする方法。

30

【請求項 29】

試験化合物の毒性を算定する方法であって、

(a)核酸を含む生物学的サンプルを前記試験化合物で処理する過程と、

(b)処理した前記生物学的サンプルの核酸と、請求項12のポリヌクレオチドの少なくとも20の連続するヌクレオチドを持つプローブをハイブリダイズさせる過程であって、このハイブリダイゼーションゼーションが、前記プローブと前記生物学的サンプル中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12のポリヌクレオチドまたはその断片のポリヌクレオチド配列を持つポリヌクレオチドである、前記過程と、

(c)ハイブリダイゼーション複合体の収量を定量する過程と、

(d)前記処理された生物学的サンプル中のハイブリタイゼーション複合体の量を、処理されていない生物学的サンプル中のハイブリタイゼーション複合体の量と比較する過程とを含み、前記処理された生物学的サンプル中のハイブリタイゼーション複合体の量の差が、前記試験化合物の毒性を標示するような方法。

40

【請求項 30】

生物学的サンプル中のKPPの発現に関連する症状または疾患に対する診断試験法であって、

(a)前記生物学的サンプルと請求項11の抗体との混合を、前記抗体が前記ポリペプチドに結合し、抗体とポリペプチドとの複合体を形成するのに適した条件下で行う過程と、

(b)前記複合体を検出する過程とを含み、前記複合体の存在が、前記生物学的サンプル中

50

の前記ポリペプチドの存在と相関することを特徴とする方法。

【請求項 3 1】

請求項11の抗体であって、

- (a)キメラ抗体
- (b)単鎖抗体
- (c)Fab断片
- (d)F(ab')₂断片
- (e)ヒト化抗体のいずれかである抗体。

【請求項 3 2】

請求項11に記載の抗体と、許容できる賦形剤とを含む組成物。

10

【請求項 3 3】

被検者中のKPPの発現に関連する病状又は疾患の診断方法であって、請求項32に記載の組成物の有効量を前記被検者に投与する過程を含むことを特徴とする方法。

【請求項 3 4】

前記抗体が標識されることを特徴とする請求項32に記載の組成物。

【請求項 3 5】

被検者中のKPPの発現に関連する病状又は疾患の診断方法であって、請求項34に記載の組成物の有効量を前記被検者に投与する過程を含むことを特徴とする方法。

【請求項 3 6】

請求項11に記載の抗体の特異性を有するポリクローナル抗体を調製する方法であって、

20

- (a)抗体反応を誘発する条件下で、SEQ ID NO:1-8からなる群から選択したアミノ酸配列またはその免疫原性断片を含むポリペプチドを用いて動物を免疫化する過程と、
- (b)前記動物から抗体を単離する過程と、
- (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、それによって、SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチドに特異結合するポリクローナル抗体を同定する過程とを含むような方法。

【請求項 3 7】

請求項36に記載の方法で産出したポリクローナル抗体。

【請求項 3 8】

請求項37に記載のポリクローナル抗体及び適切なキャリアを含む組成物。

30

【請求項 3 9】

請求項11に記載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)抗体反応を誘発する条件下で、SEQ ID NO:1-8からなる群から選択したアミノ酸配列またはその免疫原性断片を含むポリペプチドを用いて動物を免疫化する過程と、
- (b)前記動物から抗体産出細胞を単離する過程と、
- (c)前記抗体産出細胞と不死化した細胞とを融合して、モノクローナル抗体を産出するハイブリドーマ細胞を形成する過程と、
- (d)前記ハイブリドーマ細胞を培養する過程と、
- (e)SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチドに特異結合するようなモノクローナル抗体を前記培養物から単離する過程とを含むことを特徴とする方法。

40

【請求項 4 0】

請求項39に記載の方法で産出したモノクローナル抗体。

【請求項 4 1】

請求項40に記載のモノクローナル抗体及び適切なキャリアを含む組成物。

【請求項 4 2】

Fab発現ライブラリのスクリーニングにより前記抗体を産出することを特徴とする請求項11に記載の抗体。

【請求項 4 3】

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前記抗体を産出することを特徴

50

とする請求項11に記載の抗体。

【請求項44】

サンプル中のSEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチドを検出する方法であって、

(a)請求項11に記載の抗体と前記ポリペプチドとの特異結合を許容する条件下で、前記抗体と1サンプルとをインキュベートする過程と、

(b)特異結合を検出する過程とを含み、該特異結合が、SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドがサンプル中に存在することを標示することを特徴とする方法。

【請求項45】

SEQ ID NO:1-8 からなる群から選択したアミノ酸配列を含むポリペプチドを精製する方法であって、

(a)請求項11に記載の抗体と前記ポリペプチドとの特異結合を許容する条件下で、前記抗体と1サンプルとをインキュベートする過程と、

(b)前記サンプルから前記抗体を分離し、SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含む精製したポリペプチドを得る過程とを含むことを特徴とする方法。

【請求項46】

マイクロアレイの少なくとも1つのエレメントが請求項13に記載のポリヌクレオチドであるマイクロアレイ。

【請求項47】

ポリヌクレオチドを含むサンプルの発現プロファイルを作製する方法であって、

(a)サンプル中のポリヌクレオチドを標識化すること

(b)ハイブリダイゼーション複合体が形成されるのに適した条件下で請求項46に記載のマイクロアレイのエレメントとサンプル中の標識化ポリヌクレオチドを接触させる方法。

(c)サンプル中のポリヌクレオチドの発現量を定量するステップを含む。

【請求項48】

固体上の別個の物理的位置に固定した異なるヌクレオチド分子を含むアレイであり、該ヌクレオチド分子の少なくとも1つが標的ポリヌクレオチドの少なくとも30の連続するヌクレオチドと特異的にハイブリダイゼーション可能な最初のオリゴヌクレオチドまたはポリヌクレオチド配列を含み、標的ポリヌクレオチドは請求項12に記載のポリヌクレオチドである。

【請求項49】

請求項48に記載のアレイであり、最初のオリゴヌクレオチドまたはポリヌクレオチド配列は標的ポリヌクレオチドの少なくとも30の連続するヌクレオチドに完全に相補的である。

【請求項50】

請求項48に記載のアレイであり、最初のオリゴヌクレオチドまたはポリヌクレオチド配列は標的ポリヌクレオチドの少なくとも60の連続するヌクレオチドに完全に相補的である。

【請求項51】

請求項48に記載のアレイであり、最初のオリゴヌクレオチドまたはポリヌクレオチド配列は標的ポリヌクレオチドに完全に相補的である。

【請求項52】

請求項48に記載のアレイであり、マイクロアレイである。

【請求項53】

請求項48に記載のアレイであり、最初のオリゴヌクレオチドまたはポリヌクレオチド配列を含むヌクレオチド分子にハイブリダイズされた標的ポリヌクレオチドをも含む。

【請求項54】

請求項48に記載のアレイであり、リンカーは少なくとも1つのヌクレオチド分子を固体基板に接合する。

【請求項55】

請求項48に記載のアレイであり、基板上のそれぞれ別個の物理的位置には複数のヌクレオ

10

20

30

40

50

チド分子を含む。またいかなる単一な別個の物理位置の複数のヌクレオチド分子は同じ配列を有する。基板上の各別個の物理的位置には基板上の別の物理的位置でのヌクレオチド分子の配列と異なる配列を含むヌクレオチド分子が含まれる。

【請求項 5 6】

SEQ ID NO:1のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 5 7】

SEQ ID NO:2のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 5 8】

SEQ ID NO:3のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 5 9】

SEQ ID NO:4のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 0】

SEQ ID NO:5のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 1】

SEQ ID NO:6のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 2】

SEQ ID NO:7のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 3】

SEQ ID NO:8のアミノ酸配列を有する請求項 1 に記載のポリペプチド。

【請求項 6 4】

SEQ ID NO:9のポリヌクレオチド配列を有する請求項 12 に記載のポリヌクレオチド。

【請求項 6 5】

SEQ ID NO:10のポリヌクレオチド配列を有する請求項 12 に記載のポリヌクレオチド。

【請求項 6 6】

SEQ ID NO:11のポリヌクレオチド配列を有する請求項 12 に記載のポリヌクレオチド。

【請求項 6 7】

SEQ ID NO:12のポリヌクレオチド配列を有する請求項 12 に記載のポリヌクレオチド。

【請求項 6 8】

SEQ ID NO:13のポリヌクレオチド配列を有する請求項 12 に記載のポリヌクレオチド。

【請求項 6 9】

SEQ ID NO:14のポリヌクレオチド配列を有する請求項 12 に記載のポリヌクレオチド。

【請求項 7 0】

SEQ ID NO:15のポリヌクレオチド配列を有する請求項 12 に記載のポリヌクレオチド。

【請求項 7 1】

SEQ ID NO:16のポリヌクレオチド配列を有する請求項 12 に記載のポリヌクレオチド。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、キナーゼ及びホスホタ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心血管疾患、免疫系の疾患、神経の疾患、成長および発達に影響を及ぼす疾患、脂質異常、細胞増殖異常及び癌の診断・治療・予防に関する。本発明はさらに、キナーゼ及びホスホタ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。

【背景技術】

【0002】

可逆的なタンパク質のリン酸化は、真核細胞の細胞内での様々な現象を調節するために広く利用されている。一般的な哺乳動物細胞における活性化タンパク質の 10% 以上がリン酸化されていると推定される。キナーゼは、高エネルギーのリン酸基をアデノシン三リン酸(ATP)から標的タンパク質のヒドロキシアミノ酸残基であるセリン、トレオニン、またはチロシンに転移する。これとは対照的に、ホスファターゼはこれらのリン酸基を除去す

10

20

30

40

50

る。ホルモン、神経伝達物質、および成長因子、および分化因子を含む細胞内外シグナルは、キナーゼを活性化し、それによって最終的なエフェクタータンパク質のアクティベーターまたは細胞表面受容体が発生し、シグナル伝達経路に沿っても発生し得る。キナーゼのカスケードが生じ、またセカンドメッセンジャー分子に対し感受性のあるキナーゼも存在する。このシステムによって、弱いシグナル（例えば、わずかにしか存在しない成長因子分子）の増幅や、多くの弱いシグナルの全か無かの応答への統合が行われる。次に、ホスファターゼは細胞におけるリン酸化の程度の決定に必須であり、キナーゼと共に代謝酵素活性、細胞増殖、細胞の成長および分化、細胞接着、および細胞周期の進行などの重要な細胞プロセスを調節する。

【0003】

10

キナーゼ

キナーゼは最も大きな既知の酵素スーパーファミリーを構成し、標的分子が多種多様である。キナーゼは、高エネルギーリン酸基のリン酸供与体からリン酸受容体への移動を触媒する。ヌクレオチドは通常、これらの反応においてリン酸供与体として働き、ほとんどのキナーゼはアデノシン三リン酸（ATP）を利用する。リン酸受容体には様々な分子が可能であり、その中には、ヌクレオシド、ヌクレオチド、脂質、炭水化物、およびタンパク質が含まれる。タンパク質は、ヒドロキシアミノ酸においてリン酸化される。リン酸基が付加されると、受容体分子における局所電荷が変化し、それによって内部構造が変化し、分子間接触を引き起こし得る。可逆的なタンパク質リン酸化は、真核細胞におけるタンパク質活性の調節の主な方法である。一般に、タンパク質は、ホルモン、神経伝達物質、成長因子、および分化因子などの細胞外シグナルに応答したリン酸化により活性化される。活性化されたタンパク質は細胞の細胞応答を、細胞内シグナル伝達経路、並びにタンパク質リン酸化を調節するサイクリックヌクレオチド、カルシウム-カルモジュリン、イノシトール、および様々なマイトジェンなどのセカンドメッセンジャー分子によって開始させる。

20

【0004】

キナーゼは、解糖などの基本的な代謝プロセスから細胞周期の調節、分化、およびシグナル伝達カスケードによる細胞外環境との情報交換に至る細胞機能の全てに参与する。細胞内におけるタンパク質の不適切なリン酸化は、細胞周期の進行および細胞分化における変化に関連する。細胞周期における変化は、アポトーシスの誘発や癌の誘発に関連することが示されている。細胞分化における変化は、生殖系、免疫系、および骨格筋の、疾患や障害に関連することが示されている。

30

【0005】

プロテインキナーゼには2つのクラスがある。一方のクラスのプロテインチロシンキナーゼ（PTK）はチロシン残基をリン酸化し、他方のクラスであるプロテインセリン/トレオニンキナーゼ（STK）はセリン残基およびトレオニン残基をリン酸化する。ある種のPTKおよびSTKは、両方のファミリーの構造的な特徴を有し、チロシン残基およびセリン/トレオニン残基の両方に対して特異性（二重特異性）を有する。ほとんど全てのキナーゼは、特定の残基およびキナーゼファミリー特有の配列モチーフを含む保存された250~300のアミノ酸からなる触媒ドメインを含む。プロテインキナーゼ触媒ドメインは、更に11のサブドメインに分類することができる。N末端サブドメインI-IVは、ATP供与体分子に結合して方向性を与える2葉構造（two lobed structure）に折り畳まれ、サブドメインVは2葉にまたがる。C末端サブドメインVI-XIはタンパク質基質と結合し、リン酸をATPからチロシン残基、セリン残基、またはトレオニン残基のヒドロキシル基に移動させる。11のサブドメインのそれぞれは、サブドメインに特徴的なアミノ酸モチーフ或いは特定の触媒残基を含む。例えば、サブドメインIは、8個のアミノ酸からなる高グリシンATP結合共通モチーフを含み、サブドメインIIは、触媒活性の最大化に必要な重要なリシン残基を含み、サブドメインVIからIXは高度に保存された触媒中心を含む。PTKおよびSTKはまた、ヒドロキシアミノ酸特異性を付与し得るサブドメインVIおよびVIIIに固有の配列モチーフを含む。

40

50

【0006】

更に、キナーゼは、キナーゼドメイン内に含まれるか或いは隣接する、通常は5～100の範囲の残基からなる追加のアミノ酸配列によって分類され得る。これらの追加のアミノ酸配列はキナーゼ活性を調節し、基質特異性を決定する(Hardie, G.及びS. Hanks (1995) The Protein Kinase Facts Book, Vol 1, 17-20ページ Academic Press, San Diego CA. を参照)。具体的には、2つのプロテインキナーゼシグネチャ配列が、キナーゼドメインで同定された。第1のプロテインキナーゼシグネチャ配列は、ATP結合に関与するリシン残基活性部位を含み、第2のプロテインキナーゼシグネチャ配列は触媒活性に重要なアスパラギン酸残基を含む。分析するタンパク質がこの2つのプロテインキナーゼシグネチャを含む場合、そのタンパク質がプロテインキナーゼである確率は100%に近い(PROSITE: PDOC00100, November 1995)。

10

【0007】

プロテインチロシンキナーゼ

プロテインチロシンキナーゼ(PTK)は、膜貫通受容体PTKタンパク質または非膜貫通非受容体PTKタンパク質のいずれかに分類され得る。膜貫通チロシンキナーゼは、ほとんどの成長因子に対する受容体として機能する。成長因子が受容体チロシンキナーゼ(RTK)に結合し、それによって受容体が、自己(自己リン酸化)および特定の細胞内セカンドメッセンジャータンパク質をリン酸化する。受容体PTKに結合する成長因子(GF)は、上皮成長因子、血小板由来成長因子、繊維芽細胞成長因子、肝細胞成長因子、インスリンおよびインスリン様成長因子、神経成長因子、血管内皮成長因子、およびマクロファージコロニー刺激因子が含まれる。

20

【0008】

非膜貫通型非受容体PTKは、膜貫通領域を含まない代わりに細胞膜受容体の細胞質ドメインとシグナル伝達複合体を形成する。非受容体PTKを介して機能する受容体には、サイトカイン受容体およびホルモン受容体(成長ホルモンおよびプロラクチン)や、Tリンパ球およびBリンパ球における抗原特異的受容体が含まれる。

【0009】

多くのPTKは、PTK活性が正常な細胞制御を受けない癌細胞における腫瘍遺伝子産物として初めに同定された。実際に、知られている腫瘍遺伝子の約3分の1がPTKをコードする。更に、細胞形質転換(発癌)はチロシンリン酸化活性の上昇を伴う場合が多い(Charbonneau, H.及びTonks, N. K. (1992) *Annu. Rev. Cell Biol.* 8:463-93)。従って、PTK活性の調節は、ある種の癌を制御するための重要な手段となり得る。

30

【0010】

プロテインセリン/トレオニンキナーゼ

プロテインセリン/トレオニンキナーゼ(STK)は非膜貫通型タンパク質である。STKのサブクラスは、ERK(細胞外シグナル調節キナーゼ)またはMAP(マイトジェン活性化プロテインキナーゼ)として知られ、様々なホルモンや成長因子による細胞刺激によって活性化される。細胞刺激は、MEK(MAP/ERKキナーゼ)のリン酸化に導くシグナル伝達カスケードを誘導し、MEKのリン酸化の後に、セリンおよびトレオニンのリン酸化によりERKが活性化される。多数のタンパク質が活性化ERKに対する下流のエフェクターであることから、細胞増殖および分化の調節や、細胞骨格の調節に関与すると思われる。ERKの活性化は通常一過性であり、細胞は、そのダウンレギュレーション(下方制御)に関係する二重特異性ホスファターゼを有する。また様々な研究から、ERK活性の上昇がある種の癌に関連することが分かった。その他のSTKには、サイクリックAMP依存性プロテインキナーゼ(PKA)、カルシウム-カルモジュリン(CaM)依存性プロテインキナーゼ、およびマイトジェン活性化プロテインキナーゼ(MAP)などのセカンドメッセンジャー依存性プロテインキナーゼや、サイクリン依存性プロテインキナーゼ、チェックポイントおよび細胞周期キナーゼ、Numb関連キナーゼ(Nak)、ヒト融合(hFu)、増殖関連キナーゼ、5'-AMP-活性化プロテインキナーゼ、およびアポトーシスに関与するキナーゼが含まれる。

40

【0011】

50

MAPキナーゼのERKファミリーの1メンバーであるERK7は、ERK1とERK2へのモチーフ類似性を有する新規61kDaタンパク質であるが、ERK1及びERK2のように細胞外刺激によってまた共通活性化剤であるcJun N末端キナーゼ(JNK)及びp38キナーゼによっても活性化されない。ERK7はその核局在化及び、MAPキナーゼでは典型的なキナーゼドメインではなく、そのC末端テイルにより成長の阻害を調節する(Abe, M.K. (1999) Mol. Cell. Biol. 19:130113-12)。

【0012】

セカンドメッセンジャー依存性プロテインキナーゼは、サイクリックAMP(cAMP)、サイクリックGMP、イノシトール三リン酸、ホスファチジルイノシトール、3,4,5-三リン酸、サイクリックADPリボース、アラキドン酸、ジアシルグリセロールおよびカルシウム-カルモジュリンなどのセカンドメッセンジャーの効果を主に仲介する。PKAはホルモン誘導性細胞応答の仲介に関与し、ホルモン刺激に応答して細胞内で生成されるcAMPによって活性化される。cAMPは、研究した全ての動物細胞におけるホルモン作用の細胞内メディエーターである。ホルモン誘導性細胞応答には、甲状腺ホルモン分泌、コルチゾル分泌、プロゲステロン分泌、グリコーゲン分解、骨再吸収、心拍数の調節、および心筋収縮力の調節が含まれる。PKAは全ての動物細胞に見られ、これらほとんどの細胞におけるcAMPの効果に関係すると思われる。PKA発現の変化は、癌、甲状腺疾患、糖尿病、アテローム性動脈硬化症、および心血管疾患等様々な障害や疾患に関連する(Isselbacher, K.J.他. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, 416-431, 1887ページ)。

10

20

【0013】

カゼインキナーゼI (CKI) 遺伝子ファミリーは、セリン/トレオニンプロテインキナーゼの別のサブファミリーである。この引き続き拡大するキナーゼ群は、細胞代謝、DNAの複製、およびDNAの修復を含む様々な細胞プロセスおよび核プロセスの調節に関与する。CKI酵素は、真核細胞の膜、核、細胞質および細胞骨格や、哺乳動物細胞の紡錘体上に存在する(Fish, K. J.他(1995) J. Biol. Chem. 270:14875-14883.)。

【0014】

CKIファミリーメンバーの全ては、9~76のアミノ酸からなる短いアミノ末端ドメイン、284のアミノ酸からなる高度に保存されたキナーゼドメイン、および24~200を超える範囲のアミノ酸からなる可変カルボキシル末端ドメインを含む(Cegielska, A.他(1998) J. Biol. Chem. 273:1357-1364)。このCKIファミリーは高度に関連するタンパク質からなる。これは、様々な試料からのカゼインキナーゼIのイソ型の同定によって確認されている。 、 、 、 および の少なくとも5つの哺乳動物イソ型が存在する。Fishらが、ヒト胎盤cDNAライブラリからCKI を同定した。このCKI は416のアミノ酸からなる塩基性タンパク質であり、CKI に類似している。CKI が既知のCKI基質をリン酸化し、CKI特異的インヒビターであるCKI-7によって抑制されることが組み替え発現によって分かった。CKI のヒト遺伝子は、酵母CKI遺伝子座であるHRR250の欠失によって起こる成長が遅い表現型(slow-growth phenotype)を有する酵母を助けることができる(Fish 他 前出)。

30

【0015】

哺乳動物突然変異概日tauが、シリアンハムスターにおける概日リズムの周期を著しく短くするCKI の半優性染色体対立遺伝子であることが分かった。tau遺伝子座はカゼインキナーゼI によってコードされ、ショウジョウバエ概日遺伝子double-timeの相同体でもある。野生型およびtau突然変異CKI 酵素の両方の研究から、突然変異酵素はその最大速度が著しく低下し、自己リン酸化状態であることが分かった。更に、*in vitro*におけるCKI は哺乳動物PERIODタンパク質と相互作用する能力を有するが、突然変異酵素はPERIODをリン酸化する能力に欠ける。Lowrey他が、CKI が概日機構の中心をなす転写-翻訳系の自己調節ループ内の負のフィードバックシグナルを遅らせる重要な因子であると主張した。従って、CKI 酵素は、概日リズム、時差ぼけおよび睡眠、更に概日調節下のその他の生理的プロセスおよび代謝プロセスに影響を与える医薬組成物の理想的な候補である(Lo 50

40

50

wrey, P. L.他(2000) Science 288:483-491)。

【0016】

ホメオドメイン相互作用プロテインキナーゼ(HIPK)はセリン/スレオニンキナーゼであり、DYRKキナーゼサブファミリーの新しいメンバーである(Hofmann, T.G. 他(2000) Biochimie 82:1123-7).HIPKはホメオプロテインと相互作用するドメインから別れた、保存されたプロテインキナーゼドメインを含む。HIPKは核内キナーゼであり、またHIPK2は神経組織においてよく発現される(Kim, Y.H. 他(1998) J. Biol. Chem. 273:25875-25879、Wang, Y. 他(2001) Biochim. Biophys. Acta 1518:168-172)。HIPKはホメオドメイン転写因子の補抑制体として作用する。この補抑制体の活性は、細胞タンパク質機能の調節に重要なユビキチン化やリン酸化のような翻訳後修飾において見られる(Kim, Y.H. 他(1999) Proc. Nat. Acad. Sci. U.S.A. 96:12350-12355)。

10

【0017】

ショウジョウバエいぼ腫瘍抑制因子遺伝子の相同体であるヒトh-wartsタンパク質は、染色体6q24.25.1にマッピングする。それはセリン/スレオニンキナーゼドメインを有しており、間期細胞の中心体に局在する。有糸分裂に関与しており、分裂装置の成分として機能する(Nishiyama, Y. 他(1999) FEBS Lett. 459:159165)。

【0018】

カルシウム - カルモジュリン依存性プロテインキナーゼ

カルシウム - カルモジュリン (CaM) 依存性キナーゼは、平滑筋の収縮、グリコーゲン分解(ホスホリラーゼキナーゼ)、および神経伝達(CaMキナーゼIおよびCaMキナーゼII)の調節に関与する。CaM依存性プロテインキナーゼは、細胞内の遊離カルシウムの濃度に応答して細胞内カルシウム受容体であるカルモジュリンによって活性化される。多くのCaMキナーゼはまた、リン酸化によって活性化される。ある種のCaMキナーゼはまた、自己リン酸化またはその他の調節キナーゼによって活性化される。CaMキナーゼIは、神経伝達物質関連タンパク質であるシナプシンIおよびII、遺伝子転写調節因子であるCREB、および嚢胞性繊維コンダクタンス調節タンパク質であるCFTRを含む様々な基質をリン酸化する(Haribabu, B.他(1995) EMBO Journal 14:3679-3686)。また、CaMキナーゼIIは、様々な部位においてシナプシンをリン酸化し、チロシンヒドロキシラーゼのリン酸化および活性化によって脳におけるカテコールアミンの合成を調節する。CaMキナーゼIIはまた、チロシンヒドロキシラーゼおよびトリプトファンヒドロキシラーゼのリン酸化/活性化によってカテコールアミンおよびセロトニンの合成を調節する(Fujisawa, H. (1990) BioEssays 12:27-29)。カルモジュリン結合プロテインキナーゼ様タンパク質をコードするmRNAが哺乳動物前頭葉で多量に見つかった。このタンパク質は軸索および樹状突起の双方における小胞に結合し、主に出生後に蓄積される。このタンパク質のアミノ酸配列はCaM依存性STKに類似しており、このタンパク質はカルシウムの存在下でカルモジュリンと結合する(Godbout, M.他(1994) J. Neurosci. 14:1-13)。

20

30

【0019】

マイトジェン活性化プロテインキナーゼ

マイトジェン活性化プロテインキナーゼ(MAP)は、リン酸化カスケードによって細胞表面から核へのシグナル伝達を仲介する。MAPは、細胞内シグナル伝達経路を調節する別のSTKファミリーである。いくつかのサブグループが同定されており、それぞれが異なる基質特異性を有し、固有の細胞外刺激に応答する(Egan, S. E.及びWeinberg, R. A. (1993) Nature 365:781-783)。MAPキナーゼカスケードを構成する三つのキナーゼモジュールがあり、それらはMAPK(MAP)、MAPKキナーゼ(MAP2K、MAPKKまたはMKK)およびMKKキナーゼ(MAP3K、MAPKKKまたはMEKK)である(Wang, X.S. 他(1998) Biochem. Biophys. Res. Commun. 253:33-37)。細胞外信号制御キナーゼ(ERK)経路は、例えば、上皮細胞成長因子(EGF)、紫外線、高浸透圧性培養液、熱ショックまたは内毒素リポ多糖(LPS)のような成長因子やマイトジェン(分裂促進因子)によって活性化される。密接に関連しているが、異なる平行経路である、c-Jun N-末端キナーゼ(JNK)、即ちストレスによって活性化されるキナーゼ(SAPK)経路、およびP38キナーゼ経路は、ストレスによる刺激および腫瘍壊死因子(TN

40

50

F)やインターロイキン-1(IL-1)のような炎症誘発性サイトカインによって活性化される。MAPキナーゼ発現の変化は、癌、炎症、免疫疾患、および成長および発達に影響を及ぼす疾患を含む様々な疾患に關与する。MAPキナーゼシグナリング経路は哺乳類細胞および酵母に存在する。

【0020】

サイクリン依存性プロテインキナーゼ

サイクリン依存性プロテインキナーゼ(CDK)は、細胞周期を介して細胞の進行を調節するSTKである。細胞は、サイクリンと呼ばれる活性化タンパク質ファミリーの合成および分解による調節により、有糸分裂に入ったり有糸分裂から出たりする。小さな調節タンパク質であるサイクリンはCDKに結合してそのCDKを活性化させる。それによって、有糸分裂プロセスに關与する選択されたタンパク質がリン酸化され活性化される。CDKは、活性化するために多数の入力が必要であるという点でユニークである。サイクリンの結合に加えて、CDKの活性化には、CDKにおける特定のトレオニン残基のリン酸化および特定のチロシン残基の脱リン酸化が必要である。

10

【0021】

NIMA(有糸分裂に決して入らない)関連キナーゼ(Neks)は細胞周期に關連するSTKの別のファミリーである。CDKおよびNekの両方は、動物細胞における複製、成熟、微小管形成中心である中心体の分離に關与する(Fry, A. M.他(1998)EMBO J. 17:470-481)。

【0022】

チェックポイントおよび細胞周期キナーゼ

細胞分化のプロセスでは、細胞周期移行の順番およびタイミングは細胞周期チェックポイントの制御下にある。この細胞周期チェックポイントは、DNA複製および染色体分離などの重要なプロセスが正確に実行されるようにする。例えば放射線などによってDNAが損傷を受けた場合、チェックポイント経路が活性化され、修復の時間を与えるべく細胞周期が停止される。損傷が大きい場合には、アポトーシスが誘導される。このようなチェックポイントが存在しないと、損傷したDNAが異常な細胞に受け継がれ、癌などの増殖異常が引き起こされ得る。プロテインキナーゼは、このプロセスにおいて重要な役割を果たす。例えば、特定のキナーゼすなわちチェックポイントキナーゼ1(Chk1)が酵母および哺乳動物で同定された。このChk1は、酵母におけるDNA損傷によって活性化される。Chk1が活性化されると、G2/M移行において細胞が停止する(Sanchez, Y.他(1997)Science 277:1497-1501)。具体的には、Chk1が細胞分裂サイクルホスファターゼCDC25をリン酸化し、それによってサイクリン依存性キナーゼCdc2を脱リン酸化して活性化するCDC25の正常な機能が阻害される。Cdc2の活性化によって細胞が有糸分裂に入るのが調節される(Peng, C-Y他(1997)Science 277:1501-1505)。従って、Chk1の活性化によって、損傷した細胞が有糸分裂に入るのが阻止される。Chk1のようなチェックポイントキナーゼの欠損によっても、G2/Mのような他のチェックポイントにおいてDNAが損傷した細胞が停止できないことにより癌が引き起こされ得る。

20

30

【0023】

細胞増殖関連キナーゼ

細胞増殖関連キナーゼは、ヒト巨核球細胞における細胞周期および細胞増殖の調節に關与する血清/サイトカイン誘導性STKである(Li, B.他(1996)J. Biol. Chem. 271:19402-19408)。増殖関連キナーゼは、細胞分裂に關与するSTKのpolo(シヨウジョウバエpolo遺伝子に由来する)ファミリーに關連する。増殖関連キナーゼは肺腫瘍組織においてダウンレギュレートされ、プロトオンコジーンであると思われる。プロトオンコジーンが正常な組織において制御を受けないで発現すると、正常な組織が癌化し得る。

40

【0024】

5'-AMP-活性化プロテインキナーゼ

あるリガンド活性化STKプロテインキナーゼは、5'-AMP-活性化プロテインキナーゼ(AMPK)である(Gao, G.他(1996)J. Biol. Chem. 271:8675-8681)。哺乳動物AMPKは、酵素であるアセチル-CoAカルボキシラーゼおよびヒドロキシメチルグルタリル-CoAレダクターゼ

50

のリン酸化による脂肪酸およびステロールの合成の調節因子であって、熱ショックやグルコースおよびATPの枯渇などの細胞内ストレスに対するこれらの経路の応答を仲介する。AMPKは、触媒サブユニットと、このサブユニットの活性を調節すると考えられている2つの非触媒サブユニットおよびサブユニットからなるヘテロ三量体複合体である。AMPKのサブユニットは、脳、心臓、脾臓、および肺などの非脂肪性組織において予想以上に広く分布している。この分布から脂質の代謝の調節以外にもある役割を果たしていると考えられる。

【0025】

アポトーシスにおけるキナーゼ

アポトーシスは、細胞死を導く高度に制御されたシグナル伝達経路であって、組織の発達および恒常性に極めて重要な役割を果たしている。このプロセスの調節不全は、自己免疫疾患、神経変性疾患、および癌を含む様々な疾患の病原に関連する。様々なSTKがこのプロセスにおいて重要な役割を果たす。ZIPキナーゼは、N末端プロテインキナーゼドメインに加えてC末端ロイジンジッパードメインを含むSTKである。このC末端ドメインは、ホモ二量体化およびキナーゼの活性化、ならびに転写因子のサイクリックAMP応答性エレメント結合タンパク質(ATF/CREB)ファミリーのメンバーである活性化転写因子ATF4などの転写因子との相互作用を仲介する(Sanjo, H.他(1998) J. Biol. Chem, 273:29066-29071)。DRAK1およびDRAK2は、死関連プロテインキナーゼ(DAPキナーゼ)と相同性を共有するSTKであって、インターフェロン誘導性アポトーシスにおいて機能することが知られている(Sanjo他前出)。ZIPキナーゼと同様に、DAPキナーゼも、N末端キナーゼドメインに加えてアンキリンリピート型のC末端タンパク質-タンパク質相互作用ドメインを含む。ZIP、DAP、およびDRAKキナーゼは、NIH3T3細胞に形質転換されるとアポトーシスに関連する形態変化を誘導する(Sanjo他, 前出)。しかしながら、これらのタンパク質のN末端キナーゼ触媒ドメインか或いはC末端ドメインのいずれかが欠失すると、アポトーシス活性がなくなる。このことから、キナーゼ活性に加えてC末端ドメインにおける活性も、アポトーシスに必要であり、調節因子または特定の基質と相互作用するドメインの可能性がある。

【0026】

RICKは、死受容体CD95を含む特定のアポトーシス経路を仲介するとして近年同定された別のSTKである(Inohara, N.他(1998) J. Biol. Chem. 273:12296-12300)。CD95は腫瘍壊死因子受容体スーパーファミリーのメンバーであって、免疫系の調節および恒常性において重要な役割を果たす(Nagata, S. (1997) Cell 88:355-365)。CD95受容体シグナル伝達経路は、様々な細胞内分子の受容体複合体への補充、およびそれに続くリガンド結合を伴う。このプロセスは、システインプロテアーゼカスパーゼ-8の補充を伴い、それによってカスパーゼカスケードが活性化され細胞死が起こる。RICKは、カスパーゼ様ドメインと相互作用するC末端「カスパーゼ動員」ドメインおよびN末端キナーゼ触媒ドメインとからなる。このことから、RICKがカスパーゼ-8の動員においてある役割を果たしていると思われる。この解釈は、ヒト293T細胞におけるRICKの発現がカスパーゼ-8の活性化を促進し、CD95アポトーシス経路に関与する様々なタンパク質によるアポトーシスの誘導を増強するという事実によって補強される(Inohara他前出)。

【0027】

ミトコンドリアプロテインキナーゼ

原核生物ヒスチジンプロテインキナーゼに対する配列によって関連する真核生物キナーゼの新規のクラスであるミトコンドリアプロテインキナーゼ(MPK)は、他の真核生物プロテインキナーゼとは配列類似性を有していないようである。これらのプロテインキナーゼは、ミトコンドリアマトリックス空間のみに存在し、原始的真核細胞によってエンドサイトーシスによって取り込まれた呼吸依存性細菌に存在した遺伝子が進化したものと思われる。MPKは、分岐鎖ケト酸デヒドロゲナーゼおよびピルビン酸デヒドロゲナーゼ複合体のリン酸化および不活化に関係する(Harris, R. A.他(1995) Adv. Enzyme Regul. 34:147-162)。5つのMPKが同定された。4つのメンバーはピルビン酸デヒドロゲナーゼイソ

10

20

30

40

50

酵素に対応し、解糖とクエン酸回路との中間において重要な調節酵素であるピルビン酸デヒドロゲナーゼ複合体の活性を調節する。5番目のメンバーは分岐鎖ケト酸デヒドロゲナーゼキナーゼに対応し、分岐鎖アミノ酸の除去のための経路の調節に重要である(Harris, R.A.他(1997) Adv. Enzyme Regul. 37:271-293)。飢餓状態および糖尿病状態において、ラットの肝臓、心臓および筋肉におけるピルビン酸デヒドロゲナーゼキナーゼの活性が著しく上昇することが知られている。この活性の上昇が、両方の病態におけるピルビン酸デヒドロゲナーゼ複合体のリン酸化および不活化、並びに糖新生のためのピルビン酸および乳酸の保存に貢献する(Harris他 前出)。

【0028】

(非タンパク質基質に対するキナーゼ)

脂質キナーゼおよびイノシトールキナーゼ

脂質キナーゼは、脂質ヘッドグループのヒドロキシル残基をリン酸化する。ホスファチジルイノシトール(PI)のリン酸化に参与するキナーゼのファミリーが記載されており、それぞれのメンバーがイノシトール環の特定の炭素をリン酸化する(Leevers, S.他(1999) Curr. Opin. Cell. Biol. 11:219-225)。ホスファチジルイノシトールのリン酸化は、プロテインキナーゼCシグナル伝達経路の活性化に参与する。イノシトールホスホリピッド(ホスホイノシチド)細胞内シグナル伝達経路は、細胞膜においてシグナル伝達分子がGタンパク質結合受容体に結合することによって始まる。これによって、イノシトールキナーゼによる細胞膜の内側のホスファチジルイノシトール(PI)残基のリン酸化が起こり、それによってPI残基が二リン酸状態(PIP_2)に変換される。次に PIP_2 がイノシトール三リン酸(IP_3)およびジアシルグリセロールに切断される。これらの2つの生成物は、シグナル伝達経路を分けるためのメディエーターとして作用する。これらの経路によって仲介される細胞応答には、パソプレッシンに回答する肝臓におけるグリコーゲン分解、アセチルコリンに回答する平滑筋の収縮、およびトロンピン誘導性血小板凝集がある。

【0029】

PIおよびその誘導体のD3部位をリン酸化するPI3-キナーゼ(PI3K)は、細胞増殖、分化、および代謝に参与する成長因子シグナルカスケードにおいて重要な役割を果たす。PI3Kは、アダプターサブユニットおよび触媒サブユニットからなるヘテロ二量体である。このアダプターサブユニットは足場タンパク質として作用し、特定のチロシン-リン酸化タンパク質、脂質成分、およびその他の細胞質因子と相互作用する。アダプターサブユニットがインスリン応答性基質(IRS)-1などのチロシンリン酸化標的と結合すると、触媒サブユニットが活性化され、PI(4,5)二リン酸(PIP_2)をPI(3,4,5)三リン酸(PIP_3)に変換する。次に PIP_3 が、PKA、プロテインキナーゼB(PKB)、プロテインキナーゼC(PKC)、グリコーゲンシンターゼキナーゼ(GSK)-3、およびP70リボソームs6キナーゼを含むその他の幾つかのタンパク質を活性化する。PI3Kもまた、細胞骨格形成タンパク質であるRac、rhoおよびcdc42と直接相互作用する(Shepherd, P. R.他(1998) Biochem. J. 333:471-490)。obeseマウスおよびfatマウスなどの糖尿病動物モデルは、PI3kアダプターサブユニットのレベルが変更されている。アダプターサブユニットにおける特定の変異が、インスリン抵抗性のデンマーク人集団に見られることから、PI3Kが2型糖尿病においてある役割を果たしていると思われる(Shepherd、前出)。

【0030】

脂質キナーゼリン酸化活性の例は、D-エリトロ-スフィンゴシンのスフィンゴ脂質代謝産物であるスフィンゴシン-1-リン酸(SPP)へのリン酸化である。SPPは、細胞外作用および細胞内作用の両方を持つ新規の脂質セカンドメッセンジャーであることが分かった(Kohama, T.他(1998) J. Biol. Chem. 273:23722-23728)。細胞外では、SPPはGタンパク質結合受容体EDG-1(内皮由来Gタンパク質共役受容体)のリガンドである。細胞内では、SPPは、細胞の成長、生存、運動性、および細胞骨格の変化を調節する。SPPのレベルは、D-エリトロ-スフィンゴシンを特異的にリン酸化してSPPにするスフィンゴシンキナーゼによって調節される。細胞シグナル伝達におけるスフィンゴシンキナーゼの重要性は、血小板由来成長因子(PDGF)、神経成長因子、およびプロテインキナーゼCの活性化を含む様々

10

20

30

40

50

な刺激が、スフィンゴシンキナーゼの活性化によりSPPの細胞内レベルを上昇させるという事実、および酵素の競合阻害剤がPDGFによって誘導された細胞増殖を選択的に阻害するという事実から証明された (Kohama他前出)。

【0031】

プリンヌクレオチドキナーゼ

プリンヌクレオチドキナーゼであるアデニル酸キナーゼ (ATP: AMPホスホトランスフェラーゼ、もしくはAdK) およびグアニル酸キナーゼ (ATP: GMPホスホトランスフェラーゼ、もしくはGuK) が、ヌクレオチド代謝において重要な役割を果たし、ATPおよびGTPの合成および細胞内レベルの調節のそれぞれにおいて重要である。これらの2つの分子は、成長している細胞におけるDNA合成およびRNA合成の前駆物質であって、細胞における主な生化学エネルギーの主な供給源 (ATP) であり、シグナル伝達経路 (GTP) を提供する。これらの2つの分子の合成における様々なステップを妨げることが、癌および抗ウイルス治療のための多くの抗増殖剤の原理である (Pillwein, K. 他, (1990) *Cancer Res.* 50:1576-1579)。

10

【0032】

AdKはほとんどの細胞型に見られ、高い速度でATPを合成する骨格筋などの細胞に特に多く存在する。これらの細胞において、AdKは、細胞レベル下構造であるミトコンドリアおよび筋原繊維と物理的に関連する。ミトコンドリアはエネルギーを生産し、筋原繊維はそのエネルギーを利用する。近年の研究により、ATPを生成する代謝プロセスからATPを消費する細胞成分への高エネルギーのホスホリルの転移においてAdKが重要な役割を果たしていることが実証された (Zeleznikar, R. J.他 (1995) *J. Biol. Chem.* 270:7311-7319)。従って、AdKは、細胞、特に癌細胞などの増殖や代謝速度が速い細胞におけるエネルギー生産の維持において中心的な役割を演じると考えられ、ある種の癌の治療のためにその活性を阻害するための候補を提供し得る。別法では、AdK活性の低下が、心不全や呼吸器不全を引き起こす様々な代謝筋エネルギー疾患の原因と考えられ、AdKの活性を上昇させて治療できるであろう。

20

【0033】

RNAおよびDNA合成のためのGTPの合成における重要なステップを提供するのに加えて、GuKはまた、GDPおよびGTPの調節によって細胞のシグナル伝達経路において重要な役割を果たす。具体的には、膜関連Gタンパク質に結合するGTPは細胞受容体の活性化を仲介し、アデニルシクラーゼが細胞内で活性化され、セカンドメッセンジャーであるサイクリックAMPが生産される。Gタンパク質に結合するGDPはこれらのプロセスを阻害する。また、GDPおよびGTPのレベルによって、細胞増殖の調節および発癌に関与するとして知られるp21^{ras}などのある種の腫瘍タンパク質の活性が調節される (Bos, J. L. (1989) *Cancer Res.* 49:4682-4689)。GuKの抑制によって起こるGDPに対するGTPの比率が高いと、p21^{ras}が活性化され発癌が促進される。GDPのレベルを上昇させGDPに対するGTPのレベルを低下させるためにGuKの活性を高めることが、発癌を抑制する治療となり得る。

30

【0034】

GuKは、ヘルペスウイルス感染の治療に有用なある種の抗ウイルス剤のリン酸化および活性化における重要な酵素である。これらの薬剤には、グアニン相同体アシクロビルおよびbuciclovirが含まれる (Miller, W. H.及びMiller R. L. (1980) *J. Biol. Chem.* 255:7204-7207, Stenberg, K.他 (1986) *J. Biol. Chem.* 261:2134-2139)。感染細胞においてGuKの活性を高めることが、これらの薬剤の効果を促進させる治療方法となり得るため、薬剤の服用を減らすことが可能であろう。

40

【0035】

ピリミジンキナーゼ

ピリミジンキナーゼは、デオキシチジンキナーゼ、およびチミジンキナーゼ1および2を含む。デオキシチジンキナーゼは核に存在し、チミジンキナーゼ1および2は細胞質に見られる (Johansson, M.他 (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94:11941-11945)。ピリミジンキナーゼによるデオキシリボヌクレオシドのリン酸化によって、DNA前駆物質

50

の *de novo* 合成のための代替の経路が提供される。リン酸化におけるピリミジンキナーゼの役割はプリンキナーゼと同様に、化学療法に重要な幾つかのヌクレオシド類似体の活性化に重要である (Arner E. S. および Eriksson, S. (1995) *Pharmacol. Ther.* 67:155-186)。

【0036】

ホスファターゼ

一般的に、プロテインホスファターゼは、ホスホアミノ酸基質選択性に基づいて、セリン/トレオニン特異性またはチロシン特異性のいずれかによって特徴づけられる。しかしながら、或る種のホスファターゼ (DSP: 二重特異性ホスファターゼ) は、リン酸化したチロシン、セリン、またはトレオニン残基に作用し得る。プロテインセリン/トレオニンホスファターゼ (PSP) は、細胞内での多くの cAMP 仲介性ホルモン応答の重要な制御因子である。プロテインチロシンホスファターゼ (PTP) は、細胞周期および細胞内シグナル伝達プロセスにおいて重要な役割を果たす。別のファミリーのホスファターゼは、酸性ホスファターゼ、またはヒスチジン酸性ホスファターゼ (HAP) ファミリーで、そのメンバーは酸性の pH 条件下でリン酸エステルを加水分解する。

10

【0037】

PSP はサイトゾル、核、およびミトコンドリアに見られ、多くの組織、特に脳において細胞骨格および膜構造に結合している。或る種の PSP は、活性のために Ca^{2+} または Mn^{2+} などの 2 価の陽イオンを必要とする。PSP は、糖代謝、筋収縮、蛋白合成、T 細胞機能、神経作用、卵母細胞成熟、および肝代謝において重要な役割を果たす (Cohen, P. (1989) *Annu. Rev. Biochem.* 58:453-508 の概説を参照)。PSP は 2 つのクラスに分類することができる。PPP クラスには、PP1、PP2A、PP2B/カルシニューリン、PP4、PP5、PP6、および PP7 が含まれる。このクラスのメンバーは、高度に保存されたシグネチャ配列を有する相同な触媒サブユニットに、1 つ以上の調節サブユニットが組み合わさって構成されている (PROSITE PDOC00115)。更に、足場分子および固着分子との相互作用によって、PSP の細胞内局在性および基質特異性が決定される。PPM クラスは、PP2C に密接に関連した幾つかのアイソフォームから成り、PPP クラスにとは進化的な関連性が存在しない。

20

【0038】

PP1 は、サイクリック AMP 依存性プロテインキナーゼ (PKA) によってリン酸化した多くのタンパク質を脱リン酸化し、細胞における多くの cAMP 仲介性ホルモン応答の重要な制御因子である。幾つかのアイソフォームが同定されているが、 α 型および β 型は同じ遺伝子の選択的スプライシングによって生成される。PP1 に関して、遍在性の標的タンパク質および組織特異的標的タンパク質が同定された。脳において、ドーパミンおよび 32kDa のアデノシン 3', 5'-リン酸調節性ホスホプロテイン (DARPP-32) による PP1 活性化の阻害が、新線条体ニューロンにおける正常なドーパミン応答に必要である (Price, N. E. 及び M. C. Mumby (1999) *Curr. Op. Neurobiol.* 9:336-342 の概説を参照)。PP1 および PP2A が微少血管内皮細胞の細胞運動を制限することが分かっており、脈管形成の阻害に PSP がある役割を果たしていると思われる (Gabel, S. 他 (1999) *Otolaryngol. Head Neck Surg.* 121:463-468)。

30

【0039】

PP2A は、主なセリン/トレオニンホスファターゼである。コア PP2A 酵素は、更なる調節サブユニット (B) を動員する 65 kDa の足場サブユニット A) と結合した 1 つの 36 kDa の触媒サブユニット (C) から成る。B サブユニットをコードする 3 つの遺伝子ファミリー (PR55、PR61、および PR72) が知られており、そのそれぞれが複数のアイソフォームを含み、更なるファミリーが存在し得る (Millward, T. A. 他 (1999) *Trends Biosci.* 24:186-191)。これらの「B 型」サブユニットは細胞型特異的および組織特異的であって、基質特異性、酵素活性、およびホロ酵素の細胞内局在性を決定する。PR55 ファミリーは高度に保存されていて、保存されたモチーフを有する (PROSITE PDOC00785)。PR55 によって、マイトジェン活性化プロテインキナーゼ (MAPK) および MAPK キナーゼ (MEK) に対する PP2A 活性が上昇する。PP2A は、MAPK 活性化部位を脱リン酸化し、細胞の有糸分裂への移行を阻害する。

40

50

CKIIキナーゼ触媒サブユニット、ポリオーマウイルスの中型T抗原および小型T抗原、およびSV40小型t抗原を含む幾つかのタンパク質が、PP2Aのコア酵素結合についてPR55と競合し得る。ウイルスはこのメカニズムを利用してPP2A活性を徴用し、細胞の細胞周期進行を刺激する (Pallas, D. C.他(1992) J. Virol. 66:886-893)。MAPキナーゼの発現の変化はまた、癌、炎症疾患、免疫系疾患、および成長および発達に影響を与える疾患を含む様々な症状に関与すると考えられている。実際、PP2Aは *in vitro* で30種以上のプロテインキナーゼを脱リン酸化してその活性を調節する。また、その他の研究結果から、PKB、PKC、カルモジュリン依存性キナーゼ、ERKファミリーMAPキナーゼ、サイクリン依存性キナーゼ、およびI Bキナーゼなどのキナーゼについても *in vivo* で同様の結果が得られることが示された (前出のMillward他に概説)。PP2A自体が、CKIキナーゼおよびCKIIキナーゼの基質であり、ポリカチオン巨大分子によって刺激され得る。PP2A様ホスファターゼは、哺乳動物サイクリンAおよびBのG1期破壊を維持するために必要である (Bastians, H.他(1999) Mol. Biol. Cell 10:3927-3941)。PP2Aは脳において主に作用し、神経フィラメントの安定性および正常な神経機能の調節、特に微少管結合タンパク質tauのリン酸化に関係すると考えられる。tauの過剰なリン酸化がアルツハイマー病に見られる神経変性を引き起こすと提案された (前出のPrice およびMumbyの概説を参照)。

10

【0040】

PP2Bすなわちカルシニューリンは、Ca²⁺活性化二量体ホスファターゼであって、特に脳に多量に存在する。PP2Bは、触媒サブユニットおよび調節サブユニットから成り、カルシウム/カルモジュリン複合体の結合によって活性化される。カルシニューリンは、免疫抑制薬シクロスポリンおよびFK506の標的である。その他の細胞因子と共にこれらの薬剤は、カルシニューリンと相互作用してホスファターゼ活性を阻害する。T細胞では、この作用によって、転写因子のNF-ATファミリーのカルシウム依存性活性化が阻害され、免疫抑制が起こる。このファミリーは広く分布し、カルシニューリンがその他の組織においても遺伝子の発現を調節する可能性が高い。ニューロンでは、カルシニューリンは、神経伝達物質放出の阻害から後シナプスNMDA受容体結合カルシウムチャネルの脱感作、および長期の記憶に至るまでの機能を調節する (PriceおよびMumby, 前出)。

20

【0041】

PPPクラスのその他のメンバーが近年同定された (Cohen, P. T. (1997) Trends Biochem. Sci 22:245-25)。その1つであるPP5は、テトラトリコペプチド リポートを有する調節ドメインを含む。PP5は、*in vitro* で陰イオン性リン脂質および多価不飽和脂肪酸によって活性化され、心房性ナトリウム利尿ペプチドまたはステロイドホルモンによって調節されるシグナル伝達経路を含む幾つかのシグナル伝達経路に関与すると思われる (Andreeva, A. V.およびM. A. Kutuzov (1999) Cell Signal. 11:555-562の概説を参照)。

30

【0042】

PP2Cは、広い基質特異性を有する最大42 kDaの単量体であって、その活性が二価の陽イオン (主にMn²⁺またはMg²⁺) に依存する。PP2Cタンパク質は、陽イオン結合に関係するアスパラギン酸残基を含む不変なDGHモチーフを有する保存されたN末端領域を共有する (PROSITE PDOC00792)。標的タンパク質とPP2Cの作用を調節するメカニズムは、同定されていない。PP2Cは、ストレス応答性p38およびJunキナーゼ(JNK)経路を阻害することが分かっている (Takekawa, M.他(1998) EMBO J. 17:4744-4752)。

40

【0043】

PSPとは対照的に、チロシン特異的ホスファターゼ(PTP)は、大きさおよび構造が様々であって (20 kDa ~ 100 kDa以上) 一般に単量体のタンパク質であり、主に細胞膜を通過するシグナル伝達において機能する。PTPは、可溶性ホスファターゼか、或いはホスファターゼドメインを含む膜貫通受容体タンパク質のいずれかに分類される。全てのPTPは、活性部位を含む約300のアミノ酸の保存された触媒ドメインを共有する。この活性部位の共通配列は、触媒作用中にリン酸部分に求核攻撃を加えるシステイン残基を含む (Neel, B.G. およびN.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204)。受容体PTPは、可変長のN末端細胞外ドメインと、膜貫通領域と、通常は2つの触媒ドメインを有する細胞質内

50

領域とから成る。第一の触媒ドメインのみが酵素活性を有するよう見えるが、第2の触媒ドメインは第1の触媒ドメインの基質特異性に影響を与えられる。幾つかの受容体PTPの細胞外ドメインは、フィブロネクチン様リピート、免疫グロブリン様ドメイン、MAMドメイン（接着機能を有する可能性が高い細胞外モチーフ）、または炭酸脱水素酵素様ドメイン（PROSITE PDOC 00323）を含む。このような様々な構造モチーフによって、PTPの大きさおよび特異性が多様になっている。

【0044】

PTPは、細胞接着、リンパ球活性化、および細胞増殖などの生体プロセスにおいて重要な役割を果たす。PTP μ およびPTP δ は、細胞-細胞接触に関与し、カドヘリン/カテニン機能を調節するであろう。幾つかのPTPが細胞の進展、局所接着、および細胞運動に影響を与えているが、そのほとんどがインテグリン/チロシンキナーゼシグナル伝達経路によって行なわれる（前出のNeelおよびTonksを参照）。CD45ホスファターゼは、シグナル伝達およびリンパ球活性化を調節する（Ledbetter, J. A.他(1988) Proc. Natl. Acad. Sci. USA 85:8628-8632）。Src-ホモロジー-2ドメインを有する可溶性PTP(SHP)が同定された。このことは、これらの分子が受容体チロシンキナーゼと相互作用することを示唆するものである。SHP-1は、造血細胞におけるJanusファミリーPTKの調節によって、およびT細胞受容体およびc-Kitによるシグナル伝達によって、サイトカイン受容体シグナル伝達を調節する（前出のNeelおよびTonksの概説を参照）。M期誘導ホスファターゼはPTK CDC2を脱リン酸化して活性化し、有糸分裂の誘導において重要な役割を果たし、細胞分裂を導く（Sadhu, K.他(1990) Proc. Natl. Acad. Sci. USA 87:5139-5143）。加えて、少なくとも8種のPTPをコードする遺伝子が、リンパ腫、小細胞肺癌、白血病、腺癌、および神経芽腫を含む様々な腫瘍症状において転座された染色体領域或いは再構成された染色体領域にマッピングされた（Charbonneau, H.およびN. K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493の概説を参照）。PTP酵素の活性化部位は、MTM1遺伝子ファミリーのコンセンサス配列を含む。MTM1遺伝子は、X染色体に関連した劣性のミオチューブ様ミオパシーに関与している。この疾患は、Xq28に関連される先天的な筋疾患である（Kioschis, P. 他 (1998) Genomics 54:256-266）。多くのPTKが発癌遺伝子によってコードされ、発癌にチロシンリン酸化活性の上昇を伴う場合が多いことは良く知られている。従って、PTPは、細胞におけるチロシンリン酸化のレベルを調節して細胞形質転換および様々な癌の成長を防止或いは逆転し得る可能性がある。この考えは、PTPの過剰な発現により細胞の形質転換を抑制でき、PTPの特異的阻害により細胞形質転換を促進できるという研究結果によって支持される（Charbonneau および Tonks, 前出）。

【0045】

二重特異性ホスファターゼ(DSP)は、PSPよりもPTPに構造的に類似している。DSPは、追加の7個のアミノ酸残基を有する拡張されたPTP活性化部位モチーフを有する。DSPは、主に細胞増殖に関与しており、細胞周期調節因子であるcdc25A、cdc25B、およびcdc25Cが含まれる。ホスファターゼDUSP1およびDUSP2は、チロシン残基およびトレオニン残基の両方においてMAPKファミリーメンバーERK（細胞外シグナル調節キナーゼ）、JNK(c-Jun N-末端キナーゼ)、およびp38を不活化する（PROSITE PDOC 00323, 前出）。活性化された状態では、これらのキナーゼは、ニューロンの分化、増殖、癌への形質転換、血小板凝集、およびアポトーシスに関与する。従って、DSPはこれらのプロセスの適切な調節に必要な（Muda, M.他(1996) J. Biol. Chem. 271:27205-27208）。腫瘍抑制因子PTENは、脂質ホスファターゼ活性も示すDSPである。PTENは、細胞外マトリックスとの相互作用を負に調節してアポトーシス感受性を維持すると思われる。PTENは、血管形成の防止に関係する（Giri, D.およびM. Ittmann (1999) Hum. Pathol. 30:419-424）。PTENの発現の異常は、様々な癌に関連する（Tamura, M.他(1999) J. Natl. Cancer Inst. 91:1820-1828の概説を参照）。

【0046】

ヒスチジン酸ホスファターゼ（HAP、EXPASY EC 3.1.3.2）もまた、酸性ホスファターゼとして知られ、低いpH下で、アルキル基、アリール基、アシル基のオルトリン酸モノエステ

ル、およびリン酸化されたタンパク質を含む広範囲な基質を加水分解する。HAPは、2つの領域で保存された配列を共有し、どちらも触媒作用に関与するヒスチジン残基のあたりに位置する。HAPファミリーのメンバーには、リソソーム酸性ホスファターゼ (LAP) と前立腺酸性ホスファターゼ (PAP) が含まれ、両者ともL-酒石酸塩による抑制に感受性を示す (PROSITE PDOC00538)。

【0047】

エンドソーム/リソゾーム内のオルトリン酸モノエステルであるLAPは、管理を行っている遺伝子で、その酵素活性は試験されたすべての組織で検知されている (Geier, C.他. (1989) Eur. J. Biochem. 183:611616)。LAP欠損マウスは、進行性の骨格疾患をもち、全身性発作を起こしやすくなる (Saftig, P.他 (1997) J. Biol. Chem. 272:18628-18635)。LAP欠損患者には次の臨床的特徴があることがわかっている。断続的に起こる嘔吐、緊張低下、無気力、強直性発作、末梢出血、発作、幼児期での死亡 (Online Mendelian Inheritance in Man (OMIM) *200950)。

【0048】

前立腺によって生成される前立腺上皮に特異的な分化抗原である、PAPは、前立腺癌を診断するために用いられてきた。前立腺癌内では、PAPの酵素活性が正常又は良性の前立腺肥大細胞に比べると減少していることがわかっている (Foti, A.G.他 (1977) Cancer Res. 37:41204124)。分泌されるタイプと細胞内に留まるタイプの二種類のPAPが同定されている。成熟した分泌されるPAPは精液に検出され、約100Kダルトンの分子量をもつ、グリコシル化されたホモ二量体として活性をもつ。細胞内のPAPは、内在的なチロシン磷酸化タンパク質ホスファターゼ活性を示し、前立腺細胞の成長の調節に関与することが知られている (Meng, T.C.及びM.F. Lin (1998) J. Biol. Chem. 34:2209622104)。

【0049】

イノシトールリン酸ホスファターゼであるシナプトジャニン₁は、イノシトールリン酸をイノシトール環の3、4、5の位置で脱リン酸する。シナプトジャニン₁は神経末端でクラスリンで覆われたエンドソーム中間物に見つけられるシナプス前部の主要タンパク質であり、クラスリンコートに関与するタンパク質であるEPS15と結合する。この結合はシナプトジャニン₁ - 170のC末端領域 (Asp - Pro - Pheのアミノ酸の3回繰り返し) で媒介される。更に、この3残基繰り返しは、EPS15のEHドメインの結合部位であることがわかっている (Haffner, C. 他 (1997) FEBS Lett. 419:175-180)。加えて、シナプトジャニン₁はエンドソームタンパク質と原形質膜との相互作用を調節し、シナプス小胞のリサイクリングに関与している可能性もある (Brodin, L. 他 (2000) Curr. Opin. Neurobiol. 10:312-320)。シナプトジャニン₁ 遺伝子 (Synj1) の標的破壊を受けたマウスの研究は、野生型マウスに見られるより効果的にエンドソーム小胞のコートが形成されていることを支持しており、Synj1が膜とコートタンパク質の相互作用の負調整として機能していることを示唆する。これらの発見は、シナプス小胞リサイクリングでイノシトールリン酸代謝が重要な役割を果たしている遺伝的証拠を提供する (Cremona, O. 他 (1999) Cell 99:179-188)。

【0050】

新規のキナーゼおよびホスファターゼ、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、心血管疾患、免疫系の疾患、神経の疾患、成長および発達に影響を及ぼす疾患、脂質異常、細胞増殖異常及び癌の診断・治療・予防に関する。本発明はさらに、キナーゼ及びホスホタ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。

【発明の開示】

【発明の効果】

【0051】

本発明は、総称して「KPP」、個別にはそれぞれ「KPP-1」、「KPP-2」、「KPP-3」、「KPP-4」、「KPP-5」、「KPP-6」、「KPP-7」、「KPP-8」と呼ぶキナーゼおよびホスファターゼである精製されたポリペプチドを提供する。或る実施態様において本発明は、(a)

SEQ ID NO:1-8からなる群から選択したアミノ酸配列からなるポリペプチド、(b) SEQ ID NO:1-8を有する群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を含むポリペプチド、(c) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片からなる群から選択した単離されたポリペプチドを提供する。一実施態様では、SEQ ID NO:1-8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。

【0052】

また、本発明は(a) SEQ ID NO:1-8からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-8からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を持つ或る天然アミノ酸配列を有するポリペプチド、(c) SEQ ID NO:1-8からなる群から選択した或るアミノ酸配列を持つポリペプチドの生物学的活性断片、または(d) SEQ ID NO:1-8からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片からなる群から選択した或る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。一実施態様では、該ポリヌクレオチドは、SEQ ID NO:1-8からなる群から選択した或るポリペプチドをコードする。別の実施態様では、ポリヌクレオチドはSEQ ID NO:9-16からなる群から選択される。

10

【0053】

本発明は更に、(a) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチド、(b) SEQ ID NO:1-8からなる群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を含むポリペプチド、(c) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片からなる群から選択したポリペプチドをコードするようなポリヌクレオチドと機能的に連結したプロモーター配列を有する組換えポリヌクレオチドを提供する。一実施態様では、本発明は組換えポリヌクレオチドを用いて形質転換した細胞を提供する。別の実施態様では、本発明は組換えポリヌクレオチドを含む遺伝形質転換体を提供する。

20

【0054】

また、本発明は、(a) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチド、(b) SEQ ID NO:1-8からなる群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を有するポリペプチド、(c) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片から構成される群から選択したポリペプチドを製造する方法を提供する。製造方法は、(a) 組換えポリヌクレオチドを用いて形質転換した細胞をポリペプチドの発現に適した条件下で培養する過程と、(b) そのように発現したポリペプチドを回収する過程とを有し、組換えポリヌクレオチドは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連結したプロモーター配列を有する。

30

【0055】

本発明は更に、(a) SEQ ID NO:1-8からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-8からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する或る天然アミノ酸配列を持つポリペプチド、(c) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を持つポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-8からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択した或るポリペプチドに特異結合する、単離された抗体を提供する。

40

【0056】

本発明は更に、(a) SEQ ID NO:9-16からなる群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(b) SEQ ID NO:9-16からなる群から選択し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

50

チド、(c)(a)のポリヌクレオチドに相補的なポリヌクレオチド、(d)(b)のポリヌクレオチドに相補的なポリヌクレオチド、および(e)(a)~(d)のRNA等価物からな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。一実施態様では、ポリヌクレオチドは少なくとも60の連続したヌクレオチドを有する。

【0057】

本発明は更に、サンプル中の標的ポリヌクレオチドを検出する方法を提供する。ここで、標的ポリヌクレオチドは(a)SEQ ID NO:9-16を有する群から選択したポリヌクレオチド配列を有するポリヌクレオチド、(b)SEQ ID NO:9-16を有する群から選択し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有するポリヌクレオチド、(c)(a)のポリヌクレオチドに相補的なポリヌクレオチド、(d)(b)のポリヌクレオチドに相補的なポリヌクレオチド、または(e)(a)~(d)のRNA等価物を含む群から選択されたポリヌクレオチドの配列を有する。検出方法は、(a)サンプル中の上記標的ポリヌクレオチドに相補的な或る配列からなる少なくとも20の連続したヌクレオチド群からなる或るプローブを用いて該サンプルをハイブリダイズする過程と、(b)該ハイブリダイゼーション複合体の有無を検出し、複合体が存在すればオプションでその量を検出する過程からなる。該プローブと該標的ポリヌクレオチドあるい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ような条件下で、プローブは、該標的ポリヌクレオチドに対し特異的にハイブリダイズする。一実施態様では、プローブは少なくとも60の連続したヌクレオチドを含む。

10

【0058】

本発明はまた、サンプル中の標的ポリヌクレオチドを検出する方法を提供する。ここで、標的ポリヌクレオチドは、(a)SEQ ID NO:9-16からなる群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(b)SEQ ID NO:9-16からなる群から選択し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、(c)(a)のポリヌクレオチドに相補的なポリヌクレオチド、(d)(b)のポリヌクレオチドに相補的なポリヌクレオチド、および(e)(a)~(d)のRNA等価物からなる群から選択された配列のポリヌクレオチドを有する。検出方法は、(a)ポリメラーゼ連鎖反応増幅を用いて標的ポリヌクレオチドまたはその断片を増幅する過程と、(b)増幅した標的ポリヌクレオチドまたはその断片の存在・不存在を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程を含む。

20

30

【0059】

本発明は更に、有効量のポリペプチドと薬剤として許容できる賦形剤とを含む組成物を提供する。有効量のポリペプチドは、(a)SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチド、(b)SEQ ID NO:1-8からなる群から選択し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、(c)SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチドの生物学的活性断片、および(d)SEQ ID NO:1-8を有する群から選択したアミノ酸配列の免疫原性断片からなる群から選択される。一実施例では、SEQ ID NO:1-8からなる一群から選択されたアミノ酸配列を含む組成物を提供する。更に、本発明は、患者にこの組成物を投与することを含む、機能的KPPの発現の低下に関連した疾患やその症状の治療方法を提供する。

40

【0060】

本発明はまた、(a)SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチド、(b)SEQ ID NO:1-8からなる群から選択し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、(c)SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、および(d)SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片からなる群から選択されたポリペプチドのアゴニストとしての有効性を確認するために化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a)ポリペプチドを含むサンプルを化合物に曝す過程と、(b)サンプル中のアゴニスト活性を検出す

50

る過程とを含む。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KPPの発現の低下に関連した疾患やその症状の治療方法を提供する。

【0061】

本発明は更に、(a) SEQ ID NO:1-8を有する群から選択したアミノ酸配列を有するポリペプチド、(b) SEQ ID NO:1-8を有する群から選択したアミノ酸配列と少なくとも90%の同一性である天然のアミノ酸配列を有するポリペプチド、(c) SEQ ID NO:1-8を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d) SEQ ID NO:1-8を有する群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を含む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の有効性を確認するために化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) ポリペプチドを含むサンプルを化合物に曝露する過程と、(b) サンプル中のアンタゴニスト活性を検出する過程とを含む。一実施態様で本発明は、この方法によって同定したアンタゴニスト化合物と薬剤として許容できる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KPP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

【0062】

本発明は更に、(a) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を含むポリペプチド、(b) SEQ ID NO:1-8からなる群から選択したアミノ酸配列と少なくとも90%の同一性である天然のアミノ酸配列を含むポリペプチド、(c) SEQ ID NO:1-8を有する群から選択したアミノ酸配列を有するポリペプチドの生物学的活性断片、または(d) SEQ ID NO:1-8からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を含む群から選択されたポリペプチドに特異結合する化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) ポリペプチドを適切な条件下で少なくとも1つの試験化合物に混合させる過程と、(b) 試験化合物とのポリペプチドの結合を検出し、それによってポリペプチドに特異結合する化合物を同定する過程とを含む。

20

【0063】

本発明は更に、(a) SEQ ID NO:1-8を有する群から選択したアミノ酸配列を有するポリペプチド、(b) SEQ ID NO:1-8を有する群から選択し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有するポリペプチド、(c) SEQ ID NO:1-8を有する群から選択したアミノ酸配列の生物学的活性断片、または(d) SEQ ID NO:1-8を有する群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を含むポリペプチドの活性を調節する化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) ポリペプチドの活性が許容された条件下で、ポリペプチドを少なくとも1つの試験化合物と混合させる過程と、(b) ポリペプチドの活性を試験化合物の存在下で算定する過程と、(c) 試験化合物の存在下でのポリペプチドの活性を試験化合物の不存在下でのポリペプチドの活性と比較する過程とを含み、試験化合物の存在下でのポリペプチドの活性の変化は、ポリペプチドの活性を調節する化合物であることを意味する。

30

【0064】

本発明は更に、標的ポリヌクレオチドの変異発現の有効性を確認するために化合物をスクリーニングする方法を提供する。標的ポリヌクレオチドは、SEQ ID NO:9-16からなる群から選択した配列を含む。スクリーニング方法は、(a) 標的ポリヌクレオチドを含むサンプルを化合物に曝す過程と、(b) 標的ポリヌクレオチドの変異発現を検出する過程、(c) 可変量の前記化合物の存在下と前記化合物の不存在下で、前記標的ポリヌクレオチドの発現を比較する過程を含む。

40

【0065】

本発明は更に、(a) 核酸を含む生物学的サンプルを試験化合物で処理する過程と、(b) (i) SEQ ID NO:9-16を有する群から選択したポリヌクレオチド配列を含むポリヌクレオチド、(ii) SEQ ID NO:9-16を有する群から選択したポリヌクレオチド配列と少なくとも

50

も90%が同一であ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、(iii)(i)に相補的な配列を有するポリヌクレオチド、(iv)(ii)のポリヌクレオチドに相補的なポリヌクレオチド、(v)(i)~(iv)のRNA等価物を含む群から選択したポリヌクレオチドの少なくとも20の連続したヌクレオチドから構成されるプローブを用いて、処理した生物学的サンプルの核酸をハイブリダイズする過程とを含む試験化合物の毒性の算定方法を提供する。ハイブリダイゼーションは、上記プローブと生物学的サンプル中の標的ポリヌクレオチドの間に特定のハイブリダイゼーション複合体が形成されるような条件下で発生し、上記標的ポリヌクレオチドは、(i)SEQ ID NO:9-16を有する群から選択したポリヌクレオチド配列を有するポリヌクレオチド、(ii)SEQ ID NO:9-16を有する群から選択したポリヌクレオチド配列と少なくとも90%が同一である天然のポリヌクレオチド配列を有するポリヌクレオチド、(iii)(i)のポリヌクレオチドに相補的なポリヌクレオチド、(iv)(ii)のポリヌクレオチドに相補的なポリヌクレオチド、(v)(i)~(iv)のRNA等価物を含む群から選択する。或いは、標的ポリヌクレオチドは、上記(i)~(v)を含む群から選択したポリヌクレオチド配列の断片と、(c)ハイブリダイゼーション複合体の量を定量する過程と、(d)処理した生物学的サンプルのハイブリダイゼーション複合体の量を、非処理の生物学的サンプルのハイブリダイゼーション複合体の量と比較する過程を含み、処理した生物学的サンプルのハイブリダイゼーション複合体の量の差は、試験化合物の毒性を示す。

10

【発明を実施するための最良の形態】

【0066】

20

本発明のタンパク質、ヌクレオチド配列及び方法について説明するが、その前に、説明した特定の装置、材料及び方法に本発明が限定されるものではなく、改変し得ることを理解されたい。また、ここで使用する専門用語は特定の実施例を説明する目的で用いたものに過ぎず、特許請求の範囲にのみ限定される本発明の範囲を限定することを意図したのではないことも併せて理解されたい。

【0067】

請求の範囲及び明細書中で用いている単数形の「或る」及び「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数のもを指す場合もあることに注意しなければならない。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」は複数の抗体が含まれ、当業者には周知の等価物なども含まれる。

30

【0068】

本明細書中で用いる全ての専門用語及び科学用語は、特に定義されている場合を除き、当業者に一般に理解されている意味と同じ意味を有する。本明細書で説明するものと類似あるいは同等の任意の装置、材料及び方法を用いて本発明の実施または試験を行うことができるが、ここでは好適な装置、材料、方法について説明する。本発明で言及する全ての刊行物は、刊行物中で報告されていて且つ本発明に関係があるであろう細胞、プロトコル、試薬及びベクターについて説明及び開示する目的で引用しているものである。本明細書のいかなる開示内容も、本発明が先行技術の効力によってこのような開示に対して先行する権利を与えられていないことを認めるものではない。

【0069】

40

(定義)

用語「KPP」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KPPのアミノ酸配列を指す。

【0070】

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

【0071】

50

「対立遺伝子変異体」は、KPPをコードする遺伝子の別の形態である。対立遺伝子変異体は、核酸配列における少なくとも1つの突然変異から作製し得る。また、変異RNAまたはポリペプチドを作製し得る。ポリペプチドの構造または機能は、変異することもしないこともある。遺伝子は、天然の対立遺伝子変異体を全く有しないか、1個若しくは数個の天然の対立遺伝子変異体を有し得る。一般に対立遺伝子変異体を生じさせる通常の突然変異性変化は、ヌクレオチドの自然欠失、付加または置換に帰するものである。これら各変化は、単独或いは他の変化と共に、所定の配列内で1回若しくは数回生じ得る。

【0072】

KPPをコードする「変異(altered)」核酸配列は、種々のヌクレオチドを欠失、挿入または置換する核酸配列を有し、KPPと同一またはKPPの機能的特徴を少なくとも1つ有するポリペプチドを産出する。この定義には、KPP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対立遺伝子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KPP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型性を含む。コードされたタンパク質も変異され得、サイレント変化を生じKPP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KPP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、負に帯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あり、正に帯電したアミノ酸にはリジン及びアルギニンがある。親水性値が近似している非荷電極性側鎖を有するアミノ酸には、アスパラギンとグルタミン、セリンとトレオニンがある。親水性値が近似している非荷電側鎖を有するアミノ酸には、ロイシンとイソロイシンとバリン、グリシンとアラニン、フェニルアラニンとチロシンがある。

10

20

【0073】

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。

【0074】

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。増幅は通常、当業者によく知られたポリメラーゼ連鎖反応(PCR)技術を用いて行う。

30

【0075】

用語「アンタゴニスト」は、KPPの生物学的活性を阻害或いは減弱する分子である。アンタゴニストとしては、KPPと直接相互作用して、或いはKPPが関与する生物学的経路の諸成分に作用してKPP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。

【0076】

「抗体」の語は、抗原決定基と結合することができる、無傷の免疫グロブリン分子やその断片、例えばFab、F(ab')₂及びFv断片を指す。KPPポリペプチドと結合する抗体は、免疫化抗原として、無傷のポリペプチド群を用いて、または、当該の小ペプチド群を持つ断片群を用いて作製可能である。動物(マウス、ラットあるいはウサギ等)を免疫化するために用いるポリペプチドまたはオリゴペプチドは、RNAの翻訳、または化学合成によって得られるポリペプチドまたはオリゴペプチドに由来し得るもので、好みに応じてキャリアタンパク質に抱合することも可能である。通常用いられるキャリアであってペプチドと化学結合するものは、ウシ血清アルブミン、サイログロブリン及びスカシガイのヘモシアニン(KLH)等がある。その結合ペプチドは、動物を免疫化するために用いる。

40

【0077】

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。タンパク質またはタンパク質断片を用いて宿主動物を免疫化する場合、タンパク質の多数の

50

領域が、抗原決定基（タンパク質の特定の領域または3次元構造）に特異結合する抗体の産生を誘導し得る。抗原決定基は、抗体に結合するための無損傷抗原（即ち免疫応答を誘導するために用いられる免疫原）と競合し得る。

【0078】

用語「アプタマー」は、特定の分子標的に結合する核酸またはオリゴヌクレオチド分子を指す。アプタマーは *in vitro* の進化過程（例えば米国特許番号第5,270,163号に記載された SELEX (Systematic Evolution of Ligands by EXponential Enrichment)）から由来するもので、そのような過程は大きな組み合わせライブラリから標的特異的なアプタマー配列を選択する。アプタマー組成は、2本鎖または1本鎖であってもよく、デオキシリボヌクレオチド、リボヌクレオチド、ヌクレオチド誘導體、または他のヌクレオチド様分子を含み得る。アプタマーのヌクレオチド成分は、修飾された糖基（例えばリボヌクレオチドの2'-OH基が2'-Fまたは2'-NH₂で置換し得る）を有することが可能で、そのような糖基はヌクレアーゼへの抵抗性または血液中でのより長い寿命などの望ましい性質に改善し得る。循環系からアプタマーが除去される速度を遅くするために、アプタマーを高分子量キャリア等の分子に抱合させることができる。アプタマー類は、例えば或る架橋剤の光活性化によって、それらの同種リガンド群と特異的に架橋させ得る（Brody, E.N.及びL. Gold (2000) *J. Biotechnol.* 74:5-13等を参照）。

10

【0079】

「イントラマー (intramer)」の用語は *in vivo* で発現されるアプタマーを意味するたとえば、ワクシニアウイルスに基づくRNA発現系は、白血球の細胞質で特定のRNAアプタマーが高レベルで発現するために使用されている（Blind, M. 他 (1999) *Proc. Natl Acad. Sci. USA* 96:3606-3610）。

20

【0080】

「スピーゲルマー (spiegelmer)」の語はL-DNA、L-RNAその他の左旋性ヌクレオチド誘導體またはヌクレオチド様分子を含むアプタマーを指す。左旋性のヌクレオチドを含むアプタマーは右旋性ヌクレオチドに作用する天然の酵素による分解に対して耐性がある。

【0081】

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。アンチセンス組成物には、DNAや、RNAや、ペプチド核酸（PNA）や、ホスホロチオ酸、メチルホスホン酸またはベンジルホスホン酸等の修飾されたバックボーン結合を有するオリゴヌクレオチドや、2'-メトキシエチル糖または2'-メトキシエトキシ糖等の修飾された糖類を有するオリゴヌクレオチドや、或いは5-メチルシトシン、2-デオキシウラシルまたは7-デアザ-2'-デオキシグアノシン等の修飾された塩基を有するオリゴヌクレオチドがある。アンチセンス分子は、化学合成または転写を含む任意の方法で製造することができる。相補的アンチセンス分子は、ひとたび細胞に導入されたら、細胞が形成した天然の核酸配列と塩基対を形成し、転写または翻訳を妨害する二重鎖を形成する。「負」または「マイナス」という表現は、ある参考DNA分子のアンチセンス鎖を意味し、「正」または「プラス」という表現は、ある参考DNA分子のセンス鎖を意味する。

30

40

【0082】

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KPP、合成のKPP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。

【0083】

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'A-G-T3'」は、相補配列「3'T-C-A5'」と対を形成する。

【0084】

「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

50

」は広い意味で、所定のポリ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。この組成物には、乾燥製剤または水溶液が含まれ得る。KPP若しくはKPP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は、塩（例えばNaCl）、界面活性剤（例えばドデシル硫酸ナトリウム；SDS）及びその他の構成成分（例えばデンハート液、脱脂粉乳、サケの精子のDNA等）を含む水溶液中にプローブを分散させることができる。

【0085】

「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、X 10
L-PCRキット（Applied 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オーバーラップする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。伸長及び構築の両方を行ってコンセンサス配列を決定する配列もある。

【0086】

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。下表は、タンパク質中で元のアミノ酸と置換され得るア 20
ミノ酸と、保存アミノ酸置換と認められるアミノ酸を示している。

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

30

40

【0087】

保存アミノ酸置換では通常、（a）置換領域におけるポリペプチドのバックボーン構造、例えばシートや螺旋構造、（b）置換部位における分子の電荷または疎水性、及び/または（c）側鎖の大部分を保持する。

【0088】

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。

50

【0089】

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。例えば、アルキル基、アシル基、ヒドロキシル基またはアミノ基による水素の置換は、ポリヌクレオチド配列の化学修飾に含まれ得る。ポリヌクレオチド誘導体は、天然分子の生物学的または免疫学的機能を少なくとも1つは保持しているポリペプチドをコードする。ポリペプチド誘導体は、グリコシル化、ポリエチレングリコール化(pegylation)、或いは任意の同様なプロセスであって、誘導起源のポリペプチドの少なくとも1つの生物学的若しくは免疫学的機能を保持するプロセスによって、修飾されたポリペプチドである。

【0090】

「検出可能な標識」は、測定可能な信号を発生し得る、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。 10

【0091】

「示差発現」は少なくとも2つの異なったサンプルを比較することによって決められる、増加(上方調節)、あるいは減少(下方調節)、または遺伝子発現の欠損またはタンパク発現の欠損を指す。このような比較は例えば、治療後サンプルと未治療のサンプルまたは病態のサンプルと正常サンプルの間で行われ得る。

【0092】

「エキソンシャフリング」は、異なるコード領域(エキソン)の組換えを意味する。1つのエキソンがコードされたタンパク質の1つの構造的または機能的ドメインを代表し得るため、安定した基礎構造の新規な再分類を介して新しいタンパク質が組立てられることが可能であり、新しいタンパク質機能の進化を促進できる。 20

【0093】

「断片」は、KPPまたはKPPをコードするポリヌクレオチドの固有部分であって、親配列と同一配列であるが親配列よりも長さが短い配列を指す。断片は、定義された配列の全長から1ヌクレオチド/アミノ酸残基を差し引いた長さよりも短い長さを有し得る。例えば或る断片は、5~1000の連続したヌクレオチドまたはアミノ酸残基を有し得る。プローブ、プライマー、抗原、治療用分子として、或いはその他の目的のために用いられる断片は、少なくとも5、10、15、16、20、25、30、40、50、60、75、100、150、250若しくは500の連続したヌクレオチド或いはアミノ酸残基の長さであり得る。断片は、或る分子の特定領域から優先的に選択し得る。例えば、ポリペプチド断片は、所定の配列に示すようなポリペプチドの最初の250または500アミノ酸(または最初の25%または50%)から選択された或る長さの連続したアミノ酸を有し得る。これらの長さは明らかに例として挙げているものであり、本発明の実施例では、配列表、表及び図面を含む明細書に裏付けされた任意の長さであってよい。 30

【0094】

SEQ ID NO: 9-16の断片には、固有のポリヌクレオチド配列領域が含まれる。この領域は、SEQ ID NO: 9-16を特異的に同定するものであり、例えばこの断片を得たゲノム中のその他の配列とは異なるものである。SEQ ID NO: 9-16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: 9-16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。SEQ ID NO: 9-16の断片の正確な長さ及び断片に対応するSEQ ID NO: 9-16の領域は、断片に対する意図した目的に基づき当業者が慣例的に決定することが可能である。 40

【0095】

SEQ ID NO: 1-8のある断片はSEQ ID NO: 9-16のある断片によってコードされている。SEQ ID NO: 1-8の断片はSEQ ID NO: 1-8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: 1-8の断片は、SEQ ID NO: 1-8を特異認識する抗体を産出するための免疫原性ペプチドとして有用である。SEQ ID NO: 1-8のある断片の正確な長さとその断片に対応するSEQ ID NO: 1-8での領域は、その断片の目的に基づいて当業者が慣例的に決定することが可能である。

【0096】

「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）、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。

【0097】

「相同性」の語は、2つ以上のポリヌクレオチド配列または2つ以上のポリペプチド配列の配列類似性、または配列同一性を意味する。

【0098】

ポリヌクレオチド配列についての用語「一致率」または「%一致」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、2配列間のアラインメントを最適化するために比較する配列において、標準化された再現性のある方法でギャップを挿入するので、2つの配列をより有意に比較できる。

10

【0099】

ポリヌクレオチド配列間の一致率は、MEGALIGN version 3.12e配列アラインメントプログラムに組込まれているようなCLUSTAL Vアルゴリズムのデフォルトのパラメータを用いて決定できる。このプログラムは、LASERGENE ソフトウェアパッケージ（一組の分子生物学的分析プログラム）(DNASTAR, Madison WI)の一部である。このCLUSTAL Vは、Higgins, D. G.およびP. M. Sharp (1989) CABIOS 5:151-153、Higgins, D. G. 他 (1992) CABIOS 8:189-191に詳述されている。ポリヌクレオチド配列をペアワイズでアラインメントする際の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。一致率は、アラインメントされたポリヌクレオチド配列間の「類似性パーセント」としてCLUSTAL Vによって報告される。

20

【0100】

或いは、一般的に用いられ且つ自由に入手できる配列比較アルゴリズム一式が、国立バイオテクノロジー情報センター (NCBI) Basic Local Alignment Search Tool (BLAST) から提供されており (Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410)、これはメリーランド州ベセスダにあるNCBI及びインターネット (<http://www.ncbi.nlm.nih.gov/BLAST/>) を含む幾つかの情報源から入手可能である。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を直接にペアワイズで比較するために用いられる。「BLAST 2 Sequences」は、<http://www.ncbi.nlm.nih.gov/gorf/b12.html>にアクセスして、対話形式で利用できる。「BLAST 2 Sequences」ツールは、blastn 及び blastp (以下に記載) の両方に用いることができる。BLASTプログラムは、一般的には、ギャップ及びデフォルト設定に設定された他のパラメータと共に用いる。例えば、2つのヌクレオチド配列を比較するために、デフォルトパラメータとして設定された「BLAST 2 Sequences」ツールVersion 2.0.12 (2000年4月21日) を用いてblastnを実行してもよい。デフォルトパラメータの設定例を以下に示す。

30

【0101】

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

40

一致率は、ある定義された配列の全長（例えば特定のSEQ IDナンバーで定義された配列）で測定し得る。或いは、より短い長さ、例えば、定義された、より大きな配列から得られ

50

た断片（例えば少なくとも20、30、40、50、70、100または200の連続したヌクレオチドの断片）の長さと比較して一致率を測定してもよい。ここに挙げた長さは単なる例示的なものに過ぎず、表、図及び配列リストを含めた本明細書に記載された配列に裏付けられた任意の配列長さの断片を用いて、一致率を測定し得る長さを説明し得ることを理解されたい。

【0102】

高度の同一性を示さない核酸配列が、それにもかかわらず遺伝子コードの縮重が原因で類似のアミノ酸配列をコードする場合がある。この縮重を利用して核酸配列内で変化を生じさせて、全ての核酸配列が実質上同一のタンパク質をコードするような多数の核酸配列を生成し得ることを理解されたい。

10

【0103】

ポリペプチド配列に用いられる用語「一致率」または「%一致」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公知である。保存的アミノ酸置換を考慮するアラインメント方法もある。既に詳述したこのような保存的置換は通常、置換部位の酸性度及び疎水性を保存するので、ポリペプチドの構造を（従って機能も）保存する。

【0104】

ポリペプチド配列間の一致率は、MEGALIGN version 3.12e配列アラインメントプログラムに組み込まれているようなCLUSTAL Vアルゴリズムのデフォルトのパラメータを用いて決定できる（既に説明したのでそれを参照されたい）。CLUSTAL Vを用いて、ポリペプチド配列をペアワイズアラインメントする際のデフォルトパラメータは、Ktuple=1、gap penalty=3、window=5、「diagonals saved」=5と設定される。デフォルトの残基重み付け表としてPAM250マトリクスを選択する。ポリヌクレオチドアラインメントと同様に、CLUSTAL Vは、アラインメントされたポリペプチド配列対間の「類似率」として一致率を報告する。

20

【0105】

或いは、NCBI BLASTソフトウェア一式を用いてもよい。例えば、2つのポリペプチド配列をペアワイズで比較をする場合、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（2000年4月21日）でblastpを使用するであろう。デフォルトパラメータの設定例を以下に示す。

30

【0106】

Matrix: BLOSUM62

Open Gap: 11 及び Extension Gap: ペナルティ 1

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

一致率は、ある定義されたポリペプチド配列の全長（例えば特定のSEQ IDナンバーで定義された配列）で測定し得る。或いは、より短い長さ、例えば、定義された、より大きなポリペプチド配列から得られた断片（例えば少なくとも15、20、30、40、50、70、または150の連続した残基の断片）の長さと比較して一致率を測定してもよい。ここに挙げた長さは単なる例示的なものに過ぎず、表、図及び配列リストを含めた本明細書に記載された配列に裏付けられた任意の配列長さの断片を用いて、或る長さであってその長さに対して一致率を測定し得る長さを説明し得ることを理解されたい。

40

【0107】

「ヒト人工染色体（HAC）」は、約6 kb ~ 10 MbのサイズのDNA配列を含み得る、安定した染色体複製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微小染色体である。

【0108】

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

50

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。

【0109】

「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。特異的ハイブリダイゼーションは、2つの核酸配列が高い相補性を共有することの指標である。特異的ハイブリダイゼーション複合体は許容されるアニーリング条件下で形成され、「洗浄」ステップ後もハイブリダイズされたままである。洗浄ステップは、ハイブリダイゼーションプロセスのストリンジェンシーを決定する際に特に重要であり、更にストリンジェントな条件では、非特異結合（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。核酸配列のアニーリングに対する許容条件は、本技術分野における当業者が慣例的に決定できる。許容条件はハイブリダイゼーション実験の間は一定でよいが、洗浄条件は所望のストリンジェンシーを得るように、従ってハイブリダイゼーション特異性も得るように実験中に変更することができる。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1000μg/mlのせん断して変性したサケ精子DNAが含まれる。

10

【0110】

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或る程度、洗浄ステップを実行する温度を基準にして表すことができる。このような洗浄温度は通常、所定のイオン強度及びpHにおける特異配列の融点（ T_m ）より約5～20低くなるように選択する。この T_m は、所定のイオン強度及びpHの条件下で、完全に一致するプローブに標的配列の50%がハイブリダイズする温度である。 T_m を計算する式及び核酸のハイブリダイゼーション条件はよく知られており、Sambrook 他（1989）Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NYに記載されており、特に2巻の9章を参照されたい。

20

【0111】

本発明のポリヌクレオチドとポリヌクレオチドのハイブリダイゼーションに対する高ストリンジェンシー条件には、約0.2×SSC及び約1%のSDS存在下で約68において1時間の洗浄条件が含まれる。或いは、65、60、55または42の温度で行ってもよい。SSC濃度は、約0.1%のSDS存在下で、約0.1～2×SSCの範囲で変化し得る。通常は、ブロッキング剤を用いて非特異ハイブリダイゼーションを阻止する。このような遮断試薬には、例えば、約100～200μg/mlのせん断され変性したサケ精子DNAが含まれる。例えばRNAとDNAのハイブリダイゼーションのような特定条件下では、有機溶剤、例えば約35～50%v/vの濃度のホルムアミドを用いることもできる。洗浄条件の有用なバリエーションは、当業者には自明であろう。ハイブリダイゼーションは、特に高ストリンジェント条件下では、ヌクレオチド間の進化的な類似性を示唆し得る。このような類似性は、ヌクレオチド及びヌクレオチドにコードされるポリペプチドに対する類似の役割を強く示唆している。

30

【0112】

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、溶解状態で形成し得る（ C_0t または R_0t 解析等）。或いは、一方の核酸配列が溶解状態で存在し、もう一方の核酸配列が固体支持体（例えば紙、膜、フィルタ、チップ、ピンまたはガラススライド、或いは他の適切な基板であって細胞若しくはその核酸が固定される基板）に固定されているような2つの核酸配列間に形成され得る。

40

【0113】

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。

【0114】

「免疫応答」は、炎症、外傷、免疫異常症、伝染性疾患または遺伝性疾患に関連する症状を指し得る。これらの症状は、細胞及び全身の防御系に作用し得る種々の因子、例えばサ

50

イトカイン、ケモカイン、その他のシグナル伝達分子の発現によって特徴づけることができる。

【0115】

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ことが可能なKPP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KPPのポリペプチド断片またはオリゴペプチド断片を含む。

【0116】

用語「マイクロアレイ」は、基板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。

10

【0117】

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。

【0118】

用語「調節」は、KPPの活性の変化を指す。例えば、モジュレートによって、KPPのタンパク質活性の増減、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。

【0119】

「核酸」及び「核酸配列」の語は、ヌクレオチド、オリゴヌクレオチド、ポリヌクレオチドまたはこれらの断片を指す。「核酸」及び「核酸配列」の語は、ゲノム起源または合成起源のDNAまたはRNAであって一本鎖または二本鎖であるか或いはセンス鎖またはアンチセンス鎖を表し得るようなDNAまたはRNAや、ペプチド核酸(PNA)や、任意のDNA様またはRNA様物質を指すこともある。

20

【0120】

「機能的に連結した」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及ぼす場合には、そのプロモーターはそのコード配列に機能的に連結している。同一のリーディングフレーム内で2つのタンパク質コード領域を結合する必要がある場合、一般に、機能的に連結したDNA配列は非常に近接するか、或いは連続的に隣接し得る。

【0121】

「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のバックボーンに結合した、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。末端のリシンは、組成物に溶解性を与える。PNAは、相補的一本鎖DNAまたはRNAに優先的に結合して転写の伸長を停止するものであり、ポリエチレングリコール化して細胞におけるPNAの寿命を延長し得る。

30

【0122】

KPP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、KPPの酵素環境に依存し、細胞の種類によって異なり得る。

40

【0123】

「プローブ」とは、同一配列或いは対立遺伝子核酸配列、関連する核酸配列の検出に用いる、KPP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、単離されたオリゴヌクレオチドまたはポリヌクレオチドであって、検出可能な標識またはレポーター分子に結合したものである。典型的な標識には、放射性アイソトープ、リガンド、化学発光試薬及び酵素がある。「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。プライマーは次に、DNAポリメラーゼ酵素によって標的DNA鎖に沿って延長し得る。プライマー対は、例えばポリメラーゼ連鎖反応(PCR)による核酸配列の増幅(及び同定)に用い得る。

50

【0124】

本発明に用いるようなプローブ及びプライマーは通常、既知の配列の少なくとも15の連続したヌクレオチドを含んでいる。特異性を高めるために長めのプローブ及びプライマー、例えば開示した核酸配列の少なくとも20、25、30、40、50、60、70、80、90、100または少なくとも150の連続したヌクレオチドからなるようなプローブ及びプライマーを用いてもよい。これよりもかなり長いプローブ及びプライマーもある。表、図面及び配列リストを含む本明細書に裏付けされた任意の長さのヌクレオチドを用いることができるものと理解されたい。

【0125】

プローブ及びプライマーの調製及び使用方法については、Sambrook, J.他 (1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NY、Ausubel, F.M.他, (1987) Current Protocols in Molecular Biology, Greene Publishing Assoc. & Wiley-Intersciences, New York NY、Innis他 (1990) PCR Protocols, A Guide to Methods and Applications Academic Press, San Diego CA等を参照されたい。PCRプライマー対は、その目的のためのコンピュータプログラム、例えばPrimer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA)を用いるなどして既知の配列から得ることができる。

【0126】

プライマーとして用いるオリゴヌクレオチドの選択は、そのような目的のために本技術分野でよく知られているソフトウェアを用いて行う。例えばOLIGO 4.06ソフトウェアは、各100ヌクレオチドまでのPCRプライマー対の選択に有用であり、32キロベースまでのインプットポリヌクレオチド配列からオリゴヌクレオチド及び最大5,000ヌクレオチドまでの大きめのポリヌクレオチドとオリゴヌクレオチドを分析するのにも有用である。類似のプライマー選択プログラムには、拡張能力のための追加機能が組み込まれている。例えば、PrimOUプライマー選択プログラム(テキサス州ダラスにあるテキサス大学南西部医療センターのゲノムセンターから一般向けに入手可能)は、メガベース配列から特定のプライマーを選択することが可能であり、従ってゲノム全体の範囲でプライマーを設計するのに有用である。Primer3プライマー選択プログラム(Whitehead Institute/MIT Center for Genome Research(マサチューセッツ州ケンブリッジ)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」を入力できる。Primer3は特に、マイクロアレイのためのオリゴヌクレオチドの選択に有用である(後二者のプライマー選択プログラムのソースコードは、それぞれの情報源から得てユーザー固有のニーズを満たすように修正し得る)。PrimerGenプログラム(英国ヒトゲノムマッピングプロジェクト-リソースセンター(英国ケンブリッジ)から一般向けに入手可能)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し、それによって、アラインメントされた核酸配列の最大保存領域または最小保存領域の何れかとハイブリダイズするようなプライマーの選択を可能にする。従って、このプログラムは、固有であって保存されたオリゴヌクレオチド及びポリヌクレオチドの断片の同定に有用である。上記選択方法のいずれかによって同定したオリゴヌクレオチド及びポリヌクレオチドの断片は、ハイブリダイゼーション技術において、例えばPCRまたはシーケンシングプライマーとして、マイクロアレイエレメントとして、或いは核酸のサンプルにおいて完全または部分的相補的ポリヌクレオチドを同定する特異プローブとして有用である。オリゴヌクレオチドの選択方法は、上記の方法に限定されるものではない。

【0127】

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。この人為的組合せはしばしば化学合成によって達成するが、より一般的には核酸の単離セグメントの人為的操作によって、例えばSambrook他の文献(前出)に記載されているような遺伝子工学的的手法によって達成する。組換え核酸の話は、単に核酸の一部を付加、置換または欠失した変異核酸も含む。しばしば組換え核酸には、プロモーター配列に機能的に連結した核酸配列が含まれる。このような組換

10

20

30

40

50

え核酸は、例えば細胞を形質転換するために使用されるベクターの一部とすることが可能である。

【0128】

或いはこのような組換え核酸は、ウイルスベクターの一部であって、例えばワクシニアウイルスに基づくものであり得る。そのようなワクシニアウイルスは哺乳動物に接種され、その組換え核酸が発現されて、その哺乳動物内で防御免疫応答を誘導するように使用することができる。

【0129】

「調節因子」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。調節因子は、転写、翻訳またはRNA安定性を調節する宿主タンパク質またはウイルスタンパク質と相互作用する。

10

【0130】

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

【0131】

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。

20

【0132】

用語「サンプル」は、その最も広い意味で用いられている。KPP、KPP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。

【0133】

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、タンパク質の特定の構造(例えば抗原決定基即ちエピトープ)であって結合分子が認識するものが存在するか否かに依存していることを意味している。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。

30

【0134】

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

【0135】

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

40

【0136】

用語「基板」は、任意の好適な固体或いは半固体の支持物を指し、膜及びフィルター、チップ、スライド、ウエハ、ファイバー、磁気または非磁気ビード、ゲル、チューブ、プレート、ポリマー、微小粒子、毛細管が含まれる。基板は、壁、溝、ピン、チャネル、孔等、様々な表面形態を有することができ、基板にはポリヌクレオチドやポリペプチドが結合する。

【0137】

「転写イメージ」あるいは「発現プロフィール」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。

50

【0138】

「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、本技術分野で知られている種々の方法に従って自然条件または人工条件下で生じ得るものであり、外来性の核酸配列を原核宿主細胞または真核宿主細胞に挿入する任意の既知の方法を基にし得る。形質転換の方法は、形質転換する宿主細胞の種類によって選択する。限定するものではないが形質転換方法には、バクテリオファージあるいはウイルス感染、電気穿孔法（エレクトロポレーション）、熱ショック、リポフェクション及び微粒子銃を用いる方法がある。「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、限られた時間に一過的に導入DNA若しくは導入RNAを発現する細胞も含まれる。

10

【0139】

ここで用いる「遺伝形質転換体」とは任意の生物体であり、限定するものではないが動植物を含み、生物体の1個若しくは数個の細胞が、ヒトの関与によって、例えば本技術分野でよく知られている形質転換技術によって導入された異種核酸を有する。核酸の細胞への導入は、直接または間接的に、細胞の前駆物質に導入することによって、計画的な遺伝子操作によって、例えば微量注射法によって或いは組換えウイルスの導入によって行う。遺伝子操作の語は、古典的な交雑育種或いは *in vitro* 受精を指すものではなく、組換えDNA分子の導入を指すものである。本発明に基づいて予期される遺伝形質転換体には、バクテリア、シアノバクテリア、真菌及び動植物がある。本発明の単離されたDNAは、本技術分野で知られている方法、例えば感染、形質移入、形質転換またはトランス接合によって宿主に導入することができる。本発明のDNAをこのような有機体に移入する技術はよく知られており、前出の Sambrook 他(1989)等の参考文献に記載されている。

20

【0140】

特定の核酸配列の「変異体」は、核酸配列1本全部の長さに対して特定の核酸配列と少なくともとも40%の同一性を有する核酸配列であると定義する。その際、デフォルトパラメータに設定した「BLAST 2 Sequences」ツールVersion 2.0.9(1999年5月7日)を用いてblastnを実行する。このような核酸対は、所定の長さに対して、例えば少なくとも50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%またはそれ以上の配列同一性を示し得る。或る変異体は、例えば「対立遺伝子」変異体(前述)、「スプライス」変異体、「種」変異体または「多型性」変異体として説明し得る。スプライス変異体は参照分子とかなりの同一性を有し得るが、mRNAプロセッシング中のエキソンの選択的スプライシングによって通常、より多くまたはより少数のポリヌクレオチドを有することになる。対応するポリペプチドは、追加機能ドメインを有するか或いは参照分子に存在するドメインが欠落していることがある。種変異体は、種によって異なるポリヌクレオチド配列である。結果的に生じるポリペプチドは通常、相互にかなりのアミノ酸同一性を有する。多型性変異体は、与えられた種の個体間で特定の遺伝子のポリヌクレオチド配列中での変異である。多型変異配列はまた、ポリヌクレオチド配列の1つのヌクレオチド塩基が異なる「1塩基多型性」(SNP)も含み得る。SNPの存在は、例えば特定の個体群、病状または病状性向を示し得る。特定のポリペプチド配列の「変異体」は、ポリペプチド配列の1本の長さ全体で特定のポリペプチド配列に対して少なくともとも40%の配列同一性を有するポリペプチド配列として画定される。ここで、デフォルトパラメータに設定した「BLAST 2 Sequences」ツールVersion 2.0.9(1999年5月7日)を用いてblastpを実行する。このようなポリペプチド対は、所定の長さに対して、例えば少なくともとも50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%またはそれ以上の配列同一性を示し得る。

30

40

【0141】

(発明)

本発明は、新規のキナーゼおよびホスファターゼ(KPP)、およびKPPをコードするポリヌク

50

レオチドの発見、及び心血管疾患、免疫系の疾患、神経の疾患、成長および発達に影響を及ぼす疾患、脂質異常、細胞増殖異常及び癌の診断・治療あるいは予防のためのこれらの組成物の使用に基づく。

【0142】

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命名の概略である。各ポリヌクレオチド及びその対応するポリペプチドは、1つのIncyteプロジェクト識別番号（IncyteプロジェクトID）と相関する。各ポリペプチド配列は、ポリペプチド配列識別番号（ポリペプチドSEQ ID NO）とIncyteポリペプチド配列番号（IncyteポリペプチドID）によって表示した。各ポリヌクレオチド配列は、ポリヌクレオチド配列識別番号（ポリヌクレオチドSEQ ID NO）とIncyteポリヌクレオチドコンセンサス配列番号（IncyteポリヌクレオチドID）によって表示した。

10

【0143】

表2は、GenBankタンパク質（genpept）データベースに対するBLAST分析によって同定されたような、本発明のポリペプチドとの相同性を有する配列を示している。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（ポリペプチドSEQ ID NO:）と、それに対応するIncyteポリペプチド配列番号（IncyteポリペプチドID）を示す。列3は、最も近いGenBank相同体のGenBank識別番号（Genbank ID NO）を示す。列4は、各ポリペプチドとその相同体との間の一致を表す確率スコアを示す。列5は、該当箇所には適当な引用を示すとともにGenBank相同体1つ以上の注釈（annotation）を示し、これらはすべて本明細書では参考文献に含まれる。

20

【0144】

表3は、本発明の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Incyteポリペプチド配列番号（Incyteポリペプチド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、リン酸化およびグリコシル化の可能性のある部位を示す。列6は、シグネチャ配列、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。

30

【0145】

表2及び表3は共に、本発明の各々のポリペプチドの特性を要約しており、それらの特性が請求の範囲に記載されたポリペプチドがキナーゼおよびホスファターゼであることを確立している。例えば、SEQ ID NO:1は、Basic Local Alignment Search Tool (BLAST)で決定されるようにヒトMAPキナーゼホスファターゼ6(GenBank ID g6840994)と33%同一である（表2参照）。BLAST確率スコアは $1.3e-19$ であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:1はまた、二重特異性ホスファターゼ触媒ドメインを有する。これは、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データベースにおいて、統計的に有意な一致を検索して決定された。（表3参照）。DOMOデータベースを使ったBLAST解析よりのデータは、SEQ ID NO:1がMAPキナーゼホスファターゼであることをさらに確証する証拠を提供する。

40

【0146】

別の例では、SEQ ID NO:2はM1残基からM2299残基までラット系球体メサングウム細胞受容体タンパク質 - チロシンホスファターゼ前駆体(GenBank ID g3300096)と86%同一であることがBasic Local Alignment Search Tool (BLAST)によって示された。（表2参照）。BLAST確率スコアは0.0であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:2はまた、プロテインチロシンホスファターゼドメインを有するが、これは、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データベースにおいて、統計的に有意な一致を検索して決

50

定された。(表3参照)。BLIMPS、MOTIFS、及びPROFILESCAN解析よりのデータは、SEQ ID NO:2 がプロテインチロシンホスファターゼである、さらに確証的な証拠を提供する。

【0147】

また他の例として、SEQ ID NO:3はマウスのタウチュープリンキナーゼ(GenBank ID g15341198)とM1残基からQ316残基まで96%同一であることがBasic Local Alignment Search Tool (BLAST)によって示された(表2参照)。BLAST確率スコアは $1.0e-179$ であり、これは観察されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:3はまた、真核生物プロテインキナーゼドメインを有するが、これは、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ドメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された。(表3参照)。MOTIFS及び付加的なBLAST解析よりのデータは、SEQ ID NO:3 がプロテインキナーゼである、さらに実証的な証拠を提供する。

10

【0148】

代替りの例として、SEQ ID NO:4はH234残基からC1863残基までヒト核内二重特異性ホスファターゼ(GenBank ID g3015538)と98%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2参照)。BLAST確率スコアは 0.0 であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:4はまた、DENN(AEX3)ドメインとプレックストリン相同性ドメインを有するが、これは、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ドメイン群のPFAMデータベースにおいて、統計的に有意な一致を検索して決定された(表3参照)。

20

【0149】

また他の例として、SEQ ID NO:5はマウスのセリン/スレオニンキナーゼ33(GenBank ID g14148952)とM1残基からT382残基まで76%同一であることがBasic Local Alignment Search Tool (BLAST)によって示された(表2参照)。BLAST確率スコアは $1.0e-163$ であり、これは観察されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:5はまた、プロテインキナーゼドメインを有するが、これは、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ドメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された。(表3参照)。BLIMPS、MOTIFS、及びPROFILESCAN解析よりのデータは、SEQ ID NO:5がプロテインキナーゼである、さらに実証的な証拠を提供する。

30

【0150】

別の例として、SEQ ID NO:6はV107残基からL331残基までプロテインホスファターゼ2Cファミリ-と類似の線虫のタンパク質(GenBank ID g2804429)と42%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2参照)。BLAST確率スコアは $5.4e-41$ であるが、これは観測されたポリペプチド配列が偶然に得られる確率を示す。SEQ ID NO:6はまた、プロテインホスファターゼ2Cドメインを有するが、これは、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ドメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された。(表3参照)。BLIMPS及びBLAST_DOMO解析よりのデータは、SEQ ID NO:6がプロテインホスファターゼ2Cであることをさらに確証する証拠を提供する。

40

【0151】

別の例において、SEQ ID NO:7 はマウスのPFTAIREキナーゼ(GenBank ID g2392814)の残基F46からV292に対して75%の同一性があることがBasic Local Alignment Search Tool (BLAST)によって確認された。(表2参照)。BLAST確率スコアは $1.3e-103$ であり、これは観測されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:7はまた、F46残基からR296残基までヒトのセリン/トレオニンプロテインキナーゼPFTAIRE-1(GenBank ID g12002201)に73%の同一性があり、BLAST確率スコア $9.0e-103$ である。SEQ ID NO:7はまた、プロテインキナーゼドメインを持ち、これは隠れマルコフモデル(HMM)

50

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データベースにおいて、統計的に有意な一致を検索して決定された。(表3参照)。MOTIFS及びPROFILESCAN解析よりのデータは、SEQ ID NO:7がプロテインキナーゼであることをさらに確証する証拠を提供する。

【0152】

別の例において、SEQ ID NO:8はK2残基からL281残基までヒトのピリミジン5'ヌクレオチダーゼ(GenBank ID g11245474)に58%の同一性を有するが、これはBasic Local Alignment Search Tool (BLAST)によって同定される。(表2参照)。BLAST確率スコアは $1.1e-90$ であり、これは観察されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:1-8の解析のためのアルゴリズム及びパラメータが表7で記述されている。

10

【0153】

表4に示すように、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)配列を用いて、或いはこれら2種類の配列を任意に組み合わせで構築した。列1は本発明の各ポリヌクレオチドに対するポリヌクレオチド配列識別番号(ポリヌクレオチドSEQ ID NO)および対応するIncyteポリヌクレオチドコンセンサス配列番号(Incyte ID)、および塩基対の各ポリヌクレオチド配列の長さを示している。列2は、本発明の全長ポリヌクレオチド配列を構築するために使われたcDNA配列および/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')位置を示し、またSEQ ID NO:9-16を同定するための、あるいはSEQ ID NO:9-16と関連するポリヌクレオチド配列の間で区別するための技術(例えば、ハイブリダイゼーション技術または増幅技術)に有用なポリヌクレオチド配列の断片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。

20

【0154】

表4の列2で記述されたポリヌクレオチド断片は特に、例えば組織特異的cDNAライブラリあるいはプールしたcDNAライブラリに由来するIncyte cDNAを指す場合もある。或いは列2のポリヌクレオチド断片は、完全長ポリヌクレオチド配列のアセンブリに寄与したGenBank cDNAまたはESTを指す場合もある。さらに、列2のポリヌクレオチド断片は、ENSEMBL(The Sanger Centre、(英国ケンブリッジ))データベースから由来した配列を同定し得る(例えば、「ENST」命名を含む配列)。或いは、列2で記述されたポリヌクレオチド断片は、NCBI RefSeq Nucleotide Sequence Records データベースから由来する場合もあり(即ち「NM」または「NT」の命名を含む配列)、またNCBI RefSeq Protein Sequence Recordsから由来する場合もある(即ち「NP」の命名を含む配列)。または列2のポリヌクレオチド断片は、「エキソンスティッチング(exon-stitching)」アルゴリズムにより結び合わせたcDNA及びGenscan予想エキソンの両方からなる群を意味する場合がある。例えば、FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄と同定されるポリヌクレオチドは、アルゴリズムが適用される配列のクラスターの識別番号がXXXXXXであり、アルゴリズムにより生成される予測の番号がYYYYYであり、(もし存在すれば)N_{1,2,3...}が解析中に手動で編集された可能性のある特定のエキソンであるような「縫合された」配列である(実施例5参照)。または、列2のポリヌクレオチド断片は「エキソンストレッチング(exon-stretching)」アルゴリズムにより結び合わせたエキソンの集合を指す場合もある。例えば、FL_XXXXXX_gAAAAA_gBBBBB_1_Nとして同定されるポリヌクレオチド配列は、「ストレッチされた」配列である。XXXXXXはIncyteプロジェクト識別番号、gAAAAAは「エキソンストレッチング」アルゴリズムを適用したヒトゲノム配列のGenBank識別番号、gBBBBBは一番近いGenBankタンパク質相同体のGenBank識別番号またはNCBI RefSeq 識別番号であり、Nは特定のエキソンに言及している(実施例5を参照)。あるRefSeq配列が「エキソンストレッチング」アルゴリズムのためのタンパク質相同体として使用された場合では、RefSeq識別番号(「NM」、「NP」、または「NT」によって表される)が、GenBank識別(即ち、gBBBBB)の代わりに使用される場合もある。

30

40

【0155】

50

或いは、接頭コードは、手動で編集された構成配列、ゲノムDNA配列から予測された構成配列、または組み合わされた配列解析方法から由来する構成配列を同定する。次の表は、構成配列の接頭コードと、接頭コードに対応する配列分析方法の例を列記する（実施例 4 と 5を参照）。

接頭コード	解析タイプやプログラムの例
GNN 、 GFG 、 ENST	例えば、GENSCAN (Stanford University, CA, USA) または FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)を用いたゲノム配列からのエクソン予測
GBI	手動で編集されたゲノム配列の解析
FL	スティッチまたはストレッチゲノム配列(実施例 5 参照)
INCY	EST 配列のゲノムへのマッピングからの全長転写とエクソンの予想エクソンと結果の転写を予想するために、ゲノム位置と EST 構成データが組み合わされる。

10

【 0 1 5 6 】

場合によっては、最終コンセンサスポリヌクレオチド配列を確認するために表4に示すような配列の適用範囲と重複する Incyte cDNAの適用範囲が得られたが、それに関連する Incyte cDNA識別番号は示さなかった。

20

【 0 1 5 7 】

表 5 は、Incyte cDNA配列を用いて構築された完全長ポリヌクレオチド配列のための代表的なcDNAライブラリを示している。代表的なcDNAライブラリは、上記のポリヌクレオチド配列を構築及び確認するために用いられる Incyte cDNA配列によって、最も頻繁に代表される Incyte cDNAライブラリである。cDNAライブラリを作製するために用いた組織及びベクターを表 5 に示し、表 6 で説明している。

【 0 1 5 8 】

本発明はまた、KPPの変異体も含む。好適なKPPの変異体のアミノ酸配列は、KPPアミノ酸配列と少なくとも約 80%、または少なくとも約 90%、或いは少なくとも約 95%もの一致率を有し、KPPの機能的若しくは構造的特徴を少なくとも1つ有するような変異体である。

30

【 0 1 5 9 】

本発明はまた、KPP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、KPPをコードするSEQ ID NO:9-16からなる一群から選択された配列を含むポリヌクレオチド配列を提供する。SEQ ID NO:9-16のポリヌクレオチド配列は、配列表に示されているように等価RNA配列と同等の価値を有しているが、窒素塩基チミンの出現はウラシルに置換され、糖のバックボーンはデオキシリボースではなくてリボースから構成されている。

40

【 0 1 6 0 】

本発明はまた、KPP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、KPPをコードするポリヌクレオチド配列に対して少なくとも約 70%のポリヌクレオチド配列同一性、或いは少なくとも約 85%のポリヌクレオチド配列同一性、更には少なくとも約 95%ものポリヌクレオチド配列同一性を有する。本発明の或る実施態様では、SEQ ID NO:9-16からなる群から選択されたアミノ酸配列と少なくとも約 70%、或いは少なくとも約 85%、または少なくとも約 95%もの一致率を有するようなSEQ ID NO:9-16からなる群から選択された配列を有するポリヌクレオチド配列の変異配列を含む。上記の任意のポリヌクレオチドの変異体は、KPP

50

の機能的若しくは構造的特徴を少なくとも1つ有するアミノ酸配列をコードし得る。

【0161】

更にあるいは別法では、本発明のポリヌクレオチドの変異体はKPPをコードするポリヌクレオチド配列のスプライス変異体である。スプライス変異体はKPPをコードするポリヌクレオチド配列とかなりの同一性を有し得るが、mRNAプロセッシング中のエキソンの選択的スプライシングによる配列ブロックの追加または削除により、通常、ポリヌクレオチドがより多くまたはより少数の塩基を有することになる。スプライス変異体はその全長についてKPPをコードするポリヌクレオチド配列とのポリヌクレオチド配列同一性が約70%以下、あるいは約60%、あるいは約50%以下でありえるが、そのスプライス変異体のある部分はKPPをコードするポリヌクレオチド配列のある部分とのポリヌクレオチド配列同一性が少なくとも約70%、あるいは少なくとも約85%、あるいは少なくとも約95%、あるいは100%になっている。上記したスプライス変異配列は何れも、KPP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。

10

【0162】

遺伝暗号の縮重により作り出され得るKPP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、可能コドン選択に基づく組合せの選択によって産出し得るようなありとあらゆる可能性のあるポリヌクレオチド配列変異体を網羅し得る。これらの組合せは、天然のKPPのポリヌクレオチド配列に適用されるような標準トリプレット遺伝暗号を基に作られるものであり、このような変異は全て明確に開示されているものと考えられる。

20

【0163】

KPP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KPPのヌクレオチド配列とハイブリダイズ可能であるが、非天然のコドンを含めるなどの実質的に異なった使い方のコドンを含むKPP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。宿主が特定のコドンを利用する頻度に基づいて、特定の真核宿主又は原核宿主に発生するペプチドの発現率を高めるようにコドンを選択することが可能である。コードされたアミノ酸配列を変えることなくKPP及びその誘導体をコードするヌクレオチド配列を実質上変更する別の理由には、天然の配列から作製される転写物より好ましい、例えば半減期が長いなどの特性を有するRNA転写物の作製がある。

30

【0164】

本発明はまた、KPP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、当分野で周知の試薬を用いて、この合成配列を任意の様々な入手可能な発現ベクター及び細胞系中に挿入し得る。更に、合成化学を用いてKPPまたはその任意の断片をコードする或る配列に突然変異を誘導し得る。

【0165】

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:9-16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.およびS.L. Berger (1987) *Methods Enzymol.* 152:399-407、Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) アニール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。

40

【0166】

DNAシーケンシングの方法は当分野では公知であり、本発明のいずれの実施例もDNAシーケンシング方法を用いて実施可能である。DNAシーケンシング方法には酵素を用いることができ、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)を用いることができる。或いは、例えばELONGASE増幅システム (Life Technologies, Gaithersburg MD)において見られるように、ポリメラーゼと校正エキソヌクレアーゼを併用することができる。好適には、MICROLAB2200

50

液体転移システム (Hamilton, Reno, NV)、PTC200サーマルサイクラー (MJ Research, W atertown MA) 及びABI CATALYST 800サーマルサイクラー (Applied Biosystems) 等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 (Applied Biosystems)、MEGABACE 1000 DNAシーケンシングシステム (Molecular Dynamics, Sunnyvale CA) または当分野でよく知られている他の方法を用いてシーケンシングを行う。結果として得られた配列を当分野でよく知られている種々のアルゴリズムを用いて分析する (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, 856-853ページ等を参照)。

【 0 1 6 7 】

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、KPP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、使用し得る方法の1つである制限部位PCR法は、ユニバーサルプライマー及びネステッドプライマー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方法である (例えば、Sarkar, G. (1993) PCR Methods Applic. 2 :318-322を参照)。別の方法に逆PCR法があり、これは広範な方向に伸長させたプライマーを用いて環状化した鋳型から未知の配列を増幅する方法である。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酵素断片から得る (例えば、Triglia, T. 他 (1988) Nucleic Acids Res. 16:8186を参照)。第3の方法としてキャプチャPCR法があり、これ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断片をPCR増幅する方法に 20
関与している。(Lagerstrom, M.他(1991) PCR Methods Applic 1:111-119等を参照)。この方法では、PCRを行う前に複数の制限酵素の消化及びライゲーション反応を用いて未知の配列領域内に組換え二本鎖配列を挿入することが可能である。また、未知の配列を検索するために用い得る別の方法については当分野で知られている (Parker, J.D.他 (1991) Nucleic Acids Res. 19:3055-3060等を参照)。(1991) Nucleic Acids Res. 19:30553060)。更に、PCR、ネステッドプライマー及びPromoterFinder (商標) ライブラリ (Clontech, Palo Alto CA) を用いてゲノムDNAをウォーキングすることができる。この手順は、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/エキソン接合部を見付けるのに 30
有用である。全てのPCRベースの方法に対して、市販されているソフトウェア、例えばOLIGO 4.06プライマー分析ソフトウェア (National Biosciences, Plymouth MN) 或いは別の好適なプログラムを用いて、長さが約22~30ヌクレオチド、GC含有率が約50%以上、温度約68 ~ 72 で鋳型に対してアニーリングするようにプライマーを設計し得る。

【 0 1 6 8 】

完全長cDNAをスクリーニングする際は、より大きなcDNAを含むようにサイズ選択されたライブラリを用いるのが好ましい。更に、ランダムプライマーのライブラリは、しばしば遺伝子の5'領域を有する配列を含み、オリゴd(T)ライブラリが完全長cDNAを作製できない状況に対して好適である。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。

【 0 1 6 9 】

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産物のサイズを分析し、またはそのヌクレオチド配列を確認することができる。具体的には、キャピラリーシーケンシングは、電気泳動による分離のための流動性ポリマーと、4つの異なるヌクレオチドに特異的であるような、レーザで活性化される蛍光色素と、発光された波長の検出に利用するCCDカメラとを有し得る。出力/光の強度は、適切なソフトウェア (Applied Biosystems社のGENOTYPER、SEQUENCE NAVIGATOR等) を用いて電気信号に変換し得る。サンプルのロードからコンピュータ分析及び電子データ表示までの全プロセス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ようなDNA小断片のシーケンシングに特に適している。

【 0 1 7 0 】

10

20

30

40

50

本発明の別の実施態様では、適切な宿主細胞内で、KPP、その断片または機能的等価物を発現させる組換えDNA分子群に、KPPをコードするポリヌクレオチド配列群またはその断片群をクローニングし得る。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配列群を作ることができ、これらの配列をKPPの発現に利用できる。

【0171】

種々の目的で、KPPをコードする配列群を改変するために、当分野で一般的に既知の方法を用いて、本発明のヌクレオチド配列群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。遺伝子断片及び合成オリゴヌクレオチドのランダムなフラグメンテーション及びPCR再アセンブリによるDNAシャッフリングを用い、ヌクレオチド配列を組み換えることが可能である。例えば、オリゴヌクレオチドを介した部位特異的変異誘導を利用して、新規な制限部位の作製、グリコシル化パターンの変更、コドン優先の変更、スプライス変異体の生成等を起こす突然変異を導入し得る。

10

【0172】

本発明のヌクレオチドは、MOLECULARBREEDING (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に記載)等のDNAシャッフリング技術の対象となり、KPPの生物学的特性、例えば生物活性、酵素活性、或いは他の分子や化合物との結合力等を変更または向上させ得る。DNAシャッフリングは、遺伝子断片のPCRを介する組換えを用いて遺伝子変異体のライブラリを作製するプロセスである。ライブラリはその後、その遺伝子変異体を所望の特性に同定するような選択またはスクリーニングにかける。次にこれらの好適な変異体をプールし、更に反復してDNAシャッフリング及び選択/スクリーニングを行ってもよい。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、ランダムポイント突然変異を有する単一の遺伝子の断片を組み換えて、スクリーニングし、その後所望の特性が最適化されるまでシャッフリングすることができる。或いは、所定の遺伝子を同種または異種のいずれかから得た同一遺伝子ファミリーの相同遺伝子と組み換え、それによって天然に存在する複数の遺伝子の遺伝多様性を、指図された制御可能な方法で最大化させることができる。

20

30

【0173】

別の実施例によれば、KPPをコードする配列は、当分野で周知の化学的方法を用いて、全体或いは一部が合成可能である(例えば、Caruthers, M.H.他 (1980) Nucl. Acids Res. Symp. Ser 7:215-223及びHorn, T.他 (1980) Nucl. Acids Res. Symp. Ser7.225-232を参照)。別法として、化学的方法を用いてKPP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、種々の液相または固相技術を用いてペプチド合成を行うことができる(Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, 55-60ページ、Roberge, J.Y.他 (1995) Science 269:202-204等を参照)。自動合成はABI 431Aペプチドシンセサイザ(Applied Biosystems)を用いて達成し得る。更にKPP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を用いた他のタンパク質の配列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。

40

【0174】

このペプチドは、分離用高速液体クロマトグラフィーを用いて実質的に精製し得る(Chiez, R.M.及びF.Z. Regnier (1990) Methods Enzymol. 182:392-421等を参照)。合成ペプチドの組成は、アミノ酸分析またはシーケンシングによって確認することができる(前出のCreighton, 28-53ページ等を参照)。

【0175】

生物学的に活性なKPPを発現させるために、KPP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

50

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KPP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。このような要素は、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、KPP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開始コドンと、コザック配列などの近傍の配列が含まれる。KPP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るようにすべきである。外来性の翻訳要素及び開始コドンは、様々な天然物及び合成物を起源とし得る。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 (Scharf, D.他 (1994) *Results Probl. Cell Differ.* 20:125-162.等を参照)。

10

【0176】

当業者に周知の方法を用いて、KPPをコードする配列群と、好適な転写及び翻訳制御エレメント群とを持つ発現ベクター群を作製できる。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, Plainville NY, 4章及び8章及び16-17章、Ausubel, F.M.他 (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, 9章、13章及び16章等を参照)。

20

【0177】

種々の発現ベクター/宿主系を利用して、KPPをコードする配列群の保持及び発現ができる。限定するものではないがこのような発現ベクター/宿主系には、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せた細菌などの微生物等や、酵母菌発現ベクターで形質転換させた酵母菌や、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または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プラスミドまたはpBR322プラスミド)で形質転換させた植物細胞系または動物細胞系がある。(前出のSambrook、前出のAusubel、Van Heeke, G.及びS.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509、Engelhard, E.K.他 (1994) *Proc. Natl. Acad. Sci. U* 30
SA 91:3224-3227、Sandig, V.他 (1996) *Hum. Gene Ther.* 7:1937-1945、Takamatsu, N. (1987) *EMBOJ.* 6:307-311、; 『マグローヒル科学技術年鑑』(The McGraw Hill Yearbook of Science and Technology) (1992) McGraw Hill New York NY, 191-196ページ、Logan, J.及びT. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659、Harrington, J.J.他 (1997) *Nat. Genet.* 15:345-355等を参照)。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 40
、Verma, I.M.及び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。

【0178】

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、KPP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、KPPをコードするポリヌクレオチド配列の慣例的なクローニング、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 (Life Technologies) など、多機能の大腸菌ベクターを用い得る。ベクターのマルチクローニング部位に、KPP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、組換え分子を持つ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は、クローニ 50

ングされた配列における *in vitro* 転写、ジデオキシのシークエンシング、ヘルパーファージによる一本鎖のレスキュー、入れ子状態の欠失の生成にも有用であろう (Van Heeke, G. 及び S.M. Schuster (1989) J. Biol. Chem. 264:55035509等を参照)。例えば抗体類の産生などに多量のKPPが必要な場合は、KPPの発現をハイレベルで指示するベクター類が使用できる。例えば、強力な誘導SP6バクテリオファージプロモーターまたは誘導T7バクテリオファージプロモーターを含むベクターが使用できる。

【0179】

KPPの産生には、酵母発現系の使用が可能である。因子、アルコールオキシダーゼ、PGHプロモーター等の構成型或いは誘導型のプロモーターを含む多数のベクターが、出芽酵母菌 (*Saccharomyces cerevisiae*) またはピキア酵母 (*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.他(1987) Methods Enzymol.153:516-544、Scorer, C. A. 他 (1994) Bio/Technology 12:181-184等を参照)。

10

【0180】

植物系を使用してKPPを発現することも可能である。KPPをコードする配列の転写は、ウイルスプロモーター、例えば単独であるいはTMV由来のオメガリーダー配列と組み合わせて用いられるようなCaMV由来の35Sおよび19Sプロモーターによって促進し得る (Takamatsu, N. (1987) EMBO J 6:307-311)。或いは、RUBISCOの小サブユニット等の植物プロモーターまたは熱ショックプロモーターを用いてもよい (例えば、Coruzzi, G. 他 (1984) EMBO J. 3 : 1671-1680、Broglie, R.他 (1984) Science 224 : 838-843、Winter, J.他(1991) Results Probl. Cell Differ. 17 : 85-105等を参照)。これらの構成物は、直接DNA形質転換または病原体を媒介とする形質移入によって、植物細胞内に導入可能である (『マグローヒル科学技術年鑑』(The McGraw Hill Yearbook of Science and Technology) (1992) McGraw Hill New York NY, 191-196ページ等を参照)。

20

【0181】

哺乳動物細胞においては、多数のウイルスベースの発現系を利用し得る。アデノウイルスが発現ベクターとして用いられる場合、後期プロモーターと3連リーダー配列とからなるアデノウイルス転写/翻訳複合体に、KPPをコードする配列群をライゲーションし得る。非必須のE1またはE3領域へウイルスのゲノムを挿入し、宿主細胞でKPPを発現する感染ウイルスを得ることが可能である。(例えば、Logan, J.及びT. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:36553659等を参照)。更に、ラウス肉腫ウイルス (RSV) エンハンサー等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得る。SV40またはEBVをベースにしたベクターを用いてタンパク質を高レベルで発現させることもできる。

30

【0182】

ヒト人工染色体 (HAC) を用いて、プラスミドに含まれ且つプラスミドから発現するものより大きなDNAの断片を輸送することもできる。治療のために約6 kb ~ 10 MbのHACを作製し、従来の送達方法 (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 (Harrington, J.J.他 (1997) Nat. Genet. 15:345-355等を参照)。

40

【0183】

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KPP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、KPP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。ベクターの導入後、選択培地に移す前に強化培地で約1 ~ 2日間細胞を増殖させることができる。選択可能マーカーの目的は選択培地への抵抗性を与えることであり、選択可能マーカーが存在することにより、導入された配列をうまく発現するような細胞の成長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、その細胞型に適した組織培養技術を用いて増殖可能である。

50

【0184】

任意の数の選択系を用いて、形質転換細胞株を回収できる。限定するものではないがこのような選択系には、 tk^{-} 細胞のために用いられる単純ヘルペスウイルスチミジンキナーゼ遺伝子と、 apr^{-} 細胞のために用いられる単純ヘルペスウイルスのアデニンホスホリボシルトランスフェラーゼ遺伝子がある(Wigler, M. 他 (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823等を参照)。また、選択の基礎として代謝拮抗物質、抗生物質或いは除草剤への耐性を用いることができる。例えばdhfrはメトトレキセートに対する耐性を与え、neoはアミノグリコシッドネオマイシン及びG-418に対する耐性を与え、alsはクロルスルフロンに対する耐性を、patはホスフィノトリシンアセチルトランスフェラーゼに対する耐性を各々与える(Wigler, M他 (1980) Proc. Natl. Acad. Sci. USA 77:35 673570、ColbereGarapin, F. 他 (1981) J. Mol. Biol. 150:114等を参照)。その他の選択可能な遺伝子、例えば、代謝のための細胞の必要条件を変えるtrpB及びhisDは、文献に記載されている(Hartman, S.C.およびR.C. Mulligan (1988) Proc. Natl. Acad. Sci. US A 85:80478051等を参照)。可視マーカー、例えばアントシアニン、緑色蛍光タンパク質(GFP; Clontech)、グルクロニダーゼ及びその基質グルクロニド、またはルシフェラーゼ及びその基質ルシフェリン等を用いてもよい。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(Rhodes, C.A. (1995) Methods Mol. Biol. 55:121131等を参照)。

【0185】

マーカー遺伝子発現の存在/不存在によって目的の遺伝子の存在が示唆されても、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、KPPをコードする配列がマーカー遺伝子配列内に挿入された場合、KPPをコードする配列を含む形質転換細胞は、マーカー遺伝子機能の欠落により同定することが可能である。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KPPをコードする配列とタンデム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。

【0186】

一般に、KPPをコードする核酸配列を含み、KPP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。

【0187】

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KPPの発現の検出及び計測のための免疫学的な方法は、当分野で周知である。このような技法の例には、酵素に結合した免疫吸着剤検定法(ELISA)、ラジオイムノアッセイ(RIA)、フローサイトメーター(FACS)などがある。KPP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これ以外のアッセイは、当分野で公知である(Hampton, R. 他(1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Sect. IV, Coligan, J. E.他(1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY、Pound, J.D. (1998) Immunochemical Protocols, Humans Press, Totowa NJ等を参照)。

【0188】

多岐にわたる標識方法及び抱合方法が、当業者に知られており、様々な核酸アッセイおよびアミノ酸アッセイにこれらの方法を用い得る。KPP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プローブを産出する方法には、オリゴ標識化、ニックトランスレーション、末端標識化、または標識ヌクレオチドを用いるPCR法がある。或いは、KPPをコードする配列またはその任

意の断片を、mRNAプローブを産出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。このようなベクターは、当分野において知られており、市販もされており、T7、T3またはSP6等の好適なRNAポリメラーゼ及び標識されたヌクレオチドを加えて、*in vitro*でRNAプローブの合成に用いることができる。このような方法は、例えばAmersham Pharmacia Biotech、Promega (Madison WI)、U.S. Biochemical等から市販されている種々のキットを用いて実行することができる。検出を容易にする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、補助因子、インヒビター、磁気粒子のほか、放射性核種、酵素、蛍光剤、化学発光剤、発色剤等がある。

【0189】

KPPをコードするヌクレオチド配列を用いて形質転換した宿主細胞は、細胞培地からのタンパク質の回収及び発現に適した条件下で培養し得る。形質転換細胞から製造されたタンパク質が分泌されるか細胞内に留まるかは、使用される配列、ベクター、或いはその両者に依存する。KPP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KPP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。

10

【0190】

更に、宿主細胞株の選択は、挿入した配列の発現を調節する能力または発現したタンパク質を所望の形に処理する能力によって行い得る。限定するものではないが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、脂質化及びアシル化がある。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、タンパク質の標的への誘導、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固有の細胞装置及び特徴のある機構を有する種々の宿主細胞（例えばCHO、HeLa、MDCK、MEK293、WI38等）は、American Type Culture Collection (ATCC, Manassas, VA) から入手可能であり、外来タンパク質の正しい修飾及び処理を確実にするように選択し得る。

20

【0191】

本発明の別の実施態様では、KPPをコードする天然の核酸配列、修飾された核酸配列、または組換えの核酸配列を、上記した任意の宿主系内の融合タンパク質の翻訳を生じる或る異種配列にライゲーションし得る。例えば、市販されている抗体を用いて認識可能な異種部分を含むキメラKPPタンパク質は、KPP活性阻害剤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も、市販されている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。限定されるものではないが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)、チオレドキシシン (Trx)、カルモジュリン結合ペプチド (CBP)、6-His、FLAG、c-myc、赤血球凝集素 (HA) がある。GSTは固定化グルタチオン上で、MBPはマルトース上で、Trxはフェニルアルシンオキシド上で、CBPはカルモジュリン上で、そして6-Hisは金属キレート樹脂上で、同族の融合タンパク質の精製を可能にする。FLAG、c-myc及び赤血球凝集素 (HA) は、これらのエピトープ標識を特異的に認識する市販されているモノクローナル抗体及びポリクローナル抗体を用いて、融合タンパク質の免疫親和性精製を可能にする。また、KPPが精製後に異種部分から切断できるように、KPPコード配列と異種タンパク質配列の間にタンパク質分解性切断部位を含めるように融合タンパク質を遺伝子操作することもできる。融合タンパク質の発現及び精製方法は、前出のAusubel (1995) 10章に記載されている。市販されている種々のキットを用いて融合タンパク質の発現及び精製を促進することもできる。

30

40

【0192】

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 (Promega) を用いて *in vitro* で放射能標識したKPPの合成が可能である。これらの系は、T7、T3またはSP6プロモーターと機能的に連結したタンパク質コード配列の転写及び翻訳をカプセルさせる。翻訳は、例えば³⁵Sメチオニンのような放射能標識したアミノ酸前駆体の存在下で起こる。

50

【0193】

本発明のKPPまたはその断片を用いて、KPPに特異結合する化合物をスクリーニングできる。少なくとも1つまたは複数の試験化合物を用いて、KPP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。

【0194】

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KPP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、この化合物は、KPPが結合する天然受容体に、あるいは例えばリガンド結合部位など少なくともこの受容体の或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KPP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、ショウジョバエ、大腸菌からの細胞が含まれる。KPPを発現する細胞、またはKPPを含有する細胞膜断片を試験化合物と接触させて、結合や、KPPまたは該化合物の何れかの、刺激または活性の阻害を分析する。

10

【0195】

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、蛍光色素、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KPPと結合させるステップと、KPP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、アッセイ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、無細胞再構成標本、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。

20

【0196】

本発明のKPPまたはその断片を用いて、KPPの活性を調節する化合物をスクリーニングできる。このような化合物には、アゴニスト、アンタゴニスト、部分的アゴニスト、または逆アゴニスト等が含まれる。一実施例では、KPPの活性が許容される条件下でアッセイを実施し、そのアッセイでは少なくとも1つの試験化合物をKPPと混合し、試験化合物の存在下でKPPの活性を試験化合物不在下でのKPPの活性と比較する。試験化合物の存在下でのKPPの活性の変化は、KPPの活性をモジュレートする化合物の存在を示唆する。別法では、試験化合物を、KPPの活性に適した条件下で、KPPを有する *in vitro*系すなわち無細胞系と混合し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、KPP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、または複数の試験化合物をスクリーニングすることができる。

30

【0197】

別の実施態様では、胚性幹細胞（ES細胞）における相同組み換えを用いて動物モデル系内で、KPP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産生に有用である（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

40

50

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、可能性のある治療薬や毒性薬剤で検査することができる。

【0198】

KPP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)。

【0199】

KPP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、KPP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。形質転換細胞を胚に注入し、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、可能性のある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KPPを乳汁内に分泌するなどKPP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)。

【0200】

(治療)

KPPのある領域とキナーゼおよびホスファターゼの領域との間に、例えば配列およびモチーフの内容における化学的および構造的類似性が存在する。更にKPPを発現する組織の例には、脳、精巣、子宮、乳房、リンパ節の組織があり、表6にも見出すことができる。ゆえに、KPPは心血管疾患、免疫系の疾患、神経の疾患、成長および発達に影響を及ぼす疾患、脂質異常、細胞増殖異常及び癌に関与するように思われる。KPP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、KPPの発現または活性を低下させることが望ましい。また、KPP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、KPPの発現または活性を増大させることが望ましい。

【0201】

従って、一実施例において、KPP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KPP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心血管障害も含まれ、その中には動脈硬化、アテローム性動脈硬化症、高血圧、脈管炎、レイノー病、動脈瘤、動脈解離、静脈瘤、血栓静脈炎および静脈血栓、血管系腫瘍、血栓溶解の合併症、バルーン血管形成術、血管置換術、冠動脈バイパス移植、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、僧帽弁輪部石灰化、僧帽弁逸脱、リウマチ熱及びリウマチ性心疾患、感染性心内膜炎、非細菌性血栓性心内膜炎、全身性紅斑性狼瘡の心内膜炎、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、心臓移植の合併症、先天性肺異常、肺拡張不全、肺うっ血および肺水腫、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、慢性閉塞性肺疾患、肺気腫、慢性気管支炎、気管支喘息、気管支拡張症、細菌性肺炎、ウイルス性肺炎およ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、塵肺症、サルコイド症、特発性肺線維症、剥離性間質性肺炎、過敏性肺炎、肺好酸球増加、閉塞性細気管支炎 器質化肺炎(bronchiolitis obliterans-organizing pneumonia)、びまん性肺出血症候群、グッドパスチャー症候群、特発性肺血鉄症、膠原血管病併発性肺疾患、肺胞タンパク症、肺腫瘍、炎症性および非炎症性胸水、気胸症、胸膜腫瘍、薬物性肺疾患、放射線による肺疾患、および肺移植の合併症などが含まれる。免疫系疾患の中には、後天性免疫不全症候群(AIDS)及びアジソン病、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、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

10

20

30

40

50

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、胎児赤芽球症、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、骨粗しょう症、膝炎、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、血小板減少症、潰瘍性大腸炎、ブドウ膜炎、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、神経障害として、癲癇、虚血性脳血管障害、脳卒中、脳腫瘍、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病その他の錐体外路障害、筋萎縮性側索硬化その他の運動ニューロン障害、進行性神経性筋萎縮症、網膜色素変性症 (色素性網膜炎)、遺伝性運動失調、多発性硬化症その他の脱髄疾患、細菌性およびウイルス性髄膜炎、脳膿瘍、硬膜下膿瘍、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎および神経根炎、ウイルス性中枢神経系疾患と、クールー、クロイツフェルト ヤコブ病およびゲルストマン シュトロイスラー シャインカー症候群を含むプリオン病と、致死性家族性不眠症、神経系の栄養病および代謝病、神経線維腫症、結節硬化症、小脳網膜性血管腫症 (cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、精神遅滞、ダウン症候群を含むその他の中枢神経系発達障害、脳性麻痺、神経骨格異常、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィーその他の神経筋疾患、末梢神経疾患、皮膚筋炎および多発性筋炎と、遺伝性、代謝性、内分泌性および中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分障害、不安障害および統合失調症 (精神分裂病) を含む精神障害と、季節性感情障害 (SAD) と、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、遅発性ジスキネジア、ジストニー、パラノイド精神病、帯状疱疹後神経痛、トゥーレット病、進行性核上麻痺、皮質基底核変性症、家族性前頭側頭痴呆が含まれ、成長および発達に影響を及ぼす疾患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)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、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、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 ツース病及び神経線維腫症などの遺伝性神経病、甲状腺機能低下症、水頭症、舞蹈病 (Sydenham's chorea) 及び脳性小児麻痺などの発作障害、脊髄二分裂、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感覚神経性聴力損失が含まれ、脂質異常の中には、脂肪肝、胆汁うっ滞、原発性胆汁性肝硬変、カルニチン欠乏症、カルニチンパルミトイルトランスフェラーゼ欠乏症、ミオアデニレートデミナーゼ欠乏症、高トリグリセリド血症、ファブリー病などの脂質貯蔵病、ゴーシェ病、ニーマン ピック病、変染色性白質ジストロフィー、副腎性白質ジストロフィー、 G_{M2} ガングリオシドーシス、セロイドリポフスチン症、無リポ蛋白血症、タンジアー病、リポ蛋白過剰血症、糖尿病、脂肪異常栄養症、脂肪腫症、急性皮下脂肪組織炎、播種性脂肪組織壊死症、有痛脂肪症、リポイド副腎過形成、リポイドネフロゼ、微小変化型ネフロゼ症候群、脂肪腫、アテローム性動脈硬化症、高コレステロール血症、高トリグリセリド血症を伴った高コレステロール血症、原発性低リポ蛋白血症、甲状腺症機能低下症、腎臓病、肝疾患、レシチン-コレステロールアシルトランスフェラーゼ欠乏症、脳髄黄色腫症、シトステロール血症、低コレステロール血症、テイ サックス病、サンドホフ病、高脂血症、脂肪過剰血症、脂質筋障害、肥満症があり、また細胞増殖異常として、日光性角化症、動脈硬化、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)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症があり、また癌として腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌等の癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副

10

20

30

40

50

甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌、多発性骨髄腫などの白血病、ホジキン病などのリンパ腫が含まれる。

【0202】

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KPP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、KPP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。

【0203】

更に別の実施態様では、限定するものではないが上に列記した疾患を含む、KPP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KPPを有する組成物を、好適な医薬用キャリアと共に被験者に投与し得る。

10

【0204】

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

【0205】

更なる実施例では、KPP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KPP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、上記した心血管疾患、免疫系の疾患、神経の疾患、成長および発達に影響を及ぼす疾患、脂質異常、細胞増殖異常及び癌が含まれる。一実施態様においては、アンタゴニストとして直接的に、或いはKPPを発現する細胞または組織に薬剤を輸送するターゲティングまたは輸送機構として間接的にKPPと特異結合する抗体を用いることができる。

20

【0206】

別の実施例では、KPPをコードするポリヌクレオチドの相補体を発現するベクターを患者に投与して、限定するものではないが上記した疾患を含むKPPの発現または活性の増大に関連した疾患を治療または予防することも可能である。

【0207】

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、相補配列、またはベクターを、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。併用療法で用いる好適な治療薬は、当業者が従来の医薬原理に従って選択し得る。治療薬と組み合わせることにより、上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。この方法を用いることにより少量の各薬剤で医薬効果をあげることが可能となり、それによって副作用の可能性を低減し得る。

30

【0208】

KPP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し得る。詳しくは、精製されたKPP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KPPと特異的に結合する薬の同定が可能である。KPP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断片、及びFab発現ライブラリによって作られた断片が含まれる。但し、これらに限定されるものではない。中和抗体（即ち二量体の形成を阻害する抗体）は通常、治療用に好適である。

40

【0209】

抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ヒト、その他を含む種々の宿主は、KPP、またはKPPの任意の断片またはオリゴペプチドであって免疫原性の特性を有するものを注入することによって免疫化することができる。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。限定するものではないがこのようなアジュバントには、フロイントアジュバントと、水酸化アルミニウム等のミネラルゲルアジュバントと、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、スカシガイのヘモシニアン、ジニトロフェノール等の界面活性剤とがある。ヒトに用いられるアジュバントの中では、BCG（カルメット グラン杆菌）及びコリネバクテリウム

50

パルバム (*Corynebacterium parvum*) が特に好ましい。

【0210】

KPP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る、一般的には約10個以上のアミノ酸からなるアミノ酸配列を有するものが好ましい。これらのオリゴペプチド、ペプチドまたは断片は、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。KPP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。

【0211】

KPPに対するモノクローナル抗体は、抗体分子を産生する任意の技術を用いて、培地内の連続した細胞株によって作製し得る。限定するものではないがこのような技術には、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ある (Kohler, G. 他 (1975) *Nature* 256:495-497, Kozbor, D. 他 (1985) *J. Immunol. Methods* 81:31-42, Cote, R.J. 他 (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030, Cole, S. P. 他 (1984) *Mol. Cell Biol.* 62:109-120等を参照)。

10

【0212】

更に、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「キメラ抗体」作製のために開発した技術が、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.* 81:68516855, Neuberger, M.S. 他 (1984) *Nature* 312:604-608, Takeda, S. 他 (1985) *Nature* 314:452-454等を参照)。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、KPP特異性一本鎖抗体を生成する。関連特異性を有するがイディオタイプ組成が異なるような抗体を、ランダムな組合せの免疫グロブリンライブラリからチェーンシャッフリングによって産生することもできる (Burton D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137等を参照)。

20

【0213】

抗体の産生は、リンパ球集団における *in vivo* 産生の誘導によって、或いは文献に開示されているように非常に特異的な結合試薬の免疫グロブリンのライブラリまたはパネルのスクリーニングによっても行い得る (Orlandi, R. 他 (1989) *Proc. Natl. Acad. Sci. USA* 86:38333837, Winter, G. 他 (1991) *Nature* 349:293299等を参照)。

30

【0214】

KPPのための特異結合部位を有する抗体断片を産生することもできる。例えば、限定するものではないが、このような断片には、抗体分子のペプシン消化によって産出される $F(ab')_2$ 断片と、 $F(ab')_2$ 断片のジスルフィド架橋を還元することによって産生される Fab断片とがある。あるいは、Fab発現ライブラリを作製することによって、所望の特異性を持つモノクローナルFab断片を迅速且つ容易に同定することが可能となる (Huse, W.D. 他 (1989) *Science* 246:12751281等を参照)。

【0215】

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ことができる。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結合試験、または免疫放射定量測定法のための数々のプロトコルが、当分野では周知である。通常このようなイムノアッセイには、KPPとその特異性抗体との間の複合体形成の計測が含まれる。二つの非干渉性KPPエピトープに対して反応性のモノクローナル抗体を用いる、2部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合的結合アッセイも利用することができる (Pound、前出)。

40

【0216】

ラジオイムノアッセイ技術と共に Scatchard 分析などの様々な方法を用いて、KPPに対する抗体の親和性を評価する。親和性は結合定数 K_a で表す。 K_a は、平衡状態において KPP 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た値であると定義する。多数の KPP 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試薬の K_a は、KPP に対する

50

抗体の平均親和性または結合活性を表す。特定のKPPエピトープに単一特異的なモノクローナル抗体試薬の K_a は、親和性の真の測定値を表す。 K_a 値が $10^9 \sim 10^{12}$ liter/mo l の高親和性抗体試薬は、KPP抗体複合体が過酷な処理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 $10^6 \sim 10^7$ L/mo l の低親和性抗体試薬は、KPPが抗体から最終的に活性化状態で解離する必要がある免疫精製 (immunopurification) 及び類似の処理に用いるのが好ましい (Catty, D. (1988) Antibodies, Volume I: A Practical Approach. IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

【0217】

ポリクローナル抗体試薬の抗体価及び結合活性を更に評価して、後に使う或る適用例に対するこのような試薬の品質及び適性を決定することができる。例えば、少なくとも $1 \sim 2$ mg/ml の特異的な抗体、好ましくは $5 \sim 10$ mg/ml の特異的な抗体を持つポリクローナル抗体試薬は一般に、KPP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。抗体の特異性、抗体価、結合活性、様々な適用例における抗体の品質や使用に対する指針については、一般に入手可能である (前出のCattyの文献、同Coligan他の文献等を参照)。

【0218】

本発明の別の実施例では、KPPをコードするポリヌクレオチド、またはその任意の断片や相補配列が、治療目的で使用することができる。ある実施態様では、KPPをコードする遺伝子のコーディング領域や調節領域に相補的な配列やアンチセンス分子 (DNA、RNA、PNAあるいは修飾オリゴヌクレオチド) 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、アンチセンスオリゴヌクレオチドまたはより大きな断片が、KPP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。(Agrawal, S. 編集 (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ等を参照)。

【0219】

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作製する発現プラスミドの形で、細胞内に送達し得る (Slater, J.E. 他 (1998) *J. Allergy Clin. Immunol.* 102(3):469-475、Scanlon, K.J. 他 (1995) 9(13):1288-1296.等を参照)。アンチセンス配列はまた、例えばレトロウイルスやアデノ随伴ウイルスベクターなど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 (Miller, A.D. (1990) *Blood* 76:271、前出のAusubel、Uckert, W. 及び W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347等を参照)。その他の遺伝子送達機構には、リポソーム系、人工的なウイルスエンベロープ及び当分野で公知のその他の系が含まれる (Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225、Boado, R.J. 他 (1998) *J. Pharm. Sci.* 87(11):1308-1315、Morris, M.C. 他 (1997) *Nucleic Acids Res.* 25(14):2730-2736等を参照)。

【0220】

本発明の別の実施態様では、KPP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。遺伝子治療を行うことにより、(i) 遺伝子欠損症 (例えば X 染色体鎖遺伝 (Cavazzana-Calvo, M. 他 (2000) *Science* 288:669-672) により特徴付けられる重度の複合型免疫欠損 (SCID)-X1 の場合)、先天性アデノシンデアミナーゼ (ADA) 欠損症に関連する重度の複合型免疫欠損 (Blaese, R.M. 他 (1995) *Science* 270:475-480、Bordignon, C. 他 (1995) *Science* 270:470-475)、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 因子欠損に起因する血友病 (Crystal, R.G. (1995) *Science* 270:404-410、Verma, I.M. 及び Somia, N. (1

10

20

30

40

50

997) Nature 389:239-242) を治療し、(ii) 条件的致死性遺伝子産物を発現させ(例えば制御不能な細胞増殖に起因する癌の場合)、あるいは(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、Poesebla, E. 他(1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399)、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、並びに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ることができる。KPP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KPP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。

【0221】

本発明の更なる実施例では、KPPの欠損による疾患や異常症を、KPP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KPP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導入技術には、(i) 個々の細胞内への直接的なDNA微量注射法、(ii) 遺伝子銃、(iii) リポソームを介した形質移入、(iv) 受容体を介した遺伝子導入、及び(v) DNAトランスポソンの使用(Morgan, R.A.およびW.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217、Ivics, Z. (1997) Cell 91:501-510、Boulay, J-L.およびH. Recipon (1998) Curr. Opin. Biotechnol. 9:445-450)がある。

【0222】

KPP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、PCR2-TOPOTAベクター(Invitrogen, Carlsbad CA)、PCMV-SCRIPT、PCMV-TAG、PEGSH/PERV (Stratagene, La Jolla CA)、PTET-OFF、PTET-ON、PTRE 2、PTRE2-LUC、PTK-HYG (Clontech, Palo Alto CA)が含まれる。KPPを発現させるために、(i) 恒常的に活性なプロモーター(例えば、サイトメガロウイルス(CMV)、ラウス肉腫ウイルス(RSV)、SV40ウイルス、チミジンキナーゼ(TK)、若しくは - アクチン遺伝子等)、(ii) 誘導性プロモーター(例えば、市販されているT-REXプラスミド(Invitrogen)に含まれている、テトラサイクリン調節性プロモーター(Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen, M. 他(1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456)、エクジソン誘導性プロモーター(市販されているプラスミドPVGRXR及びPIND 30に含まれている: Invitrogen)、FK506/ラパマイシン誘導性プロモーター、またはRU486/ミフェプリストーン誘導性プロモーター(Rossi, F.M.V. and H.M. Blau, 前出)、または(iii) 正常な個体に由来するKPP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。

【0223】

市販のリポソーム形質転換キット(例えばInvitrogen社のPerFect Lipid 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になる。別法では、リン酸カルシウム法(Graham, F.L.及びA. J. Eb (1973) Virology 52:456-467)若しくは電気穿孔法(Neumann, B.他(1982) EMBO J. 1:841-845)を用いて形質転換を行う。初代培養細胞にDNAを導入するためには、標準化された哺乳動物の形質移入プロトコルの修飾が必要である。 40

【0224】

本発明の別の実施例では、KPP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KPP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PF B及びPFBNE0)はStratagene社から市販されており、刊行データ(Riviere, I.他(1995) Proc. Natl. Acad. Sci. U.S.A. 92:6733-6737)に基づいている。上記データを引用す 50

10

20

30

40

50

ることをもって本明細書の一部とする。ベクターは、好適なベクター産生細胞系 (VPCL) において増殖され、VPCL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またはVSVg等の汎親和性エンベロープタンパク質を発現する (Armentano, D. 他 (1987) J. Virol. 61:1647-1650、Bender, M.A. 他 (1987) J. Virol. 61:1639-1646、Adam, M.A.及び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)。RIGGに付与された米国特許第5,910,434号 (「Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant」) において、レトロウイルスパッケージング細胞系を得るための方法が開示されており、引用することをもって本明細書の一部とする。レトロウイルスベクターの増殖、細胞集団 (例えばCD4⁺ T細胞) の形質導入、及び形質導入した細胞の患者への戻しは、遺伝子治療の分野では当業者に公知の方法であり、多数の文献に記載されている (Ranga, U. 他 (1997) J. Virol. 71:7020-7029、Bauer, G.他 (1997) Blood 89:2259-2267、Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716、Ranga, U.他 (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206、Su, L. (1997) Blood 89:2283-2290)。

10

【0225】

別法では、アデノウイルス系遺伝子治療の或る送達系を用いて、KPPの発現に関連する1つ以上の遺伝子異常を有する細胞群に、KPPをコードするポリヌクレオチド群を送達する。アデノウイルス系ベクターの作製及びパッケージングについては、当業者に公知である。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内に導入するために可変性であることが証明された (Csete, M.E.他 (1995) Transplantation 27:263-268)。使用できる可能性のあるアデノウイルスベクターは、Armentanoに付与された米国特許第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、Antinozzi, P.A. 他 (1999) Annu. Rev. Nutr. 19:511-544、Verma, I.M.及びN. Somia (1997) Nature 18:389:239-242も参照されたい。両文献は、引用することをもって本明細書の一部とする。

20

【0226】

別法では、ヘルペス系遺伝子治療の送達系を用いて、KPPの発現に関して1以上の遺伝子異常を持つ標的細胞に、KPPをコードするポリヌクレオチド類を送達する。単純疱疹ウイルス (HSV) 系のベクターの利用は、HSVが向性を持つ中枢神経細胞にKPPを導入する際に特に有用たり得る。ヘルペス系ベクターの作製及びパッケージングは、当業者に公知である。複製適格性単純ヘルペスウイルス (HSV) I型系のベクターは、レポーター遺伝子を霊長類の眼に送達するために用いられてきた (Liu, X. 他 (1999) Exp. Eye Res. 169:385-395)。HSV-1ウイルスベクターの作製についても、DeLucaに付与された米国特許第5,804,413号 (「Herpes simplex virus swains for gene transfer」) に関示されており、該特許の引用をもって本明細書の一部とする。米国特許第5,804,413号には、ヒト遺伝子治療を含む目的のために好適なプロモーターの制御下において細胞に導入される少なくとも1つの外在性遺伝子を有するゲノムを含む組換えHSV d92の使用についての記載がある。上記特許はまた、ICP4、ICP27及びICP22のために除去される組換えHSV系統の作製及び使用について開示している。HSVベクターについては、Goins, W.F.他 (1999) J. Virol. 73:519-532及びXu, H. 他 (1994) Dev. Biol. 163:152-161も参照されたい。両文献は、引用をもって本明細書の一部とする。クローン化ヘルペスウイルス配列の操作、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形質移入した後の組換えウイルスの産生、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、当業者に公知の技術である。

30

40

【0227】

別法では、ウイルス (正の一本鎖RNAウイルス) ベクターを用いてKPPをコードするポリヌクレオチドを標的細胞に送達する。プロトタイプのウイルスであるセムリキ森林熱ウイルス (SFV) の生物学的研究が広範に行われており、遺伝子導入ベクターがSFVゲノムに

50

基づいていることが分かった (Garoff, H. 及び K.-J. Li (1998) *Cun. Opin. Biotech.* 9: 464-469)。 ウイルスRNAの複製中に、通常はウイルスのキャプシッ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は、完全長のゲノムRNAより高いレベルに複製されるため、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比べてキャプシッドタンパク質が過剰産生される。同様に、KPPに対するコード配列をキャプシッド領域の ウイルスゲノムに導入することにより、ベクター導入細胞において多数のKPPコードRNAが産生され、高レベルのKPPが合成される。通常は ウイルスの感染が数日以内での細胞溶解に関係する一方で、シンドビスウイルス (SIN) の変異体を有するハムスター正常腎臓細胞 (BHK-21) の持続的な感染を確立する能力は、 ウイルスの溶解複製を遺伝子治療に適用できるように好適に変更可能であることを示唆している (Dryga, S.A. 他 (1997) *Virology* 228 :74-83)。広範な宿主に ウイルスを導入できるので、様々なタイプの細胞にKPPを導入できることとなる。或る集団におけるサブセットの細胞の特定形質導入は、形質導入前に細胞の選別を必要とし得る。 ウイルスの感染性cDNAクローンの処置方法、 ウイルスのcDNA及びRNAの形質移入方法及び ウイルスの感染方法は、当業者に公知である。

【0228】

転写開始部位由来のオリゴヌクレオチドを用いて遺伝子発現を阻害することも可能である。転写開始部位とは例えば開始部位から数えて約 - 10 と約 + 10 の間である。同様に、三重らせん塩基対の形成方法を用いて阻害が可能となる。三重らせん塩基対形成は、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開くような二重らせんの能力を阻害するので、三重らせん塩基対形成は有用である。三重らせんDNAを用いる最近の治療の進歩については文献に記載がある (Gee, J.E. 他 (1994) in: Huber, B.E. 及び B.I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co., Mt. Kisco, NY, 163-177 ページ等を参照)。相補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の翻訳を阻止するべく設計することができる。

【0229】

リボザイムは酵素性RNA分子であり、RNAの特異的切断を触媒するためにリボザイムを用いることもできる。リボザイム作用のメカニズムは、相補的標的RNAへのリボザイム分子の配列特異的ハイブリダイゼーションと、その後起こる内ヌクレオチド鎖切断に参与している。例えば、人為操作されたハンマーヘッド型リボザイム分子は、KPPをコードする配列の内ヌクレオチド鎖切断を、特異的且つ効果的に触媒する可能性がある。

【0230】

任意の潜在的RNA標的内の特異的リボザイム切断部位は、GUA、GUU、GUC配列を含めたりボザイム切断部位に対する標的分子をスキャンすることによって先ず同定される。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15~20リボヌクレオチドの短いRNA配列が、そのオリゴヌクレオチドを機能不全にするような2次構造の特徴をもっていないかを評価することが可能になる。候補標的の適合性の評価も、リボヌクレアーゼ保護アッセイを用いて相補的オリゴヌクレオチドとのハイブリダイゼーションのアクセス可能性をテストすることによって行うことができる。

【0231】

本発明の相補リボ核酸分子及びリボザイムは、核酸分子合成のために当分野でよく知られている任意の方法を用いて作製し得る。任意の方法には、固相フォスフォアミダイト化学合成等のオリゴヌクレオチドを化学的に合成する方法がある。或いは、KPPをコードするDNA配列の *in vitro* 及び *in vivo* 転写によってRNA分子を産出し得る。このようなDNA配列は、T7やSP6等の好適なRNAポリメラーゼプロモーターを用いて多様なベクター内に取り込むことが可能である。或いは、相補的RNAを構成的或いは誘導的に合成するようなこれらcDNA産物を、細胞系、細胞または組織内に導入することができる。

【0232】

細胞内の安定性を高め、半減期を長くするためにRNA分子を修飾することができる。限定するものではないが可能な修飾には、分子の5'末端、3'末端、あるいはその両方におい

てフランキング配列を追加したり、分子の主鎖内においてホスホジエステラーゼ結合ではなくホスホオキチレートまたは2'-O-メチルを使用したりすることが含まれる。この概念は、PNAの産出に固有のものであり、これら全ての分子に拡大することができる。それには、内因性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、及びウリジンにアセチル -、メチル -、チオ - 及び同様の修飾をしたものの他、非従来型塩基、例えばイノシン、クエオシン (queosine)、ワイプトシン (wybutosine) 等を加えることでできる。

【0233】

本発明の更なる実施例は、KPP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。限定するものではないが特異ポリヌクレオチドの発現変化を起こすのに有効な化合物には、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、転写因子その他のポリペプチド転写制御因子、及び特異ポリヌクレオチド配列と相互作用し得る非高分子化学的実体がある。有効な化合物は、ポリヌクレオチド発現のインヒビターまたはエンハンサーのいずれかとして作用することによりポリヌクレオチド発現を変異し得る。従って、KPP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、KPP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、KPP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、KPP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。

【0234】

特異ポリヌクレオチドの変異発現における有効性に対して、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し得る。試験化合物は、当分野で通常知られている任意の方法により得られる。このような方法には、ポリヌクレオチドの発現を変異させる場合と、既存の、市販のまたは専売の、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と、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く化合物を合理的にデザインする場合と、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ことが知られているような化合物の化学修飾がある。KPPをコードするポリヌクレオチドを持つサンプルを、このようにして得た試験化合物の少なくとも1つに曝露する。サンプルには例えば、無傷細胞、透過処理した細胞、無細胞再構成系または再構成生化学系があり得る。KPP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、KPP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量を定量し、それによって1つ若しく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対する基礎を形成し得る。試験化合物に曝露されるポリヌクレオチドの発現における変化の検出は、ポリヌクレオチドの発現を変異する際に試験化合物が有効であることを示している。特異ポリヌクレオチドの変異発現に有効な化合物に対して、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 細胞等のヒト細胞系 (Clarke, M.L. 他 (2000) *Biochem. Biophys. Res. Commun.* 268:8-13) を用いてスクリーニングを実行する。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のための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、修飾オリゴヌクレオチド) の組み合わせライブラリをスクリーニングすることに関与している (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)。

【0235】

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可能であり、*in vivo*、*in vitro* 及び *ex vivo* の使用に対して同程度に適している。*ex vivo* 治療の場合、ベクターを患者から採取した幹細胞内に導入し、クローニング増殖して同一患者に自家移植で戻すことができる。トランスフェクション、リボソーム注入またはポリカチオンアミノポリマーによる

10

20

30

40

50

輸送は、当分野でよく知られている方法を用いて実行することができる (Goldman, C.K. 他 (1997) Nat. Biotechnol. 15:462-466.等を参照)。

【0236】

上記の治療方法はいずれも、例えば、ヒト、イヌ、ネコ、ウシ、ウマ、ウサギ、サル等の哺乳動物を含めて治療が必要な全ての対象に適用できる。

【0237】

本発明の追加実施例は、通常薬剤として許容できる賦形剤で処方される活性成分を有する組成物の投与に関連する。賦形剤には例えば、糖、でんぷん、セルロース、ゴム及びタンパク質がある。様々な処方通常知られており、詳細は Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA) の最新版に記載されている。このような組成物は、KPP、KPPの抗体、擬態物質、アゴニスト、アンタゴニスト、またはインヒビターなどからなる。

10

【0238】

本発明に用いられる組成物は、任意の数の経路によって投与することができ、限定するものではないが経路には、経口、静脈内、筋肉内、動脈内、骨髄内、クモ膜下腔内、心室内、肺、経皮、皮下、腹腔内、鼻腔内、腸内、局所、舌下または直腸がある。

【0239】

肺から投与する組成物は、液状または乾燥粉末状で調製し得る。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 (例えば伝統的な低分子量有機薬) の場合には、速効製剤のエアロゾル送達は当分野で公知である。高分子 (例えばより大きなペプチド及びタンパク質) の場合には、当該分野において肺の肺泡領域を介しての肺送達が最近向上したことにより、インスリン等の薬剤を実質的に血液循環へ輸送することを可能にした (Patton, J.S. 他, 米国特許第5,997,848号等を参照)。肺送達は、針注射なしに投与する点で優れており、有毒な可能性のある浸透エンハンサーの必要性をなくす。

20

【0240】

本発明での使用に適した組成物には、所定の目的を達成するために必要なだけの量の活性成分を含有する組成物が含まれる。有効投与量の決定は、当業者の能力の範囲内で行う。

【0241】

KPP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送達を促進し得る。別法では、KPP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。このようにして生成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することがわかっている (Schwarze, S.R. 他 (1999) Science 285:1569-1572)。

30

【0242】

任意の化合物に対して、細胞培養アッセイ、例えば新生物性細胞の細胞培養アッセイにおいて、或いは、動物モデル、例えばマウス、ラット、ウサギ、イヌ、サルまたはブタ等において、先ず治療の有効投与量を推定することができる。動物モデルはまた、好適な濃度範囲及び投与経路を決定するためにも用い得る。このような情報を用いて、次にヒトに対する有益な投与量及び投与経路を決定することができる。

40

【0243】

治療有効量は、症状や容態を回復させる活性成分の量、たとえばKPPまたはその断片、KPPの抗体、KPP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量を指す。薬用有効度及び毒性は、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、例えばED₅₀ (集団の50%の医薬的有効量) またはLD₅₀ (集団の50%の致死量) を測定するなどして決定することができる。毒性効果の治療効果に対する投与量の比は、治療指数であり、LD₅₀/ED₅₀比として表すことができる。高い治療指数を示すような組成物が望ましい。細胞培養アッセイ及び動物実験から得られたデータは、ヒトに用いるための投与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投与量は、毒性を殆ど或いは全く含まず、ED₅₀を含むような血中濃度の範囲にあることが好ましい。用いられる投与形態

50

、患者の感受性及び投与の経路によって、投与量はこの範囲内で様々に変わる。

【0244】

正確な投与量は、治療が必要な被験者に関する要素を考慮して、現場の医者が決定することになる。効果的なレベルの活性成分を与え、或いは所望の効果を維持するべく、投与量及び投与を調節する。被験者に関する要素としては、疾患の重症度、患者の通常の状態、患者の年齢、体重及び性別、投与の時間及び頻度、薬剤の配合、反応感受性及び治療に対する応答等を考慮する。作用期間が長い組成物は、特定の製剤の半減期及びクリアランス率によって3～4日毎に1度、1週間に1度、或いは2週間に1度の間隔で投与し得る。

【0245】

通常投与量は、投与の経路にもよるが約0.1～100,000 µgであり、合計で約1gまでとする。特定の投与量及び送達方法に関するガイダンスは文献に記載されており、現場の医者は通常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に対する処方とは異なる、ヌクレオチドに対する処方を利用することになる。同様に、ポリヌクレオチドまたはポリペプチドの送達は、特定の細胞、状態、位置等に特異的なものとなる。

【0246】

(診断)

別の実施例では、KPPに特異的に結合する抗体が、KPPの発現によって特徴付けられる疾患の診断、またはKPPやKPP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。診断目的に有用な抗体は、上記の治療の箇所で記載した方法と同じ方法で調合される。KPP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KPPを検出する方法が含まれる。この抗体は修飾されたものもされていないものも可能であり、レポーター分子との共有結合または非共有結合で標識化できる。レポーター分子としては広くさまざまな種類が本分野で知られており、また使用可能であるが、そのうちのいくつかは上記で説明されている。

【0247】

KPPを測定するための様々なプロトコル、例えばELISA、RIA、FACS等が本技術分野において知られており、KPP発現の修正レベル或いは異常レベルを診断する基準を提供する。複合体の形成に適した条件下でヒト対象等の正常な哺乳動物対象から採取した体液または細胞とKPPに対する抗体とを結合させることにより、KPP発現の正常値または標準値が決定される。標準複合体形成量は、種々の方法、例えば測光法で定量できる。被験者、対照および生検組織からの疾患サンプルで発現するKPPの量を標準値と比較する。標準値と被験者との偏差が疾患を診断するパラメータとなる。

【0248】

本発明の別の実施態様では、KPPをコードするポリヌクレオチドを、診断のために用い得る。用いることができ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RNA及びDNA分子、そしてPNAが含まれる。このポリヌクレオチドを用いて、疾患と関連し得るKPP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。診断アッセイは、KPPの不在、存在及び過剰発現を測定するために、そして治療インターベンション中にKPPレベルの調製をモニターするために用いることができる。

【0249】

一実施形態では、例えばKPPをコードまたは近縁の分子群をコードするゲノム配列群など、ポリヌクレオチド配列を検出できるPCRプローブとのハイブリダイゼーションを用いて、KPPをコードする核酸配列群を同定できる。プローブが高度に特異的な領域(例えば5'調節領域)から作られている、或いはやや特異性の低い領域(例えば保存されたモチーフ)から作られているかにかかわらず、そ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によって、そのプローブがKPPをコードする自然界の配列のみを同定するかどうか、或いは対立遺伝子や関連配列をコードする自然界の配列のみを同

10

20

30

40

50

定するかどうかが決まるであろう。

【0250】

プローブはまた、関連する配列の検出にも用いることができ、その配列はKPPをコードする任意の配列と少なくとも50%の相同性を有し得る。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:9-16の配列、或いはKPP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。

【0251】

KPPをコードするDNAに対する特異的ハイブリダイゼーションプローブを作製する方法には、KPPまたはKPP誘導体をコードするポリヌクレオチド配列を、mRNAプローブを作製するためのベクターにクローニングする方法が含まれる。mRNAプローブ作製のためのベクターは、当業者に知られており、市販されており、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得る。ハイブリダイゼーションプローブは、種々のレポーターの集団によって標識され得る。レポーター集団の例としては、³²Pまたは³⁵S等の放射性核種、或いはアビジン/ビオチン結合系を介してプローブに結合されたアルカリホスファターゼ等の酵素標識などが挙げられる。

【0252】

KPPをコードするポリヌクレオチド配列は、KPPの発現に関係する疾患の診断の為に用い得る。限定するものではないが、このような疾患には心血管障害も含まれ、その中には動脈脈瘤、アテローム性動脈硬化症、高血圧、脈管炎、レイノー病、動脈瘤、動脈解離、静脈瘤、血栓静脈炎および静脈血栓、血管系腫瘍、血栓溶解の合併症、バルーン血管形成術、血管置換術、冠動脈バイパス移植、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、僧帽弁輪部石灰化、僧帽弁逸脱、リウマチ熱及びリウマチ性心疾患、感染性心内膜炎、非細菌性血栓性心内膜炎、全身性紅斑性狼瘡の心内膜炎、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、心臓移植の合併症、先天性肺異常、肺拡張不全、肺うっ血および肺水腫、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、慢性閉塞性肺疾患、肺気腫、慢性気管支炎、気管支喘息、気管支拡張症、細菌性肺炎、ウイルス性肺炎およ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、塵肺症、サルコイド症、特発性肺線維症、剥離性間質性肺炎、過敏性肺炎、肺好酸球増加、閉塞性細気管支炎 器質化肺炎 (bronchiolitis obliterans-organizing pneumonia)、びまん性肺出血症候群、グッドパスチャー症候群、特発性肺血鉄症、膠原血管病併発性肺疾患、肺胞タンパク症、肺腫瘍、炎症性および非炎症性胸水、気胸症、胸膜腫瘍、薬物性肺疾患、放射線による肺疾患、および肺移植の合併症などが含まれる。免疫系疾患の中には、後天性免疫不全症候群 (AIDS) 及びアジソン病、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、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、胎児赤芽球症、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、骨粗しょう症、瘧疾、多発性筋炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、血小板減少症、潰瘍性大腸炎、ブドウ膜炎、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、神経障害として、癲癇、虚血性脳血管障害、脳卒中、脳腫瘍、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病その他の錐体外路障害、筋萎縮性側索硬化症その他の運動ニューロン障害、進行性神経性筋萎縮症、網膜色素変性症 (色素性網膜炎)、遺伝性運動失調、多発性硬化症その他の脱髄疾患、細菌性およびウイルス性髄膜炎、脳膿瘍、

10

20

30

40

50

硬膜下膿瘍、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎および神経根炎、ウイルス性中枢神経系疾患と、クールー、クロイツフェルト ヤコブ病およびゲルストマン シュトロイスラー シャインカー症候群を含むプリオン病と、致死性家族性不眠症、神経系の栄養病および代謝病、神経線維腫症、結節硬化症、小脳網膜性血管腫症 (cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、精神遅滞、ダウン症候群を含むその他の中枢神経系発達障害、脳性麻痺、神経骨格異常、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィーその他の神経筋疾患、末梢神経疾患、皮膚筋炎および多発性筋炎と、遺伝性、代謝性、内分泌性および中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分障害、不安障害および統合失調症 (精神分裂病) を含む精神障害と、季節性感情障害 (SAD) と、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、遅発性ジスキネジア、ジストニー、パラノイド精神病、帯状疱疹後神経痛、トゥーレット病、進行性核上麻痺、皮質基底核変性症、家族性前頭側頭痴呆が含まれ、成長および発達に影響を及ぼす疾患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)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、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、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 ツース病及び神経線維腫症などの遺伝性神経病、甲状腺機能低下症、水頭症、舞蹈病 (Sydenham's chorea) 及び脳性小児麻痺などの発作障害、脊髄二分裂、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感覚神経性聴力損失が含まれ、脂質異常の中には、脂肪肝、胆汁うっ滞、原発性胆汁性肝硬変、カルニチン欠乏症、カルニチンパルミトイルトランスフェラーゼ欠乏症、ミオアデニレートデミナーゼ欠乏症、高トリグリセリド血症、ファブリー病などの脂質貯蔵病、ゴーシェ病、ニーマン ピック病、変染色性白質ジストロフィー、副腎性白質ジストロフィー、 G_{M2} ガングリオシドーシス、セロイドリポフスチン症、無リポ蛋白血症、タンジアー病、リポ蛋白過剰血症、糖尿病、脂肪異栄養症、脂肪腫症、急性皮下脂肪組織炎、播種性脂肪組織壊死症、有痛脂肪症、リポイド副腎過形成、リポイドネフローゼ、微小変化型ネフローゼ症候群、脂肪腫、アテローム性動脈硬化症、高コレステロール血症、高トリグリセリド血症を伴った高コレステロール血症、原発性低リポ蛋白血症、甲状腺機能低下症、腎臓病、肝疾患、レシチン-コレステロールアシルトランスフェラーゼ欠乏症、脳髄黄色腫症、シトステロール血症、低コレステロール血症、テイ サックス病、サンドホフ病、高脂血症、脂肪過剰血症、脂質筋障害、肥満症があり、また細胞増殖異常として、日光性角化症、動脈硬化、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)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症があり、また癌として腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌等の癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌、多発性骨髄腫などの白血病、ホジキン病などのリンパ腫が含まれる。KPPをコードするポリヌクレオチド配列は、変容したKPP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、サザン法やノーザン法、ドットプロット法、或いはその他の膜系の技術、PCR法や、ディップスティック (dipstick)、ピン (pin)、およびマルチフォーマットのELISA式アッセイ、及びマイクロアレイに使用することが可能である。このような定性方法または定量方法は、当分野で公知である。

10

20

30

40

【0253】

或る形態では、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、KPPをコードするヌクレオチド配列が有用であり得る。KPPをコードするヌクレオチド配列は標準法で標識化され、ハイブリダイゼーション複合体の形成に好適な条件下で、患者から採取した体液または組織のサンプルに添加することができる。好適なインキュベーション期間が経過したらサンプルを洗浄し、シグナルを定量して標準値と比較する。患者サンプルの

50

シグナル量が制御サンプルと比べて著しく変化している場合は、サンプル内のKPPをコードするヌクレオチド配列の変異レベルは関連する疾患の存在を示している。このようなアッセイは、動物実験、臨床試験における特定の治療効果を推定するため、或いは個々の患者の治療をモニターするために用いることもできる。

【0254】

KPP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下で、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、KPP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。実質的に精製されたポリヌクレオチドを既知量用いて行った実験から得た値を正常な被験者から得た値と比較することにより、標準ハイブリダイゼーションを定量することができる。このようにして得た標準値は、疾患の徴候を示す患者から得たサンプルから得た値と比較することができる。標準値からの偏差を用いて疾患の存在を確定する。

10

【0255】

疾患の存在が確定されて治療プロトコルが開始されると、患者の発現レベルが正常な被検者に観察されるレベルに近づき始めたかどうかを測定するため、ハイブリダイゼーションアッセイを通常ペースで繰り返し得る。連続アッセイから得られた結果を用いて、数日から数ヶ月の期間にわたる治療の効果を示し得る。

【0256】

癌に関しては、個体からの生体組織における異常な量の転写物（過少発現または過剰発現）の存在は、疾患の発生素質を示したり、実際に臨床的症狀が現れる前に疾患を検出する方法を提供したりし得る。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または積極的な治療法を早くから利用し、それによって癌の発生または更なる進行を防止することが可能となる。

20

【0257】

KPP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、更なる診断への利用には、PCRの利用が含まれ得る。これらのオリゴマーは、化学的に合成するか、酵素により生産するか、或いは *in vitro* で産出し得る。オリゴマーは、好ましくは、KPPをコードするポリヌクレオチドの断片、或いはKPPをコードするポリヌクレオチドに相補的なポリヌクレオチドの断片を有し、最適化した条件下で、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条件下で、近縁のDNA或いはRNA配列の検出、定量、或いはその両方のため用いることが可能である。

30

【0258】

或る実施態様では、KPP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 (SNP) を検出し得る。SNPは、多くの場合にヒトの先天性または後天性遺伝病の原因となるような置換、挿入及び欠失である。限定するものではないがSNPの検出方法には、SSCP (single-stranded conformation polymorphism) 及び蛍光SSCP (fSSCP) 法がある。SSCPでは、KPPをコードするポリヌクレオチド配列に由来する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 (PCR) でDNAを増幅する。DNAは例えば、病変組織または正常組織、生検サンプル、体液その他に由来し得る。DNA内のSNPは、一本鎖形状のPCR生成物の2次及び3次構造に差異を生じさせる。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。fSSCPでは、オリゴヌクレオチドプライマーを蛍光性に標識する。それによってDNAシーケンシング機などの高処理機器でアンプリマー (amplimer) の検出が可能になる。更に、インシリコSNP (in silico SNP, isSNP) と呼ばれる配列データベース分析法は、一般的なコンセンサス配列に配列されるような個々のオーバーラップするDNA断片の配列を比較することにより、多型性を同定し得る。これらのコンピュータベースの方法は、DNAの実験室での調整及び統計モデル及びDNA配列クロマトグラムの自動分析を用いたシーケンシングのエラーに起因する配列の変異をフィルタリングして除去する。別の態様では、例えば高処理MASSARRAYシステム (Sequenom, Inc., San Diego CA) を用いた質量分析によりSNPを検出し、特徴付ける

40

50

。

【0259】

ヒト疾患の遺伝的根拠を研究するためにSNPを用いることができる。例えば、少なくとも16のありふれたSNPが非インスリン依存型真性糖尿病と関連がある。SNPは、嚢胞性線維症、鎌状赤血球貧血あるいは慢性肉芽腫症等の単一遺伝子病の疾患の現れ方の差異を研究する面でも有用である。例えば、マンノース結合レクチンの変異体であるMBL2は、嚢胞性線維の肺での有害な現れ方と関連することが示されてきた。SNPはまた、生命に関わる毒性等の薬剤への患者の反応に影響する遺伝変異体の同定という薬理ゲノミクスにおいても有用性がある。例えば、ALOX5遺伝子のコア・プロモーターにおける変異は5-リポキシゲナーゼ経路を標的とする抗喘息剤を用いた治療への臨床反応の減少につながるが、Nアセチルトランスフェラーゼの変異は抗結核薬剤イソニアジドに反応する末梢神経障害の高発生率と関連する。異なる集団におけるSNP分布の分析は、集団の起源と移動の追跡以外にも遺伝的浮動、突然変異、組換え、選択の調査において有用である(Taylor, J.G他(2001) Trends Mol. Med. 7:507-512、Kwok, P.-Y.及びZ. Gu(1999) Mol. Med. Today 5:538-543、Nowotny, P.他(2001) Curr. Opin. Neurobiol. 11:637-641)。

10

【0260】

KPPの発現を定量するために用い得る方法には、ヌクレオチドの放射標識またはビオチン標識、調節核酸の相互増幅(coamplification)及び標準曲線から得た結果の補間もある(例えば、Melby, P.C.他(1993) J. Immunol. Methods, 159:235-244、Duplaa, C.他(1993) Anal. Biochem. 212: 229-236を参照)。目的のオリゴマーまたはポリヌクレオチドが種々の希釈液中に存在し、分光光度法または比色反応によって定量が迅速になるような高処理フォーマットのアッセイを行うことによって、複数のサンプルの定量速度を加速することができる。

20

【0261】

更に別の実施例では、本明細書に記載した任意のポリヌクレオチド配列由来の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エレメントとして用いることができる。多数の遺伝子の関連発現レベルを同時にモニターする転写イメージング技術にマイクロアレイを用いることが可能である。これについては、以下に記載する。マイクロアレイはまた、遺伝変異体、突然変異及び多型性の同定に用いることができる。この情報を用いることで、遺伝子機能を決定し、疾患の遺伝的根拠を理解し、疾患を診断し、遺伝子発現の機能としての疾病の進行/後退をモニターし、疾病治療における薬剤の活性を開発及びモニターすることができる。特に、患者にとって最もふさわしく、有効的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開発することができる。例えば、患者の薬理ゲノムプロファイルに基づき、患者に対して高度に効果的で副作用を殆ど示さない治療薬を選択することができる。

30

【0262】

別の実施態様では、KPP、KPPの断片または、KPPに特異的な抗体類を、或るマイクロアレイ上のエレメントとして用い得る。マイクロアレイを用いて、上記のようなタンパク質-タンパク質相互作用、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ーまたは測定することが可能である。

40

【0263】

或る実施例は、或る組織または細胞タイプの転写イメージを生成するような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タイプ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、所与の条件下で所与の時間に発現した遺伝子の数および相対存在量を定量することにより分析し得る(Seilhame r 他米国特許第5,840,484号「Comparative Gene Transcript Analysis」を参照。当該文献は特に引用することを以って本明細書の一部となす)。従って、特定の組織または細胞タイプの転写または逆転写全体に本発明のポリヌクレオチドまたはその補体をハイブリダイズすることにより、転写イメージを生成し得る。或る実施例では、本発明のポリヌクレオチドまたはその補体がマイクロアレイ上のエレメントのサブセットを複数含むような

50

高処理フォーマットでハイブリダイゼーションを発生させる。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ィールを提供し得る。

【0264】

転写イメージは、組織、株化細胞、生検またはその生物学的サンプルから単離した転写物を用いて生成し得る。転写イメージは従って、組織または生検サンプルの場合には *in vivo*、または株化細胞の場合には *in vitro*での遺伝子発現を反映する。

【0265】

本発明のポリヌクレオチドの発現のプロフィールを作製する転写イメージはまた、工業的または天然の環境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し得る。全ての化合物は、作用及び毒性のメカニズムを暗示する、しばしば分子フィンガープリントまたは毒性シグネチャと称されるような特徴的な遺伝子発現パターンを惹起する (Nuwaysir, E.F. 他 (1999) Mol. Carcinog. 24:153-159、Steiner, S. 及び N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471、該文献は特に引用することを以って本明細書の一部となす)。試験化合物が、既知の毒性を有する化合物のシグネチャと類似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ある。フィンガープリントまたはシグネチャは、多数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る場合には、最も有用且つ正確である。理想的には、発現のゲノム全域にわたる測定が最高品質のシグネチャを提供する。たとえ、発現が任意の試験された化合物によって変化しない遺伝子があったとしても、それらの発現レベルを残りの発現データを標準化するために使用できるため、それらの遺伝子は重要である。標準化手順は、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。毒性シグネチャの要素に遺伝子機能を割り当てることで毒性メカニズムの解釈に役立つが、毒性の予測につながるシグネチャの統計的に一致させるのに遺伝子機能の知識は必要とされない (例えば2000年2月29日に米国国立環境健康科学研究所 (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中毒学的スクリーニングの際に、全ての発現した遺伝子配列を含めることは重要且つ望ましいことである。

【0266】

或る実施例では、核酸を含有する生物学的サンプルを試験化合物で処理することにより試験化合物の毒性を算定する。処理した生物学的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つ若しくは複数のプローブでハイブリダイズし、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し得る。処理した生物学的サンプル中の転写レベルを、未処理生物学的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は、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す。

【0267】

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タイプのプロテオームを分析することに関連する。プロテオームの語は、特定の組織または細胞タイプでのタンパク質発現の包括的パターンを指す。プロテオームの各タンパク質成分は、個々に更に分析の対象とすることができる。プロテオーム発現パターン即ちプロフィールは、所与の条件下で所与の時間に発現したタンパク質の数及び相対存在量を定量することにより分析し得る。従って細胞のプロテオームのプロフィールは、特定の組織または細胞タイプ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、1次元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、2次元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応じて分離するような2次元ゲル電気泳動により分離が達成される (前出のSteinerおよびAnderson)。タンパク質は、通常はクーマシーブルー、あるいは銀染色液または蛍光染色液などの物質を用いてゲルを染色することにより、分散した、独自の位置にある点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または未処理のいずれかの生物学的サンプルから得られる同等に位置する

10

20

30

40

50

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同定する。スポット内のタンパク質は、例えば化学的または酵素的切断とそれに続く質量分析を用い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、その部分配列を、好適には少なくとも5個の連続するアミノ酸残基を、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データが得られる。

【0268】

プロテオームのプロファイルは、KPPに特異的な抗体を用いてKPP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、マイクロアレイ上でエレメントとして抗体を用い、マイクロアレイをサンプルに曝して各アレイ要素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. 他(1999) Anal. Biochem. 270:103-111、Mendoze, L.G. 他(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、チオールまたはアミノ反応性蛍光化合物を用いてサンプル中のタンパク質を反応させ、各アレイのエレメントにおける蛍光結合の量を検出し得る。

10

【0269】

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に分析するべきである。或る組織の或るタンパク質に対しては、転写とタンパク質の存在量の相関が乏しいこともあるので(Anderson, N.L.及びJ. Seilhamer (1997) Electrophoresis 18:533-537)、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ィールを変化させるような化合物の分析においてプロテオーム毒性シグネチャは有用たり得る。更に、体液中の転写物の分析はmRNAの急速な分解のために困難なので、プロテオームのプロフィール作成はこのような場合により信頼し得、情報価値があり得る。

20

【0270】

別の実施例では、タンパク質を含有する生物学的サンプルを試験化合物で処理することにより試験化合物の毒性を算定する。処理された生物学的サンプル中で発現したタンパク質は、各タンパク質の量を定量し得るように分離する。各タンパク質の量を、非処理生物学的サンプル中の対応するタンパク質の量と比較する。両サンプルのタンパク質量の差は、処理サンプル中の試験化合物に対する毒性反応を示す。個々のタンパク質は、個々のタンパク質のアミノ酸残基をシーケンシングし、これら部分配列を本発明のポリペプチドと比較することにより同定する。

30

【0271】

別の実施例では、タンパク質を含有する生物学的サンプルを試験化合物で処理することにより試験化合物の毒性を算定する。生物学的サンプルから得たタンパク質は、本発明のポリペプチドに特異的な抗体を用いてインキュベートする。抗体により認識されたタンパク質の量を定量する。処理された生物学的サンプル中のタンパク質の量を、非処理生物学的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は、処理サンプル中の試験化合物に対する毒性反応を示す。

【0272】

マイクロアレイは、本技術分野でよく知られている方法を用いて調製し、使用し、そして分析する(Brennan, T.M. 他(1995)の米国特許第5,474,796号、Schna, M. 他(1996) Proc. Natl. Acad. Sci. USA 93:10614-10619、Baldeschweiler 他(1995) PCT出願第W095/251116号、Shalon, D. 他(1995) PCT出願第W095/35505号、Heller, R.A. 他(1997) Proc. Natl. Acad. Sci. USA 94:2150-2155、Heller, M.J. 他(1997) 米国特許第5,605,662号等を参照)。様々なタイプのマイクロアレイが公知であり、詳細については、DNA Microarrays: A Practical Approach, M. Schna, 編集。(1999) Oxford University Press, Londonに記載されている。該文献は、特に引用することを以って本明細書の一部となす。

40

【0273】

50

本発明の別の実施態様ではまた、KPP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。コード配列または非コード配列のいずれかを用いることができ、或る例では、コード配列全体よりも非コード配列が好ましい。例えば、多重遺伝子族のメンバー内でのコード配列の保存により、染色体マッピング中に望ましくないクロスハイブリダイゼーションが生じる可能性がある。配列は、或る特定の染色体に、または或る染色体の或る特定領域に、または人為形成の染色体、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、あるいは単一染色体cDNAライブラリ群に対してマッピングされる (Harrington, J.J他 (1997) Nat. Genet. 15:345-355、Price, C.M. (1993) Blood Rev. 7:127134、Trask, B.J. (1991) Trends Genet. 7:149154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を特定の染色体領域の遺伝または制限酵素断片長多型 (RFLP) と関連させるような遺伝子連鎖地図を発生させ得る。(例えば、Lander, E.S.およびD. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357を参照)。

10

20

30

40

50

【0274】

蛍光原位ハイブリッド形成法 (FISH) は、他の物理的及び遺伝地図データと関連し得る (前出Meyers, 965-968ページのHeinz-Ulrich, 他 (1995)を参照)。遺伝地図データの例は、種々の科学雑誌あるいはOnline Mendelian Inheritance in Man (OMIM) のウェブサイトに見ることができる。物理的な染色体地図上のKPPをコードする遺伝子の位置と、特定の疾患との関連性、あるいは特定の疾患に対する素因との関連性は、この疾患と関連するDNAの領域の決定に役立ち得るため、位置を決定するクローニングの作業を促進し得る。

【0275】

確定した染色体マーカーを用いた結合分析等の物理的マッピング技術及び染色体標本原位ハイブリッド形成法を用いて、遺伝地図を拡張することができる。例えばマウスなど別の哺乳動物の染色体上に遺伝子を配置することにより、正確な染色体上の遺伝子座が未知でも、関連するマーカー類をしばしば明らかにし得る。この情報は、位置クローニングその他の遺伝子発見技術を用いて疾患遺伝子を探す研究者にとって価値がある。いったん疾患または症候群に關与する遺伝子 (群) が、血管拡張性失調症の11q22-23領域など、特定のゲノム領域への遺伝的連鎖によって、大まかに位置決めがなされると、該領域にマップされる任意の配列は、更なる調査のための関連遺伝子あるいは調節遺伝子を提示している可能性がある (例えばGatti, R.A. 他 (1988) Nature 336:577580を参照)。転座、反転などに起因する、健常者、保有者、罹病者の三者間における染色体位置の相違を検出する場合にも、本発明のヌクレオチド配列を用い得る。

【0276】

本発明の別の実施例では、KPP、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。薬物スクリーニングに用いる断片は、溶液中に遊離しているか、固体支持物に固定されるか、細胞表面上に保持されるか、細胞内に位置することができる。KPPと、試験する薬物との結合による複合体の形成を測定し得る。

【0277】

別の薬物スクリーニング方法は、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を高い処理能力でスクリーニングするために用いられる (Geysen, 他 (1984) PCT application W084/03564等を参照)。この方法においては、多数の異なる小さな試験用化合物を固体基板上で合成する。試験用化合物は、KPP、或いはその断片と反応してから洗浄される。次に、本技術分野でよく知られている方法で、結合したKPPを検出する。精製したKPPはまた、上記した薬剤のスクリーニング技術において用いるプレート上で直接コーティングすることもできる。別法では、非中和抗体を用いてペプチドを捕捉し、ペプチドを固体支持物に固定することもできる。

【0278】

別の実施例では、KPPと結合可能な中和抗体がKPP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、KPP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。

【0279】

別の実施例では、発展途上の分子生物学技術にKPP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。

【0280】

更に詳細説明をしなくとも、当業者であれば以上の説明を以って本発明を最大限に利用できるであろう。したがって、これ以下に記載する実施例は単なる例示目的にすぎず、いかようにも本発明を限定するものではない。

【0281】

本明細書において開示した全ての特許、特許出願及び刊行物、特に米国特許第60/263,083号、第60/271,205号、第60/271,117号、第60/276,859号、第60/278,504号、第60/278,522及び第60/280,510、第60/280,266は、言及することをもって本明細書の一部となす。

【実施例】

【0282】

1 cDNAライブラリの作製

Incyte cDNA群の由来は、LIFSEQ GOLDデータベース (Incyte Genomics, Palo Alto CA) に記載されたcDNAライブラリ群である。幾つかの組織はホモジナイズしてグアニジニウムイソチオシアネート溶液に溶解し、他の組織はホモジナイズしてフェノールにまたは変性剤群の好適な混合液に溶解した。混合液の1例であるTRIZOL (Life Technologies) は、フェノールとグアニジンイソチオシアネートとの単相溶液である。結果として得られた溶解物は、塩化セシウムクッション上で遠心分離するかクロロホルムで抽出した。イソプロパノールか、酢酸ナトリウムとエタノールか、いずれか一方、或いは別の方法を用いて、溶解物からRNAを沈殿させた。

【0283】

RNAの純度を高めるため、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、DNアーゼでRNAを処理した。殆どのライブラリでは、オリゴd(T)連結常磁性粒子 (Promega)、OLIGOTEXラテックス粒子 (QIAGEN, Chatsworth CA) または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、別のRNA単離キット、例えばPOLY(A)PURE mRNA精製キット (Ambion, Austin TX) を用いて組織溶解物からRNAを直接単離した。

【0284】

場合によってはStratagene社へのRNA提供を行い、対応するcDNAライブラリをStratagene社が作製することもあった。そうでない場合は、UNIZAPベクターシステム (Stratagene) またはSUPERSCRIPトラスミドシステム (Life Technologies) を用いて本技術分野で既知の推奨方法または類似の方法でcDNAを合成し、cDNAライブラリを作製した (前出のAusubel, 1997, 5.1-6.6ユニットなどを参照)。逆転写は、オリゴd(T)またはランダムプライマーを用いて開始した。合成オリゴヌクレオチドアダプターを二本鎖cDNAに連結反応させ、好適な制限酵素または酵素でcDNAを消化した。殆どのライブラリに対して、cDNAのサイズ (300 ~ 1000 bp) 選択は、SEPHACRYL S1000、SEPHAROSE CL2BまたはSEPHAROSE CL4Bカラムクロマトグラフィー (Amersham Pharmacia Biotech)、或いは調製用アガロースゲル電気泳動法を用いて行った。cDNAは好適なトラスミドのポリリンカーの適合する制限酵素部位で連結された。好適なトラスミドは、例えばPBLUESCRIPTトラスミド (Stratagene)、PSPORT1 トラスミド (Life Technologies)、PCDNA2.1 トラスミド (Invitrogen, Carlsbad CA)、PBK-CMV トラスミド (Stratagene)、PCR2-TOPOTAトラスミド (Invitrogen)、PCM V-ICISトラスミド (Stratagene)、pIGEN (Incyte Genomics, Palo Alto CA)、pRARE (Incyte Genomics) またはpINCY (Incyte Genomics) 等およびその誘導体である。組換えトラスミドは、Stratagene社のXL1-Blue、XL1-BlueMRFまたはSOLR、或いはLife Technologies社のDH5、DH10BまたはELECTROMAX DH10Bを含むコンピテント大腸菌細胞に形質転換した。

10

20

30

40

50

【0285】

2 cDNAクローンの単離

UNIZAPベクターシステム (Stratagene) を用いた *in vivo* 切除によって、或いは細胞溶解によって、実施例1のようにして得たプラスミドを宿主細胞から回収した。プラスミドの精製には、下記の少なくとも1つを用いた。すなわち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 Plasmid、QIAWELL 8 Plus PlasmidおよびQIAWELL 8 Ultra Plasmid 精製システム、R.E.A.L. Prep 96プラスミド精製キットのいずれかである。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。

10

【0286】

別法では、高処理フォーマットにおいて直接結合PCR法を用い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)。宿主細胞の溶解および熱サイクリング過程は、単一反応混合液中で行った。サンプルを処理し、それを384ウェルプレート内で保管し、増幅したプラスミドDNAの濃度をPICOGREEN色素 (Molecular Probes, Eugene OR) 及びFluoroskan II蛍光スキャナ (Labsystems Oy, Helsinki, Finland) を用いて蛍光分析的に定量した。

【0287】

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Incyte cDNAを、以下に示すようにシークエンシングした。cDNAのシークエンス反応は、標準的方法或いは高処理装置、例えばABI CATALYST 800 サーマルサイクラー (Applied Biosystems) またはPTC-200 サーマルサイクラー (MJ Research) をHYDRAマイクロディスペンサー (Robbins Scientific) またはMICROLAB 2200 (Hamilton) 液体転移システムと併用して処理した。cDNAのシークエンス反応は、Amersham Pharmacia Biotech社が提供する試薬、またはABIシークエンシングキット、例えばABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems) の試薬を用いて準備した。cDNAのシークエンス反応の電気泳動的分離及び標識したポリヌクレオチドの検出には、MEGABACE 1000 DNAシークエンシングシステム (Molecular Dynamics) が、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 (Applied Biosystems) が、或いはその他の本技術分野でよく知られている配列解析システムを用いた。cDNA配列内のリーディングフレームは、標準的方法 (前出のAusubel, 1997, unit 7.7に概説) 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させた。

20

30

【0288】

IncyteのcDNA配列に由来するポリヌクレオチド配列は、ベクター、リンカー及びポリ(A)配列の除去し、あいまいな塩基対をマスクすることによって有効性を確認した。その際、BLAST、動的プログラミング及び隣接ジヌクレオチド頻度分析に基づくアルゴリズム及びプログラムを用いた。次に、Incyte cDNA配列またはそれらの翻訳の問い合わせを、以下のデータベース群に対して行った。すなわち、選抜した公共のデータベース群 (例えばGenBankの霊長類、げっ歯類、哺乳類、脊椎動物、真核生物のデータベースと、BLOCKS、PRINTS、DOMO、PRODOM) と、ヒト、ラット、マウス、線虫 (*Caenorhabditis elegans*)、出芽酵母 (*Saccharomyces cerevisiae*)、分裂酵母 (*Schizosaccharomyces pombe*) および*Candida albicans*からの配列群を持つPROTEOMEデータベース群 (Incyte Genomics, Palo Alto CA)、および、隠れマルコフモデル (HMM) ベースのタンパク質ファミリーデータベース群 (PFAM等) である (HMMは、遺伝子ファミリーのコンセンサス1次構造を分析する確率的アプローチである。例えば、Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6: 361-365を参照のこと)。問合せは、BLAST、FASTA、BLIMPS及びHMMERに基づくプログラムを用いて行った。Incyte cDNA配列は、完全長のポリヌクレオチド配列を産出するように構築された。あるいは、GenBank cDNA群、GenBank EST群、ステッチされた配列群、ストレッチされた配列群、またはGenscan予測コード配列群 (実施例4 および 5 を参照) を

40

50

用い、Incyte cDNAの集団を完全長まで伸長させた。PhredとPhrapとConsedとに基づくプログラムを用いて構築し、GenMarkとBLASTとFASTAとに基づくプログラムを用いて、cDNAの集団を、オープンリーディングフレームについてスクリーニングした。完全長ポリヌクレオチド配列を翻訳し、対応する完全長ポリペプチド配列を誘導した。あるいは、本発明のポリペプチドは、完全長翻訳ポリペプチドの任意のメチオニン残基で開始し得る。引き続き、GenBankタンパク質データベース (genpept)、SwissProt、P R O T E O M E データベース、BLOCKS、PRINTS、DOMO、PRODOM及びProsite等のデータベース、PFAM等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する問合せによって完全長ポリペプチド配列を分析した。完全長ポリヌクレオチド配列はまた、MA CDNASIS PROソフトウェア (日立ソフトウェアエンジニアリング, South San Francisco CA) およびLASERGENEソフトウェア (DNASTAR) を用いて分析した。ポリヌクレオチド及びポリペプチド配列アラインメントは、アラインメントした配列と配列の一致率も計算するMEGALIGNマルチシーケンスアラインメントプログラム (DNASTAR) に組み込まれているようなCLUSTALアルゴリズムによって特定されるデフォルトパラメータを用いて生成する。

【0289】

Incyte cDNA及び完全長配列の分析及びアセンブリに利用したツール、プログラム及びアルゴリズムの概略と、適用可能な説明、引用文献、閾値パラメータを表7に示す。用いたツール、プログラム及びアルゴリズムを表7の列1に、それらの簡単な説明を列2に示す。列3は好適な引用文献であり、全ての文献はそっくりそのまま引用を以って本明細書の一部となす。適用可能な場合には、列4は2つの配列が一致する強さを評価するために用いたスコア、確率値その他のパラメータを示す (スコアが高いほど、また確率値が低いほど、2配列間の同一性が高くなる)。

【0290】

完全長ポリヌクレオチド配列群とポリペプチド配列群との構築と分析とに用いた上記プログラム群は、SEQ ID NO:9-16のポリヌクレオチド配列断片群の同定にも利用した。ハイブリダイゼーション技術と増幅技術とに有用な約20~約4000ヌクレオチドの断片群を、表4の列2に示した。

【0291】

4 ゲノムDNAからのコード配列の同定及び編集

推定上のキナーゼおよびホスファターゼは、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、様々な生物からゲノムDNA配列を分析する汎用遺伝子同定プログラムである (Burge, C.及びS. Karlin (1997) J. Mol. Biol. 268:78-94、Burge, C.及びS. Karlin (1998) Cuff. Opin. Struct. Biol. 8:346-354参照)。プログラムは予測エキソンを連結し、メチオニンから停止コドンに及ぶ構築されたcDNA配列を形成する。Genscanの出力は、ポリヌクレオチドおよびポリペプチド配列のFASTAデータベースである。Genscanが一度に分析する配列の最大範囲は、30 kbに設定した。これらのGenscan推定cDNA配列の内、どの配列がキナーゼおよびホスファターゼをコードするかを決定するために、コードされたポリペプチドをPFAMモデルにおいてキナーゼおよびホスファターゼについて問合せて分析した。潜在的なキナーゼおよびホスファターゼも、キナーゼおよびホスファターゼとして注釈が付けられたIncyte cDNA配列に対する相同性を基に同定された。こうして選択されたGenscan予測配列は、次にBLAST分析により公共データベースgenpept及びgbpriと比較した。必要であれば、genpeptからのトップBLASTヒットと比較することによりGenscan予測配列を編集し、余分なまたは省略されたエキソンなど、Genscanが予測した配列におけるエラーを補正した。BLAST分析はまた、Genscan予測配列の、いかなるIncyte cDNAまたは公共cDNAカバレッジ (coverage) の発見にも用いられ、したがって転写の証拠を提供した。Incyte cDNAカバレッジが利用できた場合には、この情報を用いてGenscan予測配列を補正または確認した。完全長ポリヌクレオチド配列は、実施例3に記載した構築プロセスを用いて、Incyte cDNA配列および/または公共cDNA配列でGenscan予測コード配列を構築して得

た。或いは、完全長ポリヌクレオチド配列は編集または非編集のGenscan予測コード配列に完全に由来する。

【0292】

5 cDNA配列データを使ったゲノム配列データの構築

ステッチ配列 (Stitched Sequence)

部分cDNA配列は、実施例 4に記載のGenscan遺伝子同定プログラムにより予測されたエキソンを用いて伸長させた。実施例 3に記載されたように構築された部分cDNAは、ゲノムDNAにマッピングし、関連するcDNA及び1つ若しくは複数のゲノム配列から予測されたGenscanエキソンを含むクラスターに分解した。cDNA及びゲノム情報を統合するべくグラフ理論及び動的プログラミングに基づくアルゴリズムを用いて各クラスターを分析し、引き続いて確認、編集または伸長して完全長配列を産出するような潜在的スプライス変異体を生成した。間隔全体の長さがクラスター中の2以上の配列に存在するような配列を同定し、そのように同定された間隔は推移により等しいと考えられた。例えば、1つのcDNA及び2つのゲノム配列に間隔が存在する場合、3つの間隔は全て等しいと考えられる。このプロセスは、無関係であるが連続したゲノム配列をcDNA配列により結び合わせて架橋し得る。このようにして同定した区間を、それらの親配列 (parent sequences) に沿って現われる順にスティッチアルゴリズムで「縫い合わせ」、可能な限り最長の配列と変異配列群とを生成した。1種類の親配列に沿って発生した間隔と間隔との連鎖 (cDNA - cDNAまたはゲノム配列 - ゲノム配列) は、親の種類を変える連鎖 (cDNA - ゲノム配列) に優先した。結果として得たスティッチ配列群を翻訳し、BLAST分析で公共データベースgenpeptおよびgbpriと比較した。Genscanにより予測された不正確なエキソンは、genpeptからトップのBLASTヒットと比較することにより修正した。必要な場合には、追加cDNA配列を用いるかゲノムDNAの検査により配列を更に伸長させた。

10

20

【0293】

ストレッチ配列 (Stretched Sequence)

部分DNA配列は、BLAST分析に基づくアルゴリズムにより完全長まで伸長された。まず、BLASTプログラムを用いて、GenBankの霊長類、げっ歯類、哺乳動物、脊椎動物及び真核生物のデータベースなどの公共データベースに対し、実施例 3に記載されたように構築された部分cDNAを問い合わせた。次に、最も近いGenBankタンパク質相同体を、BLAST分析により、Incyte cDNA配列または実施例 4に記載のGenScanエキソン予測配列のいずれかと比較した。結果として得られる高スコアリングセグメント対 (HSP) を用いてキメラタンパク質を産出し、翻訳した配列をGenBankタンパク質相同体上にマッピングした。元のGenBankタンパク質相同体に対し、キメラタンパク質内では挿入または欠失が起こり得る。GenBankタンパク質相同体、キメラタンパク質またはその両方をプローブとして用い、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した。このようにして、部分的なDNA配列を、相同ゲノム配列の付加によりストレッチすなわち伸長した。結果として得られるストレッチ配列を検査し、完全遺伝子を含んでいるか否かを判定した。

30

【0294】

6 KPP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:9-16を構築するために用いた配列群を、BLAST他のSmith-Watermanアルゴリズムの実装群を用いて、Incyte LIFESEQデータベースおよび公共ドメインデータベース群の配列と比較した。SEQ ID NO:9-16と一致するこれらのデータベースの配列を、Phrapなどの構築アルゴリズム (表7) を使用して、連続しオーバーラップする配列のクラスター群に組み入れた。スタンフォード・ヒトゲノムセンター (SHGC)、ホワイトヘッド・ゲノム研究所 (WIGR)、Genethonなどの公的な情報源から入手可能な放射線ハイブリッドおよび遺伝地図データを用いて、いずれかのクラスター化された配列が既にマッピングされているかを判定した。マッピングされた配列が或るクラスターに含まれている場合、そのクラスターの全配列が、個々の配列番号と共に、地図上の位置に割り当てられた。

40

【0295】

地図上の位置は、ヒト染色体の範囲または間隔として表される。センチモルガン間隔の地

50

図上の位置は、染色体のpアームの末端に関連して測定する。(センチモルガン(cM)は、染色体マーカー間の組換え頻度に基づく計測単位である。平均して、1 cMは、ヒト中のDNAの1メガベース(Mb)にほぼ等しい。尤も、この値は、組換えのホットスポット及びコールドスポットに起因して広範囲に変化する。) cM距離は、各クラスター内に配列が含まれる放射線ハイブリッドマーカー類に対して境界を提供するGenethonによってマッピングされた遺伝マーカー群に基づく。NCBI「GeneMap99」(<http://www.ncbi.nlm.nih.gov/genemap>)などの一般個人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。

【0296】

10

7 ポリヌクレオチド発現の分析

ノーザン分析は、転写された遺伝情報の存在を検出するために用いられる実験技術であり、特定の細胞種または組織からのRNAが結合され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に関与している。(例えば前出のSambrook, 7章、同Ausubel (1995) 4章および16章を参照)。

【0297】

BLASTに適用する類似のコンピュータ技術を用いて、GenBankやLIFESEQ (Incyte Genomics) 等のcDNAデータベースにおいて同一または関連分子を検索する。ノーザン分析は、多数膜系ハイブリダイゼーションよりも非常に速い。更に、任意の特定の一致を厳密なあるいは相動的なものとして分類するか否かを決定するため、コンピュータ検索の感度を修正することができる。検索の基準は積スコアであり、次式で定義される。

20

【0298】

【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

積スコアは、2つの配列間の類似度と、配列が一致する長さとの両方を考慮している。積スコアは、0 ~ 100の正規化された値であり、次のようにして求める。BLASTスコアにヌクレオチドの配列一致率を乗じ、その積を2つの配列の短い方の長さの5倍で除する。BLASTスコアを計算するには、或る高スコアリングセグメント対(HSP)内の一致する各塩基に+5のスコアを割り当て、各不一致塩基に-4を割り当てる。2つの配列は、2以上のHSPを共有し得る(ギャップにより隔離される)。2以上のHSPがある場合には、最高BLASTスコアのセグメント対を用いて積スコアを計算する。積スコアは、断片的オーバーラップとBLASTアラインメントの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%一致する場合のみ得られる。積スコア70は、一端が100%一致し、70%オーバーラップしているか、他端が88%一致し、100%オーバーラップしているかのいずれかの場合に得られる。積スコア50は、一端が100%一致し、50%オーバーラップしているか、79%一致し、100%オーバーラップしているかのいずれかの場合に得られる。

30

40

【0299】

或いは、KPPをコードするポリヌクレオチド配列群を、由来する組織源について分析する。例えば幾つかの完全長配列は、少なくとも一部は、オーバーラップするIncyte cDNA配列群を用いて構築される(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。すなわち心血管系、結合組織、消化器系、胎芽構造、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、肝、筋骨格系、神経系、膵臓、呼吸器系、感覚器、皮膚、顎口腔系、非分類性/混合性または尿路である。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。同様に、各ヒト組織は、以下の疾患/条件カテゴリーすなわち癌、細胞株、発達、炎症、神経性、外傷、心

50

血管、プール、その他の1つに分類される。各カテゴリーのライブラリ数を数えて、全カテゴリーの総ライブラリ数で除する。得られるパーセンテージは、KPPをコードするcDNAの組織特異的および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。

【0300】

8 ポリヌクレオチドをコードするKPPの伸長

完全長のポリヌクレオチド配列はまた、完全長分子の適切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該断片を伸長させて生成した。或るプライマーは既知の断片の5'伸長を開始するべく合成し、別のプライマーは既知の断片の3'伸長を開始するべく合成した。開始プライマー群の設計にはOLIGO 4.06ソフトウェア (National Biosciences) 或いは別の適切なプログラムを用い、長さが約22~30ヌクレオチド、GC含有率が約50%以上となり、約68~約72の温度で標的配列にアニーリングするようにした。ヘアピン構造およびプライマー-プライマー二量体を生ずるようなヌクレオチド群の伸長は、全て回避した。

10

【0301】

配列を伸長するために、選択されたヒトcDNAライブラリを用いた。2段階以上の伸長が必要または望ましい場合には、付加的プライマーあるいはプライマーのネステッドセットを設計した。

【0302】

高忠実度の増幅が、当業者によく知られている方法を利用したPCR法によって得られた。PCRは、PTC-200 サーマルサイクラー (MJ Research, Inc.) を用いて96ウェルプレート内で行った。反応混合液は、鋳型DNA及び200nmol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反応バッファー、Taq DNAポリメラーゼ (Amersham Pharmacia Biotech)、ELONGASE酵素 (Life Technologies)、Pfu DNAポリメラーゼ (Stratagene) を含む。プライマーの組、PCI AとPCI Bに対して以下のパラメータで増幅を行った。ステップ1: 94 で3分間、ステップ2: 94 で15秒間、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および4を20回繰り返す、ステップ6: 68 で5分間、ステップ7: 4 で保存。別法では、プライマー対であるT7とSK+に対し、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秒間、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および4を20回繰り返す、ステップ6: 68 で5分間、ステップ7: 4 で保存。

20

30

【0303】

各ウェルのDNA濃度は、 $1 \times TE$ および $0.5 \mu l$ の希釈していないPCR産物に溶解した $100 \mu l$ のPICOGREEN定量試薬(0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR)を不透明な蛍光光度計プレート (Corning Costar, Acton MA)の各ウェルに分配し、DNAが試薬と結合できるようにして測定した。サンプルの蛍光を計測してDNAの濃度を定量すべく、プレートをFluoroskan II (Labsystems Oy, Helsinki, Finland) でスキャンした。反応混合物のアリコット $5 \sim 10 \mu l$ を1%アガロースゲル上で電気泳動法によって解析し、どの反応が配列の伸長に成功したかを判定した。

40

【0304】

伸長させたヌクレオチドは、脱塩及び濃縮して384穴プレートに移し、CviJIコレラウイルスエンドヌクラーゼ (Molecular Biology Research, Madison WI) を用いて消化し、pUC 18ベクター (Amersham Pharmacia Biotech) への再連結反応前に音波処理またはせん断した。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で分離し、断片を切除し、寒天をAgar ACE (Promega) で消化した。伸長させたクローンをT4リガーゼ (New England Biolabs, Beverly MA) を用いてpUC 18ベクター (Amersham Pharmacia Biotech) に再連結し、Pfu DNAポリメラーゼ (Stratagene) で処理して制限部位のオーバーハングを満たし、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有する培地に移し、それぞれのコロニ

50

ーを切りとってLB/2xカルベニシリン培養液の384穴プレートに37 で一晩培養した。

【0305】

細胞を溶解して、Taq DNAポリメラーゼ(Amersham Pharmacia Biotech)及びPfu DNAポリメラーゼ(Stratagene)を用いて以下の手順でDNAをPCR増幅した。ステップ1: 94 で3分間、ステップ2: 94 で15秒間、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および4を29回繰り返す、ステップ6: 72 で5分間、ステップ7: 4 で保存。上記したようにPICOGREEN試薬(Molecular Probes)でDNAを定量化した。DNAの回収率が低いサンプルは、上記と同一の条件を用いて再増幅した。サンプルは20%ジメチルスルホキシド(1:2, v/v)で希釈し、DYENAMIC エネルギートランスファー シークエンシングプライマー、及びDYENAMIC DIRECT kit(Amersham Pharmacia Biotech)またはABI PRISM BIGDYE ターミネーターサイクル シークエンシング反応キット(Terminator cycle sequencing ready reaction kit)(Applied Biosystems)を用いてシークエンシングした。

10

【0306】

同様に、上記手順を用いて完全長ポリヌクレオチド配列を検証し、或いはそのような伸長のために設計されたオリゴヌクレオチド及び適切なゲノムライブラリを用いて5'調節配列を得る。

【0307】

9 KPPをコードするポリヌクレオチドの一塩基多型の同定

一塩基多型性(SNP)として知られるありふれたDNA配列変異体は、LIFESEQデータベース(Incyte Genomics)を用いてSEQ ID NO:9-16で同定される。実施例3に記述されているように同じ遺伝子からの配列は共にクラスター化され、構築され、遺伝子内の全ての配列変異体を同定する。一連のフィルタから成るアルゴリズムは、SNPを他の配列変異体から区別するために用いられる。予備フィルタが、最小限のPhredクオリティスコア15を要求することにより大多数のベースコールのエラーを除去し、ベクター配列、キメラ、スプライス変異体の不適切なトリミングに起因する配列アラインメントエラーとエラーを除去した。進歩した染色体分析の自動化手順により、推定上のSNPの近傍の本来のクロマトグラムファイル进行分析した。クローンエラー-フィルタは、統計的に生み出されたアルゴリズムを用いて、逆転写酵素、ポリメラーゼあるいは体細胞性突然変異によって引き起こされる等の実験プロセッシング中に導入されたエラー-を同定した。クラスターリングエラー-フィルタは、統計的に生み出されたアルゴリズムを用いて、近い相同体あるいは偽遺伝子のクラスターリングに起因するエラー、または非ヒト配列によるコンタミネーションによるエラー-を同定した。フィルタの最終セットは、免疫グロブリンまたはT細胞受容体に見出される重複とSNPを除去した。

20

30

【0308】

異なる4つのヒト集団のSNP部位における対立遺伝子頻度を分析するために高処理MASSARRAYシステム(Sequenom, Inc.)を用いる質量分析によって更なる特徴付けのためにいくつかのSNPが選択された。白人集団(男性46人、女性46人)は92人から成り、そのうち83人はユタ州、4人はフランス、3人はベネズエラ、2人はアーミッシュの出身である。黒人集団は194人(男性97人、女性97人)から成り、全て米国の黒人である。ラテンアメリカ系集団は324人(男性162人、女性162人)から成り、全てメキシコ出身である。アジア系の集団は126人(男性64人、女性62人)からなり、中国人43%、日本人31%、韓国人13%、ベトナム人5%、他のアジア系8%の親の構成が報告されている。対立遺伝子頻度は最初に白人集団で分析された。この集団で対立遺伝子変異を示さないSNPの時には他の3つの集団で更に試験されない場合もあった。

40

【0309】

10 個々のハイブリダイゼーションプローブの標識及び使用

SEQ ID NO:9-16から得たハイブリダイゼーションプローブを利用して、cDNA、ゲノムDNAまたはmRNA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載するが、より大きなヌクレオチド断片に対しても本質的に同一の手順が用い

50

られる。オリゴヌクレオチドを、OLIG04.06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、50pmolの各オリゴマーと、250 μ Ciの[32 P] アデノシン三リン酸 (Amersham Pharmacia Biotech) 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。標識したオリゴヌクレオチドは、SEPHADEX G-25超細繊維分子サイズ排除デキストラン ビードカラム (Amersham Pharmacia Biotech) を用いて実質的に精製する。Ase I、Bgl II、Eco RI、Pst I、XbaI または Pvu II (DuPont NEN) のいずれか1つのエンドヌクレアーゼで消化されたヒトゲノム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、毎分 10^7 カウントの標識されたプローブを含むアリコットを用いる。

【0310】

各消化物から得たDNAは、0.7%アガロースゲル上で分画してナイロン膜 (Nytran Plus, Schleicher & Schuell, Durham NH) に移す。ハイブリダイゼーションは、40で16時間行う。非特異的シグナルを除去するため、例えば0.1 \times クエン酸ナトリウム食塩水および0.5%ドデシル硫酸ナトリウムに一致する条件下で、プロットを室温で順次洗浄する。オートラジオグラフィまたはそれに代わるイメージング手段を用いてハイブリダイゼーションパターンを視覚化し、比較する。

【0311】

11 マイクロアレイ

マイクロアレイの表面上でアレイエレメントの結合または合成は、フォトリソグラフィ、圧電印刷 (インクジェット印刷、前出のBalteschweiler等を参照)、機械的マイクロスポッティング技術及びこれらから派生したものをを用いて達成することが可能である。上記各技術において基板は、均一な、非多孔性の表面を持つ固体とすべきである (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。或いは、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、熱的、紫外線的、化学的または機械的結合手順を用いて基板の表面にエレメントを配置及び結合させてもよい。通常のアレイは、利用可能な、当業者に公知の方法と機械とを用いて作製でき、任意の適正数のエレメントを有し得る (例えば Schena, M. 他. (1995) Science 270:467-470、Shalon, D. 他. (1996) Genome Res. 6:639-645、Marshall, A. 及び J. Hodgson (1998) Nat. Biotechnol. 16:27-31等を参照)。

【0312】

完全長cDNA、発現配列タグ (EST)、またはその断片またはオリゴマーは、マイクロアレイのエレメントと成り得る。ハイブリダイゼーションに好適な断片またはオリゴマーを、LASERGENEソフトウェア (DNASTAR) 等の本技術分野で公知のソフトウェアを用いて選択することが可能である。アレイエレメント群を、生体サンプル中のポリヌクレオチド群とハイブリダイズする。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などの分子タグに抱合させる。ハイブリダイゼーション後、生体サンプルからのハイブリダイズされていない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でのハイブリダイゼーションを検出する。あるいは、レーザ脱離および質量スペクトロメトリ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或る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、相補性の度合と相対存在度とを算出し得る。一実施態様におけるマイクロアレイの調整および使用について、以下に詳述する。

【0313】

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サンプルを、MMLV逆転写酵素、0.05pg/ μ lのオリゴ(dT)プライマー (21mer)、1 \times 第一鎖合成バッファー、0.03unit/ μ lのRNアーゼ阻害因子、500 μ MのdATP、500 μ MのdGTP、500 μ MのdTTP、40 μ MのdCTP、40 μ MのdCTP-Cy3 (BDS) またはdCTP-Cy5 (Amersham Pharmacia Biotech) を用いて逆転写する。逆転写反応は、GEMBRIGHTキット (Incyte) を用い、200ngのポリ(A)⁺RNAを含

10

20

30

40

50

有する体積 2.5 ml で行う。特異的対照ポリ(A)⁺RNAは、非コード酵母ゲノムDNAから *in vitro* 転写により合成する。37 で2時間インキュベートした後、各反応サンプル(1つはCy3、もう1つはCy5標識)は、2.5 mlの0.5 M水酸化ナトリウムで処理し、85 で20分間インキュベートし、反応を停止させてRNAを分解させる。サンプルは、2つの連続するCHROMA SPIN 30ゲル濾過スピンカラム (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) を用いて精製する。混合後、2つの反応サンプルは、1 mlのグリコーゲン(1 mg/ml)、60 mlの酢酸ナトリウム及び300 mlの100%エタノールを用いてエタノール析出させる。サンプルは次に、SpeedVAC (Savant Instruments Inc., Holbrook NY) を用いて完全に乾燥させ、14 µlの5 × SSC / 0.2% SDS中で再懸濁する。

【0314】

10

マイクロアレイの調製

本発明の配列を用いて、アレイエレメントを生成する。各アレイエレメントは、クローン化cDNAインサートによりベクター含有細菌性細胞から増幅する。PCR増幅は、cDNAインサートの側面に位置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で1~2 ngの初期量から5 µgより大きい最終量までアレイエレメントを増幅する。増幅したアレイエレメントは、SEPHACRYL-400 (Amersham Pharmacia Biotech) を用いて精製する。

【0315】

精製したアレイエレメントは、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、0.1%のSDSおよびアセトン中で超音波洗浄し、処理中および処理後に多量の蒸留水で洗浄する。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、蒸留水中で非常に良く洗浄し、95%エタノール中で0.05%アミノプロピルシラン (Sigma) を用いてコーティングする。コーティングしたスライドは、110 のオーブンで硬化させる。

20

【0316】

米国特許第5,807,522号で説明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。該特許は、引用を以って本明細書の一部となす。平均濃度100 ng/µlのアレイエレメントDNA 1 µlを、高速ロボット装置 (robotic apparatus) 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。装置はここで、スライド毎に約5 nlのアレイエレメントサンプルを加える。

30

【0317】

マイクロアレイには、STRATALINKER UV架橋剤 (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度洗浄し、蒸留水で3度洗浄する。リン酸緩衝生理食塩水 (PBS) (Tropix, Inc., Bedford MA) 中の0.2% カゼイン中において60 で30分間マイクロアレイをインキュベートした後、前に行ったように0.2% SDS及び蒸留水で洗浄することにより、非特異結合部位をブロックする。

【0318】

ハイブリダイゼーション

40

ハイブリダイゼーション反応は、5 × SSC, 0.2% SDSハイブリダイゼーション緩衝液中のCy3及びCy5標識したcDNA合成生成物を各0.2 µg含む9 µlのサンプル混合体を有する。サンプル混合体は、65 まで5分間加熱し、マイクロアレイ表面上で等分して1.8 cm² のカバーガラスで覆う。アレイは、顕微鏡用スライドより僅かに大きい空洞を有する防水チェンバーに移す。チェンバーのコーナーに140 µlの5 × SSCを加えることにより、チェンバー内部を湿度100%に保持する。アレイを含むチェンバーは、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分間洗浄し、第2洗浄緩衝液中(0.1 × SSC)において45 で10分間各々3度洗浄して乾燥させる。

【0319】

50

検出

レポーター標識ハイブリダイゼーション複合体は、Cy3の励起のためには488nm、Cy5の励起のためには632nmでスペクトル線を発生し得る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。励起レーザー光の焦点をアレイ上に置くため、20×顕微鏡対物レンズ (Nikon, Inc., Melville NY) を用いる。アレイを含むスライドを、顕微鏡の、コンピュータ制御のX-Yステージに置き、対物レンズを通してラスタースキャンする。本実施例で用いる1.8cm×1.8cmのアレイは、解像度20μmでスキャンする。

【0320】

2つの異なるスキャンのうち、混合ガスマルチラインレーザーは2つの蛍光色素を連続的に励起する。発光された光は、波長に基づき分離され、2つの蛍光色素に対応する2つの光電子増倍管検出器 (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) に送られる。好適なフィルタ群をアレイと光電子増倍管との間に設置して、シグナルをフィルタする。用いる蛍光色素の最大発光の波長は、Cy3では565nm、Cy5では650nmである。装置は両方の蛍光色素からのスペクトルを同時に記録し得るが、レーザー源において好適なフィルタを用いて、蛍光色素1つにつき1度スキャンし、各アレイを通常2度スキャンする。

10

【0321】

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対照種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配列が含まれ、その位置におけるシグナルの強度をハイブリダイジング種の重量比1:100,000に相関させる。異なる源泉 (例えば試験される細胞及び対照細胞など) からの2つのサンプルを、各々異なる蛍光色素で標識し、他と異なって発現した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する場合には、その校正を、2つの蛍光色素で校正するcDNAのサンプルを標識し、ハイブリダイゼーション混合体に各々等量を加えることによって行う。

20

【0322】

光電子増倍管の出力は、IBMコンパチブルPCコンピュータにインストールされた12ビットRTI-835Hアナログ-デジタル (A/D) 変換ボード (Analog Devices, Inc., Norwood MA) を用いてデジタル化される。デジタル化されたデータは、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へのリニア20色変換を用いてシグナル強度がマッピングされたようなイメージとして表示される。データは、定量的にも分析される。2つの異なる蛍光色素を同時に励起及び測定する場合には、各蛍光色素の発光スペクトルを用いて、データは先ず蛍光色素間の光学的クロストーク (発光スペクトルの重なりに起因する) を補正する。

30

【0323】

グリッドが蛍光シグナルイメージ上に重ねられ、それによって各スポットからのシグナルはグリッドの各エレメントに集められる。各エレメント内の蛍光シグナルは統合され、シグナルの平均強度に応じた数値が得られる。シグナル分析に用いるソフトウェアは、GEMT00LS遺伝子発現分析プログラム (Incyte) である。

40

【0324】

12 相補的ポリヌクレオチド

KPP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配列は、天然のKPPの発現を検出し、低下させ、または阻害するために用いられる。約15~30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、これより小さなあるいは大きな配列の断片の場合でも、本質的に同じ手順を用いる。Oligo4.06ソフトウェア (National Biosciences) 及びKPP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防止する。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KPPをコードする転写物に結合するのを阻害する。

50

【0325】

1.3 KPPの発現

KPP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。細菌内でKPPを発現させるためには、抗生物質耐性遺伝子と、cDNA転写レベルを高める誘導性プロモーターとを持つ、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、これらに限定するものではない。組換えベクターを、BL21(DE3)等の好適な細菌宿主に形質転換する。抗生物質耐性をもつ細菌が、イソプロピル-Dチオガラクトピラノシド(IPTG)で誘発されるとKPPを発現する。真核細胞でのKPP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組換えAutographica californica核多角体病ウイルス(AcMNPV)を感染させて行う。バキュロウイルスの非必須の多角体遺伝子を、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、KPP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、強力な多角体プロモーターによって高レベルのcDNA転写が行われる。組換えバキュロウイルスは、多くの場合は夜蛾の1種Spodoptera frugiperda(Sf9)昆虫細胞への感染に用いるが、ヒト肝細胞への感染に用いることもある。後者の感染の場合は、バキュロウイルスの更なる遺伝的変更が必要になる(Engelhard, E. K.他(1994) Proc. Natl. Acad. Sci. USA 91:3224-3227, Sandig, V.他(1996) Hum. Gene Ther. 7:1937-1945等を参照)。

10

【0326】

殆どの発現系では、KPPが、例えばグルタチオンSトランスフェラーゼ(GST)と、または、FLAGや6-Hisなどペプチドエピトープ標識と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を迅速に1ステップで行い得る。GSTは日本住血吸虫からの26kDaの酵素であり、タンパク質の活性および抗原性を維持した状態で、固定化したグルタチオン上での融合タンパク質の精製を可能とする(Amersham Pharmacia Biotech)。精製後、GSTの部分を特定の開発部位においてKPPからタンパク質的に切断することが可能である。FLAGは8アミノ酸のペプチドであり、市販されているモノクローナル及びポリクローナル抗FLAG抗体(Eastman Kodak)を用いて免疫親和性精製を可能にする。6ヒスチジン残基が連続して伸長した6-Hisは、金属キレート樹脂上での精製を可能にする(QIAGEN)。タンパク質の発現および精製の方法は、前出のAusubel(1995)10章、16章に記載されている。これらの方法で精製したKPPを直接用いて以下の実施例17、18、19、20および21のアッセイを行うことができる。

20

30

【0327】

1.4 機能的アッセイ

KPP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KPPをコードする配列の発現によって評価する。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cDNAをサブクローニングする。選択されるベクターには、pCMV SPO RTプラスミド(Life Technologies)及びpCR 3.1プラスミド(Invitrogen)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有する。リポソーム製剤あるいは電気穿孔法を用いて、5~10 μ gの組換えベクターをヒト細胞株、例えば内皮由来または造血由来の細胞株に、一過的に形質移入する。更に、標識タンパク質をコードする配列を含む1~2 μ g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、形質移入細胞と非形質移入細胞を区別する手段が与えられる。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。標識タンパク質は、例えば緑色蛍光タンパク質(GFP; Clontech)、CD64またはCD64-GFP融合タンパク質から選択できる。自動化された、レーザ光学に基づく技術であるフローサイトメトリー(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、それらの細胞のアポトーシス状態や他の細胞特性を評価する。FCMは、細胞死に先行するか或いは同時に発生する現象を診断する蛍光分子の取込みを検出して計量する。このような現象として挙げられる

40

50

のは、プロピジウムヨウ化物によるDNA染色によって計測される核DNA内容物の変化、前方光散乱と90°側方光散乱によって計測される細胞サイズと顆粒状性の変化、プロモデオキシウリジンの取込み量の低下によって計測されるDNA合成の下方調節、特異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におけるタンパク質の発現の変化、及び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ある。フローサイトメトリー法については、Ormerod, M.G. (1994) *Flow Cytometry*, Oxford, New York NYに記述がある。

【0328】

遺伝子発現におけるKPPの影響は、KPP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) の保存された複数の領域に結合する。形質転換された細胞と形質転換されない細胞とは、ヒトIgGがCD64に対する抗体のどちらかで被覆された磁気ビードを用いて分離することができる (DYNAL, Lake Success, NY)。mRNAは、当分野で周知の方法で細胞から精製することができる。KPP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。

10

【0329】

15 KPP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; Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495等を参照) または他の精製技術を用いて実質上精製されたKPPを用いて、標準プロトコルでウサギを免疫化して抗体を産出する。

20

【0330】

或いは、レーザGENEソフトウェア (DNASTAR) を用いてKPPアミノ酸配列を解析し、免疫抗原性の高い領域を決定する。そして対応するオリゴペプチドを合成し、このオリゴペプチドを用いて当業者によく知られている方法で抗体を生成する。適切なエピトープ、例えばC末端付近或いは隣接する親水性領域にあるエピトープの選択については、当分野で公知である (前出のAusubel, 1995, 11章等を参照)。

【0331】

通常は、長さ約15残基のオリゴペプチドを、Fmocケミストリを用いるABI 431A ペプチドシンセサイザ (Applied Biosystems) を用いて合成し、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) を用いた反応によってKLH (Sigma-Aldrich, St. Louis MO) に結合させて、免疫原性を高める (前出のAusubel, 1995等を参照)。完全フロイント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KPP活性を検査するには、ペプチドまたはKPPを基質に結合し、1%BSAを用いてブロックし、ウサギ抗血清と反応させて洗浄し、さらに放射性ヨウ素で標識したヤギ抗ウサギIgGと反応させる。

30

【0332】

16 特異的抗体を用いる天然KPPの精製

天然または組換えKPPを、KPP特異抗体を用いたイムノアフィニティークロマトグラフィにより実質的に精製する。イムノアフィニティークラムは、抗KPP抗体を活性化クロマトグラフィ用樹脂、例えばCNBr活性化セファロース (Amersham Pharmacia Biotech) と共有結合させることにより構築する。結合後に、製造者の使用説明書に従ってこのレジンをブロックし、洗浄する。

40

【0333】

KPPを含む培養液をイムノアフィニティークラムに通し、KPPを優先的に吸着する条件下 (例えば洗浄剤が存在する高イオン強度緩衝液) でカラムを洗浄する。そのカラムを、抗体とKPP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塩で) 溶出させ、KPPを回収する。

【0334】

50

1 7 KPPと相互作用する分子の同定

KPPまたは生物学的に活性であるKPP断片を、 ^{125}I ボルトンハンター試薬で標識する。(例えばBolton A.E.およびW.M. Hunter (1973) Biochem. J. 133:529-539を参照)。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KPPと共にインキュベートし、洗浄して、標識したKPP複合体を有する全てのウェルをアッセイする。様々なKPP濃度で得られたデータを用いて、候補分子と結合したKPPの数量及び親和性、会合についての値を計算する。

【0335】

別法では、KPP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。

10

【0336】

KPPはまた、ハイスループットな方法で酵母2ハイブリッドシステムを使用するPATHCALLINGプロセス(CuraGen Corp., New Haven CT)に用いて、遺伝子の2大ライブラリにコードされる遺伝子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)。

【0337】

1 8 KPP活性の実証

通常、プロテインキナーゼ活性は、 $[\text{-}^{32}\text{P}]\text{ATP}$ の存在で、KPPによりタンパク質基質のリン酸化を定量化することで測定される。KPPを、タンパク質基質、 ^{32}P -ATP、および或る適切なキナーゼバッファと共にインキュベートする。基質に組み込まれた ^{32}P を、電気泳動法で遊離 ^{32}P -ATPより分離し、組み込まれた ^{32}P をラジオアイソトープカウンターでカウントする。組み込まれた ^{32}P の量は、KPPの活性に比例する。リン酸化された特定のアミノ酸残基の決定は、加水分解タンパク質のホスホアミド酸解析によってなされる。

20

【0338】

或る実施態様では、プロテインキナーゼ活性はガンマリリン酸塩がアデノシン三リン酸(ATP)からタンパク質基質中のセリン、トレオニン、またはチロシン残基に転移する量を定量化することによって測定される。反応はビオチン標識されたペプチド基質を有するプロテインキナーゼサンプルとガンマ ^{32}P -ATPとの間で起こる。反応に続き、溶液中のアビジンがビオチン標識された ^{32}P ペプチド生成物に結合させる目的で添加される。結合試料は、次に生成物アビジン複合体を保持し、遊離のガンマ ^{32}P -ATPの透過させる膜を用いて、遠心限外濾過プロセスを行う。残余として ^{32}P ペプチド生成物を含む遠心分離ユニットのリザーバーは、次にシンチレーションカウンタでカウントされる。この方法は、選択されたペプチド基質及びキナーゼ反応バッファによってあらゆるタイプのプロテインキナーゼサンプルのアッセイを可能とする。このアッセイは、キット形態(ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, 米国特許番号 5,869,275)で提供される。限定されてはいないが、示唆された基質及びその各々の酵素は、ヒストンH1(シグマ)及びp34^{cdc2}キナーゼ、アネキシンI、アンジオテンシン(シグマ)及びEGF受容体キナーゼ、アネキシンII及びsrcキナーゼ、ERK1及びERK2基質及びMEK、ミエリン塩基性蛋白質及びERKである(Pearson, J. D.他(1991) Methods Enzymol. 200 : 62-81)。

30

40

【0339】

別の実施態様では、KPPのプロテインキナーゼ活性は、KPP、50 μl のキナーゼバッファ、ミエリン塩基性蛋白質(MBP)若しくは合成ペプチド基質のような1 μg の基質、1mMのDTT、10 μg のATP、及び0.5. 5 μCi の $[\text{-}^{32}\text{P}]\text{ATP}$ を含むアッセイに於いてin vitroで示される。反応液を30 で30分間インキュベートし、ピペットでP81ペーパーに移して反応を停止させる。組み込まれていない $[\text{-}^{32}\text{P}]\text{ATP}$ は、洗浄によって取り除かれ、組み込まれた放射活性はシンチレーションカウンタを用いて測定される。別法では、SDSローディングバッファの存在下で100 に加熱して反応を停止し、オートラジオグラフにより

50

12% SDSポリアクリルアミドゲル上で分離させる。組み込まれた放射活性はPKINの不在下、若しくは不活性キナーゼK38Aの存在下で実行される反応のために修正されても良い。組み込まれた³²Pの量は、KPPの活性に比例する。

【0340】

さらに別の実施例では、KPPのアデニレートキナーゼ若しくはグアニル酸キナーゼ活性が、ガンマラジオアイソトープカウンターを用いて、[γ -³²P]ATPからの³²PのADP若しくはGDPへの取り込みによって測定でき得る。キナーゼバッファー中のKPPは、適切なヌクレオチドモノリン酸塩基質（AMP若しくはGMP）及びリン酸塩ドナーとしての³²p標識ATPと共にインキュベートされる。反応は37℃でインキュベートされ、トリクロロ酢酸の添加によって終了される。酸抽出物は、中和され、ゲル電気泳動にかけられて、一リン酸、二リン酸、三リン酸のヌクレオチド画分に分離する。二リン酸ヌクレオチド画分が取り除かれ、カウントされる。回復した放射活性はKPPの酵素の活性に比例する。

10

【0341】

さらに別の実施例では、KPPのための別のアッセイは、シンチレーション近接アッセイ（SPA）、シンチレーションプレート技術、及びフィルタ結合アッセイを含む。有用な基質はグルタチオントランスフェラーゼで標識された組換えタンパク質、またはビオチンで標識された合成ペプチド基質を含む。小さな有機分子、タンパク質、及びペプチドのようなKPPの活性の阻害剤は、そのようなアッセイで同定され得る。

【0342】

別の実施例では、KPPのホスファターゼ活性は、P-ニトロフェニルリン酸（PNPP）の加水分解によって測定する。KPPを、0.1%のメルカプトエタノールの存在下で、pH7.5のHEPESバッファー中にPNPPと共に37℃で60分インキュベートする。この反応は6 mlの10N NaOHを加えて停止する（Diamond, R.H.他（1994）Mol. Cell. Biol. 14:375262）。或るいは、KPPの酸性ホスファターゼ活性は、pH4.5の0.1Mクエン酸ナトリウム、及び50 μ lの40mM塩化ナトリウム中にKPPを含む抽出液と100 μ lの10mM PNPPを37℃で20分インキュベートして実証する。この反応は0.5 mlの0.4M グリシン / NaOH（pH10.4）を加えて止める（Saffig, P. 他（1997）J. Biol. Chem. 272:1862818635）。PNPPの加水分解によって起こる410nmでの光吸収の増加を分光光度計で測定する。このアッセイでは、光吸収の増加がKPPの活性に比例する。

20

【0343】

別法では、KPP活性は、リン酸化したタンパク質基質から除去されたリン酸の量を測定して決定する。反応は、60mM Tris（pH7.6）、1mM EDTA、1mM EGTA、0.1%の γ -メルカプトエタノールおよび10 μ M 基質、適量の³²P標識したセリン / トレオニン / またはチロシンを含む最終容量30 μ lの2 nM或いは4nMのKPPにおいて行なう。反応は基質で開始し、30で10~15分間インキュベートする。0.6 M HCl、90mM Na₄P₂O₇、および2mM NaH₂PO₄中の4%（w/v）活性化木炭450 μ lで停止させ、次に12,000 \times gで5分間遠心分離する。酸可溶性³²P iを液体シンチレーションカウンターで定量する（Sinclair, C. 他（1999）J. Biol. Chem. 274:2366623672）。

30

【0344】

19 キナーゼ結合アッセイ

KPPのFLAG-CD44cyt融合タンパク質への結合は、KPPを抗KPP抱合免疫親和性ビーズとインキュベートした後、ビーズ（タンパク質の10~20ngを有する）の一部を、¹²⁵I-標識化 FLAG-CD44cyt 融合タンパク質（5,000 cpm/ngのタンパク）の存在下で、0.5mlの結合バッファ（20 mM Tris-HCL（pH 7.4）、150 mMの NaCl、0.1% ウシ血清アルブミンおよび0.05% Triton X-100）で4℃で、5時間インキュベートすることによって測定できる。結合させた後、ビーズは結合バッファー中でよく洗浄し、ビーズ結合放射活性をシンチレーションカウンターで測定する（Bourguignon, L.Y.W. 他（2001）J. Biol. Chem. 276:7327-7336）。取り込まれた³²Pの量は、結合したKPPの量に比例する。

40

【0345】

20 KPPインヒビタ-と活性化因子の同定

50

実施例 18 のアッセイで説明したように、検査されるべき化合物を、好適なバッファーおよび基質と共に様々な濃度で384-ウェルプレートのウェルに注入する。KPP活性はそれぞれのウェルについて測定し、KPP活性を阻害するそれぞれの化合物の能力および容量反応キネティクスを決定することができる。KPP活性を促進する分子の同定にも、このアッセイを用いることができる。

【0346】

KPP活性化若しくは抑制のアゴニスト若しくはアンタゴニストは、実施例 18 で記述したアッセイを用いてテストしても良い。アゴニストは、KPP活性の増加を引き起こし、アンタゴニストはKPP活性の減少を引き起こす。

【0347】

10

2.1 KPP基質の同定

KPPの「基質トラッピング (substrate-trapping)」アッセイは、タンパク質チロシンホスファターゼのPTPシグネチャ配列におけるある突然変異によって付与され得る基質親和性の上昇を利用する。このような突然変異を有するKPPは、それらの基質と安定した複合体を形成し、このような複合体を生化学的に単離し得る。KPPの触媒ドメインをコードするクローンのPTPシグネチャ配列における不変な残基の特定部位の突然変異誘発を、当分野で周知の方法或いはBIO-RADが販売するMUTA-GENEキットなどの市販のキットを用いて行うことができる。大腸菌においてKPP突然変異を誘発するために、突然変異を含むDNA断片を、グルタチオンS-トランスフェラーゼ(GST)-KPP融合タンパク質或いはKPPをコードする配列を有する発現ベクターの対応する野生型配列と交換する。KPP突然変異体が大腸菌で発現され、それを生化学的に精製する。

20

【0348】

発現ベクターを、10cmのシャーレ1枚につき20 μ gのCsCl-精製DNA、6cmのシャーレ1枚につき8 μ gのCsCl-精製DNAを用いたリン酸カルシウム仲介性トランスフェクションによってCOS1または293細胞に形質転換する。トランスフェクションから48時間経過した後に、細胞を100ng/mlの上皮成長因子で刺激すると、チロシンキナーゼEGFRがCOS細胞に多量に存在するため、細胞におけるチロシンのリン酸化が増大する。細胞を、50mM Tris HCl (pH7.5) / 5mM EDTA / 150mM NaCl / 1% Triton X-100 / 5 mMヨード酢酸 / 10mM リン酸ナトリウム / 10mM NaF / 5 μ g/mlロイペプチン / 5 μ g/mlアプロチニン / 1mM benzamidineで溶解する (10cmのシャーレ1枚につき1ml、6cmのシャーレ1枚につき0.5ml)。KPPを、好適な抗体で溶解物から免疫沈降させる。GST-KPP融合タンパク質を、細胞溶解物1mgにつきそれぞれ、グルタチオン-セファロース、4 μ gのmAbまたは10 μ lのビーズで沈降させる。複合体をポリアクリルアミド電気泳動 (PAGE) で視覚化することができるが、更に精製して基質分子を同定することもできる (Flint, A. J.他(1997) Proc. Natl. Acad. Sci. USA 94: 1680-1685)。

30

【0349】

当業者には、本発明の要旨および精神から逸脱しない範囲での、記載した本発明の方法およびシステムの、種々の修正および変更の手段は自明であろう。本発明について説明するにあたり幾つかの実施例に関連して説明を行ったが、本発明の請求の範囲が、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。分子生物学または関連分野の専門家には明らかな、本明細書に記載する本発明の実施方法の様々な修正は、明確に特許請求の範囲内にあるものとする。

40

【0350】

(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の命名法の概略を示す。

【0351】

表2は、GenBank識別番号及び本発明のポリペプチドに最も近いGenBank相同体の注釈を示す。各ポリペプチドとその相同体が一致する確率スコアも併せて示す。

【0352】

50

表 3 は、予測されるモチーフ及びドメインを含む本発明のポリペプチド配列の構造的特徴を、ポリペプチドの分析に用いるための方法、アルゴリズム及び検索可能なデータベースと共に示す。

【 0 3 5 3 】

表 4 は、本発明のポリヌクレオチド配列を構築するために用いた cDNA やゲノム DNA 断片を、ポリヌクレオチド配列の選択した断片と共に示す。

【 0 3 5 4 】

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。

【 0 3 5 5 】

表 6 は、表 5 に示した cDNA ライブラリの作製に用いた組織及びベクターを説明する付表である。

【 0 3 5 6 】

表 7 は、本発明のポリヌクレオチドとポリペプチドの分析に用いたツール、プログラム、アルゴリズムを、適用可能な説明、引用文献及び閾値パラメータと共に示す。

【 0 3 5 7 】

【 表 1 】

表 1

Incyte プロジェクト ID	ポリペプチド NO.	ポリペプチド SEQ ID	Incyte ポリペプチド ID	ポリヌクレオチド SEQ ID NO.	Incyte ポリヌクレオチド ID
55074884	1		55074884CD1	9	55074884CB1
7480588	2		7480588CD1	10	7480588CB1
7482931	3		7482931CD1	11	7482931CB1
2080788	4		2080788CD1	12	2080788CB1
71918969	5		71918969CD1	13	71918969CB1
8187571	6		8187571CD1	14	8187571CB1
7494145	7		7494145CD1	15	7494145CB1
5807954	8		5807954CD1	16	5807954CB1

20

30

40

【 0 3 5 8 】

【 表 2 】

表2

ポリペプチドSEQ ID NO:	Incyte ポリペプチド ID	GenBank ID NO またはプロテオーム ID NO:	確率スコア	注釈
1	55074884CD1	956840994	1.8E-19	[ヒト] MAP キナーゼホスファターゼ6 Martí, F. 他. (2001) J. Immunol. 166:197-206
2	7480588CD1	93300096	0	[ラット] 糸球体メサンギウム細胞受容体タンパク質 - チロシンホスファターゼ前駆体 Wright, M.B. 他. (1998) J. Biol. Chem. 273:23929-23937
3	7482931CD1	915341198	1.0E-179	[マウス] タウチユープリンキナーゼ Tomizawa, K. 他. (2001) FEBS letters. 492:221-227
4	2080788CD1	93015538	0	[ヒト] 核二重特異性ホスファターゼ Cui, X. 他. (1998) Nature Genet. 18:331-337
5	71918969CD1	914148952	1.0E-163	[マウス] セリン/トレオニンキナーゼ33 Mujica, A.O. 他. (2001) Gene 280:175-181
6	8187571CD1	92804429	5.4E-41	[線虫] プロテインホスファターゼ2Cファミリーに類似
7	7494145CD1	92392814	1.8E-103	[マウス] PFTAIRE キナーゼ Lazzaro, M.A. 他. (1997) J. Neurochem. 68:348-364.
8	5807954CD1	911245474	1.1E-90	[ヒト] ピリミジン5ヌクレオチダーゼ Amici, A. 他. (2000) Blood 96:1596-1598

10

20

30

【 0 3 5 9 】

【 表 3 - 1 】

表 3-1

SEQ ID NO	Incyte プラズマ 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	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
1	55074884CD1	176	S30 S135	N50	Signal_Leavage:M1-A36 シグナルペプチド:M1-G35 二重特異性ホスファターゼ、触媒ドメイン P18-A156 VH1 タイプ II 重特異性ホスファターゼ DM03823 P28562 169-314 V20-A159 I38890 29-320:V20-L157 A56115 1-336 V20-A159 Q02256 1-174 G97-E155	SPSCAN HMMER HMMER, PFAM BLAST, DOMO
2	7480588CD1	2299	S104 S182 S186 S200 S298 S332 S350 S490 S557 S567 S649 S661 S668 S687 S715 S733 S751 S761 S772 S862 S898 S912 S927 S1277 S1312 S1338 S1381 S1385 S1411 S1530 S1558 S1590 S1628 S1836 S1900 S1941 S1975 S1979 S1984 S2018 S2237 S2261	N155 N162 N311 N349 N384 N575 N713 N 731 N765 N 770 N809 N.895 N903 N959 N988 N996 N1008 N1038 N1059 N1071 N1161 N1194 N1248 N1253 N1292 N 1380 N1399 N 1524 N1584 N1594 N1618 N1647 N1694 N1719 N1802 N2059 N2162 N2273	シグナルペプチド:M1-T17, M1-V19	HMMER

【 0 3 6 0 】
【 表 3 - 2 】

表 3 - 2

SEQ ID NO.	Incyte 番号	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	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
2(続き)			T117 T128 T120 T208 T249 T313 T328 T365 T389 T413 T511 T607 T614 T623 T633 T660 T678 T757 T767 T824 T834 T849 T870 T907 T998 T1013 T1048 T1095 T1129 T1145 T1158 T1223 T1334 T1424 T1635 T1759 T1770 T1884 T1894 T1959 T2020 T2082 T2143 T2157 Y710 Y1940		プロテインチロシンホスファターゼ N2027-D2258	HMMER_PFAM
					プロボネチンタイプIII ドメイン:P51, S138, P665-V746, P854-R936, P759-S842, P1340-S1418, P1150-S1231, I.1243-S1328, P1430-S1326, P948-S1040, P365-S651, P1538-S1620, P300-S383, P1641-V1734, P1053-S1134, P150-G286, P394-T554 推定膜貫通セグメント-II:604-A1622, R1902-I1930	HMMER_PFAM
					推定膜貫通セグメント-II:604-A1622, R1902-I1930 チロシン特異性プロテインホスファターゼタンパク質 BL00383: K2030-V2044, S2055-I2063, Q2160-P2172, V2198-G2208, R2230-F2251 糖タンパク質チロシンキナーゼ/チロシンリン酸化酵素 BL00790:G66-V91, N115-T145 チロシン特異性タンパク質ホスファターゼシグネチャ及びプロフィール tyr_phosphatase.prf.L2178-R2236	TMAP BLIMPS_BLOCKS BLIMPS_BLOCKS PROFILES SCAN BLIMPS_PRINTS
2(続き)					プロボネチンタイプIII ドメイン:P51, S138, P665-V746, P854-R936, P759-S842, P1340-S1418, P1150-S1231, I.1243-S1328, P1430-S1326, P948-S1040, P365-S651, P1538-S1620, P300-S383, P1641-V1734, P1053-S1134, P150-G286, P394-T554 推定膜貫通セグメント-II:604-A1622, R1902-I1930 チロシン特異性プロテインホスファターゼタンパク質 BL00383: K2030-V2044, S2055-I2063, Q2160-P2172, V2198-G2208, R2230-F2251 糖タンパク質チロシンキナーゼ/チロシンリン酸化酵素 BL00790:G66-V91, N115-T145 チロシン特異性タンパク質ホスファターゼシグネチャ及びプロフィール tyr_phosphatase.prf.L2178-R2236 プロボネチンタイプIII リポドシグネチャ PR00014:T1165-P1174, Y1402-D1416	HMMER_PFAM

【 0 3 6 1 】
【 表 3 - 3 】

表 3 - 3

SEQ ID NO:	Incyte 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	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
					プロテインチロシンホスファターゼ シグネチャ PR00700 D2056-I2063, F2072-E2092, M2156-E2173, P2195-L2213, V2226-C2241, M2242-L2252	BLIMPS_PRINTS
					糸球体メサンギウム細胞受容体 プロテインチロシン ホスファターゼ 前駆体 PD178547C1913-W2026 PD174021:W1528-K1642 PD184208:T145-D205 ヒドロラーゼ ホスファターゼ タンパク質 チロシン 前駆体 PD000167:N2027-N2221	BLAST_PRODOM
					プロテインチロシンホスファターゼ DM000089 A57064I896-1170 H1991-I2256 SG0613I924-1198:H1991-I2256 I49372I113-387:H1991-I2256 I49374I1-269:L1995-L2256	BLAST_PRODOM
2(続き)					ATP/GTP 結合部位モチーフ A (P ループ) A1621-S1628 チロシン特異性タンパク質ホスファターゼ活性部位 V2198-F2210	MOTIFS MOTIFS
3	7482931CD1	478	S173 S237 S263 S278 S283 S353 T118 T155 T189 T269 T300 T402 T431 T456 T463 Y471		真核生物プロテインチロシンホスファターゼドメイン:W21-S256	HMMER_PFAM
					カゼインキナーゼに類似 PD115601:F213-E298, L11-T114	BLAST_PRODOM
					プロテインチロシンホスファターゼドメイン DM00004I42169I74-330I27-E240 DM00004I48730I1-265:R25-Y273 DM00004I35506I19-273:Y23-Y273 DM00004I40233I13-267:R25-Y273	BLAST_PRODOM

10

20

30

【 0 3 6 2 】

【 表 3 - 4 】

表 3 - 4

SEQ ID NO	Incyte ポリペプチド ID	アミノ酸残基数	着在的リン酸化部位	潜在的グリコシル化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
4	2080786CD1	1867	S75 S137 S182 S359 S430 S559 S637 S659 S716 S722 S773 S803 S812 S873 S968 S1047 S1094 S1113 S1120 S1174 S1199 S1232 S1237 S1259 S1280 S1284 S1362 S1387 S1465 S1517 S1527 S1552 S1561 S1630 T84 T93 T96 T102 T313 T538 T636 T933 T971 T1069 T1599 T1747 T1827 T1838 T1839 Y1769	N1585 N1761	プロテインキナーゼ ATP 結合領域シグネチャ I27-K50 DENN (AEX-3)ドメイン L171-G310	MOTIFS HMMER, PFAM
4(続き)					PH ドメイン R1762-S1865 膜貫通セグメント A212-S240, L246-P268, L274-F292, N629-K657, L1384-L1410 N-末端は細胞質内でない ヒドロラーゼ タンパク質 MYOTUBULARIN 疾患 突然変異 F53A2.8 タンパク質 チロシン ホスファターゼ C19A8.03 CPA2NNF1 PD014611:P1379-H1470, D1129-G1266, G1474-Y1556, G927-F938 タンパク質 シナプス前作用の調節因子 セリン プロテアーゼ I シヒター RAB3 GDP/GTP PD008900:P177-D340, R3-H115, D364-R408	HMMER, PFAM TMAP BLAST, PRODOM

10

20

30

【 0 3 6 3 】
【 表 3 - 5 】

表 3-5

SEQ ID NO:	Incyte プラトID	アミノ酸残基数	潜在的な糖化部位	潜在的グリコシ化部位	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
5	71918969CD1	514	S4, S10, S20, S29, S71, S80, S87, S169, S211, S249, S274, S296, S326, S349, S388, S407, S420, S441, S473, T11, T70, T134, T184, T371, T412, T440, T450, T459, T510	N95, N213, N411, N438	ロイジンジッパー/偽-L289-L260 真核生物プロテインキナーゼドメイン: Y116-L381	MOTIFS HMMER-PFAM
					プロテインキナーゼシグネチャおよびプロファイル(C): E214-A269 プロテインキナーゼ触媒ドメインシグネチャ: PR00109.M192-D205, Y228-V246, C304-S326, T184-R206 プロテインキナーゼドメイン: DM00004.S5734721-2666.F118-T371 プロテインキナーゼATP結合領域シグネチャ, L122-K145	PROFILESCAN BLIMPS-PRINTS BLAST-DOMO MOTIFS
5(続き)					セリン/トレオニンプロテインキナーゼ活性部位シグネチャ I234-V246 プロテインホスファターゼ2C: S63-R128, V212-A282	MOTIFS HMMER-PFAM
6	8187571CD1	338	S55 S220 S281 S320 T134 T155 T246 T268 T305 Y248	N178	膜貫通ドメイン: V89-I117 N末端は細胞質内 タンパク質ホスファターゼ2C タンパク質 BL01032.L64-A8.I, N109-E148, R205-D218, D253-D265, S320-L329	TMAP BLIMPS_BLOCKS

【 0 3 6 4 】

【 表 3 - 6 】

表 3-6

SEQ ID NO:	Incyte 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	シグネチャ配列、ドメインおよびモチーフ	分析方法及びデータベース
7	7494145CD1	321	S19 S24 S228 Y63		タンパク質ホスファターゼ2C DM00377P496961-295 S55-G143, R205-S281 DM00377Q09173 1-296 Y66-L141, E188-A282 DM00377S6246211-297:Y66-L141, E188-A282 真核生物プロテインキナーゼドメイン Y52-P275 膜貫通ドメイン:D231-L248; N-末端は細胞質内	HMMER-PFAM TMAP
7(純 き)					プロテインキナーゼシグネチャおよびプロファイル: N149-A197 キナーゼトランスフェラーゼタンパク質 セリン/トレオニン タンパク質 ATP 結合 II リン酸化カゼインアルファ鎖 PD002608:Q203-P281 プロテインキナーゼドメイン: DM00004, Q00536 166-436:L53- L305	PROFILESCAN BLAST-PRODOM BLAST-DBOMO
8	5807954CD1	292	S32 S200 S223 T86 T228	N198	プロテインキナーゼ ATP 結合領域シグネチャ: L58-K81 セリン/トレオニンプロテインキナーゼ活性部位シグネチャ V169-I181 膜貫通ドメイン: N135-K163 N-末端は細胞質内	MOTIFS MOTIFS TMAP

【 0 3 6 5 】

【 表 4 - 1 】

表 4 - 1

飛行記録ID/配列長	飛行記録ID NO/ hrcyte 飛行記録ID	配列断片
9/55074884CBI/ 1605		1-361, 1-830, 4-517, 188-830, 346-731, 437-731, 440-493, 440-495, 440-516, 440-517, 440-528, 440-731, 440-1205, 440-1285, 443-731, 450-731, 480-731, 520-774, 526-774, 540-774, 541-731, 544-731, 596-731, 616-731, 645-731, 651-731, 672-731, 683-731, 777-1605, 1026-1191, 1052-1191, 1102-1377, 1337-1440
10/7480588CBI/ 7225		1-148, 1-460, 159-377, 376-894, 646-1584, 1351-2448, 1517-1882, 1700-2082, 1883-1955, 1883-2556, 1883-3153, 2242-2678, 2272-2680, 2719-3351, 2986-3351, 3154-3444, 3171-3283, 3171-3477, 3171-3513, 3171-3522, 3171-3551, 3171-3573, 3171-3604, 3171-3620, 3171-3646, 3171-3680, 3171-3689, 3179-3689, 3202-3689, 3220-3682, 3242-3689, 3244-3689, 3246-3691, 3446-4285, 3459-4284, 4234-4915, 4234-4925, 4235-4920, 4237-4925, 4245-4925, 4254-4925, 4271-4925, 4288-5687, 4291-4925, 4297-4925, 4303-4925, 4307-4925, 4324-4925, 4330-4509, 4330-4678, 4330-4685, 4382-4677, 4382-4681, 4382-4682, 4382-4684, 4382-4685, 4384-4684, 4405-4683, 4435-4684, 4467-4680, 5110-5794, 5110-5852, 5110-5854, 5110-5862, 5110-5908, 5110-5916, 5111-5854, 5230-6114, 5444-6104, 5686-6861, 6047-6781, 6047-6784, 6047-6834, 6047-6840, 6047-6854, 6047-6863, 6047-6885, 6047-6891, 6470-7225, 6495-7225, 6507-7225, 6510-7225, 6541-7225, 6545-7225, 6567-7225, 6582-7225, 6738-6864, 6861-6883
11/7482931CBI/ 1691		1-243, 1-576, 32-235, 93-230, 95-378, 186-377, 307-703, 307-763, 307-803, 307-836, 307-845, 307-1049, 307-1054, 307-1056, 307-1063, 307-1076, 307-1078, 307-1080, 307-1112, 307-1127, 308-845, 311-372, 377-787, 377-797, 427-797, 465-1132, 577-1082, 757-1630, 816-1618, 822-1608, 826-1624, 878-1567, 887-1314, 907-1614, 933-1614, 935-1167, 940-1608, 982-1085, 982-1543, 1020-1691, 1021-1691, 1038-1691, 1052-1604, 1053-1431, 1053-1614, 1053-1628, 1053-1642, 1053-1667, 1053-1680, 1053-1691, 1054-1691, 1055-1680, 1055-1691, 1063-1691, 1072-1691, 1163-1691

【 0 3 6 6 】
【 表 4 - 2 】

表 4 - 2

ボリスレオ子ド ID/配列番号	SEQ ID NO./ Incyte ボリスレオ子ド ID/配列番号	配列断片
12/2080788CB1/6146		1-540, 4-335, 4-580, 11-216, 11-464, 19-359, 19-647, 240-412, 265-595, 271-719, 343-931, 446-1064, 522-1140, 539-821, 617-1206, 661-1241, 664-977, 696-1261, 696-1293, 700-1230, 756-1259, 756-1364, 801-1374, 814-1393, 848-1376, 953-1448, 956-1526, 974-1304, 1005-1465, 1005-1655, 1192-1838, 1206-1743, 1234-1784, 1244-1439, 1258-1819, 1259-1634, 1263-1524, 1271-1543, 1302-1734, 1307-1906, 1319-1672, 1362-1865, 1363-1784, 1368-1749, 1377-1610, 1378-1589, 1378-1640, 1378-1643, 1388-2021, 1415-1689, 1415-1756, 1415-1816, 1415-1986, 1429-1972, 1466-1973, 1475-1756, 1501-1769, 1525-1926, 1527-2112, 1539-1844, 1559-2049, 1584-1782, 1584-2191, 1584-2213, 1586-1828, 1586-2154, 1590-2153, 1603-2222, 1616-2168, 1643-2102, 1671-2212, 1692-2313, 1715-2249, 1734-2404, 1743-2341, 1754-2015, 1777-2361, 1779-2378, 1800-2352, 1800-2422, 1806-2253, 1806-2271, 1812-2343, 1836-2130, 1837-2140, 1840-2110, 1845-2434, 1846-2188, 1848-2188, 1848-2387, 1863-2420, 1878-2304, 1883-2411, 1889-2475, 1890-2426, 1892-2508, 1901-2401, 1906-2401, 1909-2415, 1919-2169, 1919-2436, 1921-2533, 1927-2271, 1945-2507, 1951-2454, 1972-2417, 1987-2515, 2028-2611, 2036-2562, 2036-2641, 2041-2292, 2041-2337, 2043-2658, 2045-2620, 2068-2509, 2100-

10

20

30

【 0 3 6 7 】
【 表 4 - 3 】

表 4 - 3

ポリヌクレオチド SEQ ID NO./ Incyte ポリヌクレオチド ID/ 配列長	配列断片
12(続き)	
	2699, 2134-2674, 2149-2396, 2149-2397, 2149-2416, 2149-2425, 2149-2427, 2149-2428, 2149-2431, 2158-2360, 2170-2734, 2180-2769, 2188-2433, 2188-2550, 2188-2648, 2199-2532, 2206-2847, 2208-2847, 2213-2847, 2216-2790, 2218-2764, 2221-2693, 2226-2847, 2229-2842, 2229-2847, 2230-2602, 2231-2847, 2235-2842, 2235-2847, 2242-2847, 2247-2790, 2255-2842, 2258-2847, 2259-2847, 2259-2847, 2290-2847, 2295-2780, 2296-2846, 2308-2847, 2309-2842, 2310-2605, 2312-2847, 2314-2596, 2321-2842, 2322-2586, 2337-2847, 2339-2616, 2348-2838, 2356-2816, 2357-2847, 2360-2829, 2364-2847, 2365-2824, 2365-2847, 2366-2847, 2406-2838, 2413-2829, 2413-2841, 2417-2829, 2417-2847, 2421-2847, 2424-2756, 2425-2500, 2431-2658, 2441-2780, 2443-2847, 2452-2846, 2476-2847, 2478-2847, 2479-3129, 2482-2785, 2487-2784, 2496-2735, 2496-2847, 2507-2846, 2523-2744, 2523-2847, 2539-2842, 2540-2847, 2559-3031, 2569-2847, 2572-2847, 2575-2847, 2577-2847, 2601-2811, 2637-2847, 2719-3332, 2751-2828, 2801-3359, 2840-3228, 2840-3266, 2840-3288, 2840-3293, 2840-3303, 2840-3443, 2840-3452, 2843-3221, 2908-3491, 2928-3321, 2936-3257, 2961-3245, 3333-3753, 3354-3585, 3398-3932, 3446-

10

20

30

【 0 3 6 8 】
【 表 4 - 4 】

表 4 - 4

ポリシクロオキソド ID/ 配列長	SEQ ID NO./ Inoyte	ポリシクロオキソド 配列断片				
3942,	3447-3714,	3617-4204,	3637-3753,	3665-3887,	3748-3772,	3905-
4425,	4011-4321,	4045-4664,	4173-4445,	4186-4463,	4197-4459,	4244-
4523,	4250-4489,	4250-4794,	4255-4510,	4296-4536,	4296-4814,	4302-
4514,	4303-4596,	4304-4773,	4385-4655,	4398-4632,	4493-4768,	4519-
4771,	4526-4784,	4526-5055,	4538-4940,	4542-4689,	4542-4924,	4550-
4784,	4589-5001,	4594-4855,	4640-4899,	4659-4958,	4689-4874,	4731-
5023,	4737-4982,	4757-5017,	4764-5064,	4777-5026,	4777-5196,	4784-
5043,	4788-5087,	4793-5051,	4793-5297,	4800-5254,	4819-5107,	4832-
5140,	4846-5072,	4901-5155,	4907-5184,	4910-5177,	4960-5351,	5006-
5236,	5020-5245,	5022-5314,	5022-5632,	5040-5324,	5055-5256,	5123-
5372,	5135-5405,	5153-5394,	5183-5481,	5213-5456,	5214-5456,	5216-
5500,	5226-5424,	5282-5555,	5311-5569,	5320-5559,	5322-5524,	5322-
5798,	5368-5623,	5380-5698,	5440-5914,	5452-5654,	5463-5709,	5463-
6000,	5464-5722,	5468-5722,	5468-5815,	5484-5784,	5502-5736,	5525-
5722,	5525-5786,	5526-5892,	5539-6109,	5543-5800,	5544-6102,	5554-
6085,	5588-6077,	5601-5813,	5601-5910,	5604-5739,	5604-5864,	5604-
5871,	5604-6115,	5605-5876,	5619-6103,	5637-5889,	5638-6146,	5667-
6104,	5669-5853,	5673-5918,	5740-5970,	5760-6048,	5790-6128,	5909-
6092,	5933-6130,	5939-				
6127,	5946-6124,	5990-6124,	5981-6131			

【 0 3 6 9 】
【 表 4 - 5 】

10

20

30

表 4 - 5

ポリヌクレオチド SEQ ID NO./ Incyte ポリヌクレオチド ID / 配列長	配列断片
13/71918969CB1/ 2362	1-341, 1-450, 1-488, 1-546, 1-580, 1-590, 1-753, 25-325, 25-443, 30-341, 66-338, 249-958, 249-958, 251-597, 457-951, 457-969, 457-1091, 520-785, 523-1161, 580-1032, 606-1177, 625-1147, 630-883, 631-1368, 634-1297, 755-1422, 762-1323, 816-1493, 818-1529, 822-1520, 829-1113, 829-1339, 842-1520, 847-1120, 865-1395, 915-1584, 920-1638, 936-1575, 951-1623, 995-1622, 1003-1594, 1021-1735, 1030-1620, 1082-1820, 1086-1769, 1111-1785, 1135-1309, 1145-1309, 1169-1309, 1169-1769, 1192-1620, 1198-1897, 1205-1899, 1209-1309, 1218-1785, 1220-1920, 1222-1460, 1222-1505, 1223-1309, 1224-1309, 1235-1307, 1235-1754, 1243-1831, 1247-1309, 1248-1862, 1250-1309, 1253-1860, 1276-1961, 1277-1908, 1291-1866, 1301-1703, 1311-1929, 1358-1948, 1386-1926, 1454-1969, 1455-1698, 1468-1871, 1477-2276, 1481-1677, 1482-2128, 1498-2010, 1504-2130, 1504-2152, 1511-2169, 1511-2245, 1544-2319, 1557-2310, 1566-2134, 1580-2240, 1581-2325, 1590-2152, 1597-2132, 1602-2325, 1620-2266, 1624-2171, 1632-2277, 1643-2325, 1650-1903, 1660-2151, 1666-2054, 1685-2325, 1686-1906, 1686-1920, 1686-2163, 1686-2206, 1699-1925, 1718-2174, 1722-2171, 1748-2020, 1749-2170, 1756-2338, 1761-2325, 1783-2042, 1804-2166, 1805-2166, 1812-2212, 1827-2362, 1828-2173, 1890-2325, 1900-2325, 1901-2077, 1903-2325, 1920-2166
14/8187571CB1/ 1535	1-617, 142-828, 175-330, 178-348, 181-459, 239-814, 258-487, 270-795, 311-855, 400-640, 400-983, 557-812, 711-938, 715-988, 783-1079, 803-1371, 810-1509, 854-1431, 886-1528, 1022-1519, 1089-1533, 1124-1535, 1131-1527, 1133-1379, 1145-1373, 1182-1531, 1188-1531, 1215-1336, 1215-1518, 1215-1528, 1215-1535, 1216-1531, 1242-1533, 1254-1528, 1259-1325, 1277-1532, 1294-1526, 1330-1531, 1331-1535, 1343-1531
15/7494145CB1/ 1376	1-489, 149-469, 153-207, 161-213, 189-521, 268-417, 268-778, 288-812, 268-951, 268-982, 556-1142, 556-1145, 556-1147, 556-1149, 568-1149, 842-1147, 864-1154, 904-1376

【 0 3 7 0 】
【 表 4 - 6 】

表 4 - 6

ポリ双レオナド SEQ ID NO./ Ineyte ポリ双レオナド ID/ 配列番号	配列断片
16/5607954CBI/ 1482	1-41, 1-298, 14-485, 34-381, 69-264, 144-642, 191-891, 196-436, 238-482, 259-460, 259-470, 259-473, 271-575, 283-635, 293-581, 293-961, 359-570, 359-639, 378-639, 382-1008, 438-1020, 483-762, 562-891, 635-879, 640-895, 653-1172, 654-941, 654-1003, 691-938, 886-1117, 906-1155, 909-1156, 918-1146, 929-1191, 932-1122, 932-1178, 932-1186, 932-1238, 934-1173, 945-1136, 995-1429, 1007-1272, 1009-1227, 1009-1236, 1009-1381, 1009-1408, 1009-1447, 1138-1477, 1170-1333, 1237-1482, 1294-1477

10

20

30

【 0 3 7 1 】

【 表 5 】

表5

ポリスケレオチド SEQ ID NO:	Inevite プロジェクトID	代表的ライブラリ
9	55074884CB1	HEARNON03
10	7480588CB1	LUNGFER04
11	7482931CB1	BRAQNOT01
12	2080788CB1	BRAUNOR01
13	71918969CB1	TESTNOT03
14	8187571CB1	UTRSNOT01
15	7494145CB1	BRSTTMC01
16	5807954CB1	LNODNOT03

【 0 3 7 2 】
【 表 6 - 1 】

表 6 - 1

ライブラリ	ハンダー pINCY	ライブラリの説明
BRAQNOT01	pINCY	ライブラリは 35 才の白人男性から採取した中脳組織から単離した RNA を用いて作製した。神経病理学的には何も見つからなかった。当患者の病歴には、孤発型心筋症、鬱血性心不全、及び肥大した脾臓と肝臓が含まれる。
BRAUNOR01	pINCY	このランダムプライムされたライブラリは、アテローム性動脈硬化症による胸部大動脈破裂による出血で死亡の 81 才白人女性から得た線条体、淡室球および縦隔組織から単離した RNA を用いて作製した。病理学検査では、双方の内部類動脈に中程度のアテローム性動脈硬化、前頭皮質および海馬に微視的な梗塞、および散在広汎性のアミロイド斑、および神経原線維障害体が見られたが、これは年齢と矛盾しない。肉眼で見ても、軟膜は上矢状静脈洞に沿った軽度の硝子質化と厚化が見られるのみであった。残りの軟膜は薄く、充血した血管がいくらか見られた。主に前頭葉および前頭葉、及び側頭葉(双方)に軽度の萎縮が見られた。微視的には、アルツハイマー病 II 型状細胞のペアが新皮質の深層にあった。中央前頭皮質の灰白質の奥深部のニューロンの周りでサテライトシス(satellitosis)が増大していた。扁桃にまれな広汎性筋と神経原線維障害性を含んでいた。後部海馬は反応性グリオシスに囲まれたハモゲリアン含有マクロファージを伴う微視的な萎縮性の空洞化領域を含んでいた。
BRSTTMC01	pINCY	患者の病歴は、敗血症、胆管炎、術後無気肺、肺炎冠動脈疾患(OAD)、左心室肥大に起因する心肥大、巨脾症、縮動脈性管硬化、結節性コロイド甲状腺腫、気腫、鬱血性心不全(CHF)、甲状腺機能低下症および抹消血管疾患を含んでいた。この大型分画ライブラリは 6 人の供与者から 5 のプールされた cDNA を用いて作製した。cDNA は 40 才白人女性(供与者 A)の側縮小乳房形成術時に切除した病変した乳房組織、および 46 才白人女性(供与者 B)の乳房癌を伴う片側性拡大単純乳房切除術時に摘出された乳房組織、56 才白人女性(供与者 C)の乳房癌を伴う片側性拡大単純乳房切除術時及び側縮小乳房形成術時に切除した乳房組織、および 57 才白人女性(供与者 D)の片側性拡大単純乳房切除術時に摘出された乳房組織から単離された mRNA を用いて作製した。病理学検査では、両側性軽度線維芽細胞性変化及び増殖性変化(A)、胸部筋膜は腫瘍について陰性(B)、非増殖性線維芽細胞性変化、良性脂肪肉腫に置換された乳房癌(D)を示した。一致する腫瘍組織(B)の病理はアポクリン特徴を有する増殖性 3 の腺癌、管型を示した。一致する腫瘍組織(C)の病理は、浸潤性 3 の管内腺癌を示した。
BRSTTMC01 (続く)		患者の病歴には、乳房肥大と純粋高コレステロール血症(A); 乳癌(B); 慢性的気道閉塞と気腫(C); 良性高血圧症、高脂質血症、不整脈、良性の血腫腫瘍、孤立乳房癌、不明な行動の乳房腫瘍がある。以前に受けた手術には、直視下乳房生検がある(B)。提供者 B の服用薬剤には、Cytosol, Adriamycin がある。
HEARNON03	pINCY	この標準化した心臓組織ライブラリは、心臓組織ライブラリの 840 万個の独立クローンから作製した。開始 RNA は、顕微鏡内出で死亡した 44 才の白人男性から採取した心臓組織から作製した。血清学的検査は抗 CMV (サイトメガロウイルス)に陽性であった。患者の病歴には、腰痛と頭痛、高血圧症、肺炎、副鼻副感染症、アルコール摂取、日常的なパイプタバコ使用(3 年)がある。患者の使用薬剤には Procainamide がある。このライブラリは、極めて長時間(48 時間/1 回)の再アニーリングによるハイ

【 0 3 7 3 】
【 表 6 - 2 】

表 6 - 2

ライブラリ	ベクター	ライブラリの説明
LNODNOT03	pINCY	ライブラリの説明 フリダイゼーションを用いた他は、Soares 他、PNAS (1994) 91:3228-9232 及び Bonaldo 他、Genome Research (1996) 6:791 を適用した条件を用いて 2 回にわたリノーマライズした。 ライブラリは、気管支鏡法による肺の部分除去中の 67 才白人男性から取り除かれたリンパ腺組織から単離された RNA を用いて作製された。顕微鏡検査では、この組織は 10% の生存腫瘍のある強度な萎縮であることが発見された。関連する腫瘍組織の病理学検査では、グレード 3 から 4 の扁平上皮細胞の癌腫性病が呈された。患者の病歴には血管腫が含まれる。家歴には、アテローム硬化型冠状動脈疾患、良性の高血圧、鬱血性心不全症、およびアテローム硬化型冠状動脈疾患が含まれる。
LUNGFER04	PCDNA2.1	このランダムライブラリは胎児死亡の白人男子胎児から摘出した肺組織から単離された RNA を用いて作成した。
TESTNOT03	PBLUESCRIPT	ライブラリは肝疾患で死亡した 37 才の白人男性から採取した精巣組織から単離した RNA を用いて作製した。患者の病歴には、肝硬変、黄疸、肝不全があった。
UTRSNOT01	PSPORT1	ライブラリは、心筋梗塞で死亡した 58 才女性の子宮組織から単離された RNA を用いて作製された。患者の病歴には、心筋症、冠状動脈疾患、以前の心筋梗塞症、高コレステロール血症、低血圧症、閉塞性腎臓病がある。

【 0 3 7 4 】
【 表 7 - 1 】

10

20

30

【 0 3 7 5 】
 【 表 7 - 2 】

表 7 - 1

プログラム	説明	参考文献	パラメータ関連値
ABI FACTURA	ベクター配列を除き、核酸配列のままの短い塩基をマスキングするプログラム	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、注釈付けに有用である。	Applied Biosystems, Foster City, CA. Paracel Inc., Pasadena, CA.	不一致<5%
ABI AutoAssembler	核酸配列を構築するプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	Basic Local Alignment Search Tool はアミノ酸および核酸配列の配列類似性検索に有用であり、BLASTには5つの機能がある: blastp, blastn, blastx, tblastnおよびtblastx。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410, Altschul, S.F. 他 (1997) Nucleic Acids Res. 25:3389-3402.	ES7:標準値=1.0E-8 以下 完全長配列 標準値= 1.0E-10以下
FASTA	問合せ配列と同種の配列群配との類似性を検索する Pearson およびLipman アルゴリズム。FASTAは最低5つの機能からなる: fasta, tfasta, fastx, tfastxおよびsssearch。	Pearson, W.R.及びD.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol.163:63-96, Smith, T.F.及びM.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ES7:fasta E 値 =1.06E-6 構築された ES7:fasta 同一性=95%以上 一致した長さ=200塩基以上 fasta E 値=1.0E-8以下 完全長配列: fastx スコア=100以上
BLIMPS	配列をBLOCKS、PRINTS、DOMO、PRODOMおよびPFAM データベースの配列と対応させて遺伝子ファミリー、配列相同性および構造的プラインガープリント領域を検索するBLOCKS IMPROVED サーチャー。	Henikoff, S.及びJ.G. Henikoff (1991) Nucleic Acids Res.19:6565-6572, Henikoff, J.G.及びS. Henikoff (1996) Methods Enzymol.266:88-105, Attwood, T.K. 他.(1997) J. Chem. Inf. Comput. Sci. 37:417-424.	標準値= 1.0E-3以下
HMMER	PFAMのような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. 1-350ページ。	PFAMヒット標準値=1.0E-3 以下 シングルレブレットヒット:スコア=0以上

10

20

30

表7-2

プログラム	説明	参考文献	パラメータ閾値
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の構造的モチーフおよび配列モチーフを検索するアルゴリズム。	Gribskov, M. 他 (1988) CABIOS 4:61-66, Gribskov, M. 他 (1989) Methods Enzymol 183:146-159, Barroch, A. 他 (1997) Nucleic Acids Res. 25:217-221.	標準化された質のスコアとその特定の Prosite モチーフに対するGCG指定 HIGH 値 通常、スコア=1.4~2.1
Phred	高い感度と確率で自動配列決定機のアレースを調べる塩基読出しアルゴリズム。	Ewing, B. 他 (1998) Genome Res. 8:175-185, Ewing, B. 及び P. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的実行に基づいたプログラムである、SWATとCrossMatchを含むPhrap 改訂構築プログラムで、配列相溶性検索やDNA配列の構築に有用である。	Smith, T.F. および M.S. Waterman (1981) Adv. Appl. Math. 2:482-489, Smith, T.F. および M.S. Waterman (1981) J. Mol. Biol. 147:195-197; Green, P., University of Washington, Seattle, WA.	スコア=120以上 一致した長さ=56以上
Consed	Phrap で構築したものの表示、編集用グラフィカルツール。	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ックス解析プログラム。	Nielson, H. 他 (1997) Protein Engineering 10:1-6, Claverie, J.M. 及び S. Audic (1997) CABIOS 12:431-439.	スコア=3.5以上
TMAP	加重マトリックスを用いてタンパク質配列での膜貫通セグメントの描写および配向を決定するプログラム。	Persson, B. および P. Argos (1994) J. Mol. Biol. 237:182-192, Persson, B. および P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	隠れMarkov モデル(HMM) を使ってタンパク質配列上の膜貫通部分を描写し、方向を決定するプログラム。	Sonnhammer, E.L. 他 (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow 他, 編集, The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, 175-182ページ。	
Motifs	Prositeで定義された配列と一致したパターンのアミノ酸配列を検索するプログラム。	Barroch, A. 他 (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, 第9版, M51-59ページ, Genetics Computer Group, Madison, WI	

10

20

30

【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
14 November 2002 (14.11.2002)

PCT

(10) International Publication Number
WO 02/090530 A2

- (51) International Patent Classification: C12N 9/00
Creek Road, Fremont, CA 94555 (US). **TANG, Tom, Y.** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US).
YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US); **TRAN, Bao** [US/US]; 750 Siltberg Avenue, Santa Clara, CA 95051 (US); **DING, Li** [US/US]; 3353 Aina Street #146, Palo Alto, CA 94306 (US); **LIU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US); **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US); **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US).
- (21) International Application Number: PCT/US02/01369
- (22) International Filing Date: 16 January 2002 (16.01.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/263,083 18 January 2001 (18.01.2001) US
60/271,205 23 February 2001 (23.02.2001) US
60/271,117 23 February 2001 (23.02.2001) US
60/276,859 16 March 2001 (16.03.2001) US
60/278,504 23 March 2001 (23.03.2001) US
60/278,522 23 March 2001 (23.03.2001) US
60/280,510 29 March 2001 (29.03.2001) US
60/280,266 29 March 2001 (29.03.2001) 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, EC, EE, ES, FI, GB, GD, GE, GR, GM, HN, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NZ, NI, 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, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- (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): **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US); **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US); **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US); **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US); **GURURAJAN, Rajagopal** [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US); **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US); **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US); **JACKSON, Jennifer, L.** [US/US]; 4913 Deep
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/090530 A2

(54) Title: KINASIS AND PHOSPHATASES

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

WO 02/090530

PCT/US02/01369

KINASES AND PHOSPHATASES**TECHNICAL FIELD**

This invention relates to nucleic acid and amino acid sequences of kinases and phosphatases and to the use of these sequences in the diagnosis, treatment, and prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of kinases and phosphatases.

10

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups. Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur, as well as kinases sensitive to second messenger molecules. This system allows for the amplification of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell cycle progression.

25

KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response

35

WO 02/090530

PCT/US02/01369

to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

5 Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked
10 to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine
15 residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to
20 the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence
25 motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in
30 Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

35 **Protein Tyrosine Kinases**

WO 02/090530

PCT/US02/01369

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) *Annu. Rev. Cell Biol.* 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates

WO 02/090530

PCT/US02/01369

its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase domain as is typical with other MAP kinases (Abe, M.K. (1999) *Mol. Cell. Biol.* 19:1301-1312).

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) *J. Biol. Chem.* 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) *J. Biol. Chem.* 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, α , β , γ , δ , and ϵ . Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state.

WO 02/090530

PCT/US02/01369

Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) Science 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) Biochimie 82:1123-1127).

HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) J. Biol. Chem. 273:25875-25879; Wang, Y. et al. (2001) Biochim. Biophys. Acta 1518:168-172). HIPKs act as corepressors for homeodomain transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:12350-12355).

The human h-warts protein, a homolog of Drosophila warts tumor suppressor gene, maps to chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus (Nishiyama, Y. et al. (1999) FEBS Lett. 459:159-165).

Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and

WO 02/090530

PCT/US02/01369

dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

Mitogen-Activated Protein Kinases

5 The mitogen-activated protein kinases (MAP), which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). There are three kinase modules comprising the MAP kinase cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang, X.S. et al (1998) Biochem. Biophys. Res. Commun. 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun 15 N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

20 Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the 25 mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation 30 of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome 35 segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint

WO 02/090530

PCT/US02/01369

pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol. Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This

WO 02/090530

PCT/US02/01369

C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share
5 homology with the death-associated protein kinases (DAP kinases), known to function in interferon- γ induced apoptosis (Sanjo et al., supra). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., supra). However, deletion of either the
10 N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a
15 member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal
20 kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., supra).

25 Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the
30 mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory
35 enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to

WO 02/090530

PCT/US02/01369

a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) supra).

KINASES WITH NON-PROTEIN SUBSTRATES

Lipid and Inositol kinases

10 Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) Curr. Opin. Cell Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP₂). PIP₂ is then cleaved into inositol triphosphate (IP₃) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) biphosphate (PIP₂) to PI (3,4,5) P₃ (PIP₃). PIP₃ then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, supra).

35 An example of lipid kinase phosphorylation activity is the phosphorylation of

WO 02/090530

PCT/US02/01369

D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP
5 regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive
10 inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., supra).

Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism
15 and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in
20 transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R. J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be
30 treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenylyl cyclase, and production
35 of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and

WO 02/090530

PCT/US02/01369

GTP levels also control the activity of certain oncogenic proteins such as p21^{ras} known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) *Cancer Res.* 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21^{ras} and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buclovir (Miller, W.H. and R.L. Miller (1980) *J. Biol. Chem.* 255:7204-7207; Stenberg, K. et al. (1986) *J. Biol. Chem.* 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) *Pharmacol. Ther.* 67:155-186).

20

PHOSPHATASES

Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca²⁺ or Mn²⁺, for activity. PSPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) *Annu. Rev. Biochem.* 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly

35

WO 02/090530

PCT/US02/01369

conserved signature sequence, coupled with one or more regulatory subunits (PROSITE PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

- 5 PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-
- 10 monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) *Curr. Opin. Neurobiol.* 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) *Otolaryngol. Head Neck Surg.* 121:463-468).
- 15 PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) *Trends Biosci.* 24:186-191). These "B-type" subunits are cell type- and tissue-
- 20 specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII
- 25 kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) *J. Virol.* 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the
- 30 activities of more than 30 protein kinases *in vitro*, and other evidence suggests that the same is true *in vivo* for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases, cyclin-dependent kinases, and the I κ B kinases (reviewed in Millward et al., *supra*). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-
- 35 (Bastians, H. et al. (1999) *Mol. Biol. Cell* 10:3927-3941). PP2A is a major activity in the brain and is

WO 02/090530

PCT/US02/01369

implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, *supra*).

5 PP2B, or calcineurin, is a Ca^{2+} -activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of
10 transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, *supra*).

15 Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids *in vitro* and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell
20 Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not
25 been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble
30 phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204).
Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a
35 transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic

WO 02/090530

PCT/US02/01369

domain. Although only the first copy seems to have enzymatic activity, the second copy apparently affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, *supra*). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, *supra*). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990) Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) Genomics 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, *supra*).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, *supra*). In the activated state, these kinases have been implicated

WO 02/090530

PCT/US02/01369

in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) *J. Biol. Chem.* 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) *Hum. Pathol.* 30:419-424) and abnormalities in its expression are associated with numerous cancers (reviewed in Tamura, M. et al. (1999) *J. Natl. Cancer Inst.* 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

LAP, an orthophosphoric monoester of the endosomal/lysosomal compartment is a housekeeping gene whose enzymatic activity has been detected in all tissues examined (Geier, C. et al. (1989) *Eur. J. Biochem.* 183:611-616). LAP-deficient mice have progressive skeletal disorder and an increased disposition toward generalized seizures (Saftig, P. et al. (1997) *J. Biol. Chem.* 272:18628-18635). LAP-deficient patients were found to have the following clinical features: intermittent vomiting, hypotonia, lethargy, opisthotonos, terminal bleeding, seizures, and death in early infancy (Online Mendelian Inheritance in Man (OMIM) *200950).

PAP, a prostate epithelium-specific differentiation antigen produced by the prostate gland, has been used to diagnose and stage prostate cancer. In prostate carcinomas, the enzymatic activity of PAP was shown to be decreased compared with normal or benign prostate hypertrophy cells (Foti, A. G. et al. (1977) *Cancer Res.* 37: 4120-4124). Two forms of PAP have been identified, secreted and intracellular. Mature secreted PAP is detected in the seminal fluid and is active as a glycosylated homodimer with a molecular weight of approximately 100-kilodalton. Intracellular PAP is found to exhibit endogenous phosphotyrosyl protein phosphatase activity and is involved in regulating prostate cell growth (Meng, T.C. and Lin, M.F. (1998) *J. Biol. Chem.* 34: 22096-22104).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) *FEBS Lett.* 419:175-180). Additionally,

WO 02/090530

PCT/US02/01369

synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) *Curr. Opin. Neurobiol.* 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (*Synj1*) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that *Synj1* can act as a negative regulator of membrane-coat protein interactions. These findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling (Cremona, O. et al. (1999) *Cell* 99:179-188).

The discovery of new kinases and phosphatases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of kinases and phosphatases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, kinases and phosphatases, referred to collectively as "KPP" and individually as "KPP-1," "KPP-2," "KPP-3," "KPP-4," "KPP-5," "KPP-6," "KPP-7," and "KPP-8." 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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-8.

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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-8. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:9-16.

WO 02/090530

PCT/US02/01369

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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. 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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

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:9-16, 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:9-16, 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.

WO 02/090530

PCT/US02/01369

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:9-16, 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:9-16, 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 polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, 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:9-16, 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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

WO 02/090530

PCT/US02/01369

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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. 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 KPP, 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 amino acid sequence selected from the group consisting of SEQ ID NO:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. 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 treating a disease or condition associated with overexpression of functional KPP, 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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test

WO 02/090530

PCT/US02/01369

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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, 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.

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 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, 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:9-16, 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 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:9-16, 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:9-16, iii) a polynucleotide

WO 02/090530

PCT/US02/01369

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 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

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

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

15 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 polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

20 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.

25 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

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

35 It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a

WO 02/090530

PCT/US02/01369

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 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 connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"KPP" refers to the amino acid sequences of substantially purified KPP 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 KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

An "allelic variant" is an alternative form of the gene encoding KPP. 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 KPP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a polypeptide with at least one functional characteristic of KPP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding KPP. 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 KPP. Deliberate amino acid substitutions may be made on the basis of similarity in

WO 02/090530

PCT/US02/01369

polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

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

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

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of KPP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP 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 KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

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

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a

WO 02/090530

PCT/US02/01369

- specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include
- 5 deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system.
- 10 Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) *J. Biotechnol.* 74:5-13.)
- The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) *Proc. Natl Acad. Sci. USA* 96:3606-3610).
- 15 The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.
- The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense
- 25 molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.
- 30 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 KPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.
- 35 "Complementary" describes the relationship between two single-stranded nucleic acid

WO 02/090530

PCT/US02/01369

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 KPP or fragments of KPP 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
25	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
30	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
35	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
40	Trp	Phe, Tyr
	Tyr	His, Phe, Trp

WO 02/090530

PCT/US02/01369

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,
5 (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.

10 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.

15 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
20 diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

25 A "fragment" is a unique portion of KPP or the polynucleotide encoding KPP 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,
30 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the
35 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present

WO 02/090530

PCT/US02/01369

embodiments.

A fragment of SEQ ID NO:9-16 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:9-16, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:9-16 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:9-16 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:9-16 and the region of SEQ ID NO:9-16 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:1-8 is encoded by a fragment of SEQ ID NO:9-16. A fragment of SEQ ID NO:1-8 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-8. For example, a fragment of SEQ ID NO:1-8 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-8. The precise length of a fragment of SEQ ID NO:1-8 and the region of SEQ ID NO:1-8 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

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

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

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

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms

WO 02/090530

PCT/US02/01369

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:

Matrix: BLOSUM62
Reward for match: 1
Penalty for mismatch: -2
Open Gap: 5 and Extension Gap: 2 penalties
Gap x drop-off: 50
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.

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

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative

WO 02/090530

PCT/US02/01369

substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

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

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

Matrix: BLOSUM62
Open Gap: 11 and Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
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, 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.

"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.

WO 02/090530

PCT/US02/01369

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
5 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.

10 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
15 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,
20 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
25 circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid
30 sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_{0t} or R_{0t} 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).

35 The words "insertion" and "addition" refer to changes in an amino acid or nucleotide

WO 02/090530

PCT/US02/01369

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
5 cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of KPP 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 KPP which is useful in any of the antibody production methods disclosed herein or known in the
10 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 KPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of KPP.
15

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
20 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
25 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.
30 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 KPP 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
35 vary by cell type depending on the enzymatic milieu of KPP.

WO 02/090530

PCT/US02/01369

“Probe” refers to nucleic acid sequences encoding KPP, 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.

5 “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).

10 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.

15 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
20 can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to
25 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from
30 megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “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
35 programs may also be obtained from their respective sources and modified to meet the user’s specific

WO 02/090530

PCT/US02/01369

needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both
5 unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

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

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

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

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

30 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 KPP,
35 nucleic acids encoding KPP, or fragments thereof may comprise a bodily fluid; an extract from a cell,

WO 02/090530

PCT/US02/01369

chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

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

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

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

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic 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.

A "transcript image" or "expression profile" 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

WO 02/090530

PCT/US02/01369

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 alternate splicing of exons during mRNA processing. The corresponding polypeptide may 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.

35 THE INVENTION

WO 02/090530

PCT/US02/01369

The invention is based on the discovery of new human kinases and phosphatases (KPP), the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

5 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
10 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
15 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 scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

20 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
25 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
30 properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:1 is 33% identical to human MAP kinase phosphatase 6 (GenBank ID g6840994) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.3e-19$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a dual specificity phosphatase catalytic domain as determined
35 by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM

WO 02/090530

PCT/US02/01369

database of conserved protein family domains. (See Table 3.) Data from BLAST analysis of the DOMO database provides further corroborative evidence that SEQ ID NO:1 is a MAP kinase phosphatase.

In an alternative example, SEQ ID NO:2 is 86% identical, from residue M1 to residue
5 M2299, to *Rattus norvegicus* glomerular mesangial cell receptor protein-tyrosine phosphatase
precursor (GenBank ID g3300096) as determined by the Basic Local Alignment Search Tool
(BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of
obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a
protein-tyrosine phosphatase domain as determined by searching for statistically significant matches
10 in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains.
(See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further
corroborative evidence that SEQ ID NO:2 is a protein-tyrosine phosphatase.

In an alternative example, SEQ ID NO:3 is 96% identical, from residue M1 to residue Q316,
to a mouse tau tubulin kinase (GenBank ID g15341198) as determined by the Basic Local Alignment
15 Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.0e-179$, which indicates the
probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also
contains a eukaryotic protein kinase 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.) Data from MOTIFS, and additional BLAST analyses provide further
20 corroborative evidence that SEQ ID NO:3 is a protein kinase.

In an alternative example, SEQ ID NO:4 is 98% identical, from residue H234 to residue
C1863, to human nuclear dual-specificity phosphatase (GenBank ID g3015538) as determined by the
Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0,
which indicates the probability of obtaining the observed polypeptide sequence alignment by chance.
25 SEQ ID NO:4 also contains a DENN (AEX-3) domain and a pleckstrin homology 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.) Data from BLAST
analysis using the PRODOM database provides further corroborative evidence that SEQ ID NO:4 is a
nuclear phosphatase.

In an alternative example, SEQ ID NO:5 is 76% identical, from residue M1 to residue T382,
30 to mouse serine/threonine kinase 33 (GenBank ID g14148952) as determined by the Basic Local
Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.0e-163$, which
indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ
ID NO:5 also contains a protein kinase domain as determined by searching for statistically significant
35 matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family

WO 02/090530

PCT/US02/01369

domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:5 is a protein kinase.

In an alternative example, SEQ ID NO:6 is 42% identical, from residue V107 to residue L331, to a *Caenorhabditis elegans* protein similar to the protein phosphatases 2c family (GenBank ID g2804429) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $5.4e-41$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a protein phosphatase 2C 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.) Data from BLIMPS and BLAST_DOMO analyses provide further corroborative evidence that SEQ ID NO:6 is a protein phosphatase 2C.

In an alternative example, SEQ ID NO:7 is 75% identical, from residue F46 to residue V292, to *Mus musculus* PFTAIRE kinase (GenBank ID g2392814) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.3e-103$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 is also 73% identical, from residue F46 to residue R296, to human serine/threonine protein kinase PFTAIRE-1 (GenBank ID g12002201), with a BLAST probability score of $9.0e-103$. SEQ ID NO:7 also contains a protein kinase 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.) Data from MOTIFS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is a protein kinase.

In an alternative example, SEQ ID NO:8 is 58% identical, from residue K2 to residue L281, to human pyrimidine 5'-nucleotidase (GenBank ID g11245474) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.1e-90$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. The algorithms and parameters for the analysis of SEQ ID NO:1-8 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. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:9-16 or that

WO 02/090530

PCT/US02/01369

distinguish between SEQ ID NO:9-16 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_XXXXX_{N₁}_N₂_YYYY_{N₂}_N₄ represents a "stitched" sequence in which XXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N_{1,2,3...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXX_gAAAAA_gBBBBB₁_N is a "stretched" sequence, with XXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or PGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).

WO 02/090530

PCT/US02/01369

GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

5 In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 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 10 the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses KPP variants. A preferred KPP variant is one which has at 15 least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the KPP amino acid sequence, and which contains at least one functional or structural characteristic of KPP.

The invention also encompasses polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected 20 from the group consisting of SEQ ID NO:9-16, which encodes KPP. The polynucleotide sequences of SEQ ID NO:9-16, 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 KPP. In 25 particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding KPP. 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:9-16 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% 30 polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:9-16. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of KPP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant

WO 02/090530

PCT/US02/01369

of a polynucleotide sequence encoding KPP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding KPP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about
5 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of
10 the polynucleotide sequence encoding KPP. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of KPP.

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 KPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be
15 produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring KPP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode KPP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding KPP 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
25 occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP 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 KPP and KPP 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
30 into a sequence encoding KPP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of
35

WO 02/090530

PCT/US02/01369

hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:9-16 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

5 "Definitions."

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

10 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA

15 sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) *Short Protocols in Molecular Biology*, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, pp. 856-853.)

20 The nucleic acid sequences encoding KPP 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.)

25 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,

30 M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries

35 (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries

WO 02/090530

PCT/US02/01369

and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

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

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

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

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter KPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, 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 No. 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

WO 02/090530

PCT/US02/01369

improve the biological properties of KPP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of 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 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.

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

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

In order to express a biologically active KPP, the nucleotide sequences encoding KPP 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 inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding KPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where

WO 02/090530

PCT/US02/01369

sequences encoding KPP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate 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 KPP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding KPP. 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 KPP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding KPP can be achieved using a

WO 02/090530

PCT/US02/01369

multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPO1 plasmid (Life Technologies). Ligation of sequences encoding KPP 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 KPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of KPP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

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

Plant systems may also be used for expression of KPP. Transcription of sequences encoding KPP 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 KPP 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 KPP 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

WO 02/090530

PCT/US02/01369

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 KPP in cell lines is preferred. For example, sequences encoding KPP 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 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 *apr* 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 KPP is inserted within a marker gene sequence, transformed cells containing sequences encoding KPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP 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 KPP and that express

WO 02/090530

PCT/US02/01369

KPP 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.

5 Immunological methods for detecting and measuring the expression of KPP 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 antibodies reactive to two non-interfering epitopes on KPP is preferred, but a competitive
10 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.)

15 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 KPP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding KPP, or any fragments thereof, may be cloned into a vector for
20 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
25 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 KPP 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
30 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode KPP may be designed to contain signal sequences which direct secretion of KPP 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
35 the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation,

WO 02/090530

PCT/US02/01369

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 KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric KPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of KPP 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 KPP encoding sequence and the heterologous protein sequence, so that KPP 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 KPP 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.

KPP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to KPP. At least one and up to a plurality of test compounds may be screened for specific binding to KPP. 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 KPP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a

WO 02/090530

PCT/US02/01369

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 KPP 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 KPP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing KPP or cell membrane fractions which contain KPP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either KPP 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 KPP, either in solution or affixed to a solid support, and detecting the binding of KPP 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.

KPP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the presence of a test compound is compared with the activity of KPP in the absence of the test compound. A change in the activity of KPP in the presence of the test compound is indicative of a compound that modulates the activity of KPP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of KPP 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 KPP 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 No. 5,175,383 and U.S. Patent No. 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

WO 02/090530

PCT/US02/01369

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).

5 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 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 KPP may also be manipulated in vitro in ES cells derived from
10 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 KPP can also be used to create "knockin" humanized animals
15 (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding KPP 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.
20 Alternatively, a mammal inbred to overexpress KPP, e.g., by secreting KPP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists
between regions of KPP and kinases and phosphatases. In addition, examples of tissues expressing
25 KPP are brain, testes, uterine, breast and lymph node tissues and also can be found in Table 6. Therefore, KPP appears to play a role in cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. In the treatment of disorders associated with increased KPP expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders
30 associated with decreased KPP expression or activity, it is desirable to increase the expression or activity of KPP.

Therefore, in one embodiment, KPP 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 KPP.
Examples of such disorders include, but are not limited to, a cardiovascular disease such as
35 arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial

WO 02/090530

PCT/US02/01369

dissections, varicose 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, 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, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism.

10 pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis,

15 hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung

20 transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema,

25 episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic

30 anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other

35 extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive

WO 02/090530

PCT/US02/01369

neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-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; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, 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 glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as

WO 02/090530

PCT/US02/01369

actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocytopenia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

In another embodiment, a vector capable of expressing KPP 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 KPP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified KPP 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 KPP including, but not limited to, those provided above.

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

In a further embodiment, an antagonist of KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP. Examples of such disorders include, but are not limited to, those cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express KPP.

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

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

WO 02/090530

PCT/US02/01369

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

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

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to KPP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of KPP 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 KPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; 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.)

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

WO 02/090530

PCT/US02/01369

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

5 Antibody fragments which contain specific binding sites for KPP may also be generated. For example, such fragments include, but are not limited to, F(ab)₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab)₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. 10 et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between KPP and its specific 15 antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an 20 association constant, K_a , which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a 25 particular KPP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the KPP-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 KPP, preferably in active form, from the antibody 30 (Catty, D. (1988) *Antibodies, Volume I: A Practical Approach*, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) *A Practical Guide to Monoclonal Antibodies*, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For 35 example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml,

WO 02/090530

PCT/US02/01369

preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of KPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

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

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

In another embodiment of the invention, polynucleotides encoding KPP may be used for
25 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),
30 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), thalassamias, 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
35 cell proliferation), or (iii) express a protein which affords protection against intracellular parasites

WO 02/090530

PCT/US02/01369

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

Expression vectors that may be effective for the expression of KPP include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA 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). KPP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the 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 H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding KPP from a normal individual.

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

WO 02/090530

PCT/US02/01369

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to KPP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding KPP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and 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 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 No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, C. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. 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 No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP to cells of the central nervous system, for which HSV has a

WO 02/090530

PCT/US02/01369

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 No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 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 KPP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (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., protease and polymerase). Similarly, inserting the coding sequence for KPP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding RNAs and the synthesis of high levels of KPP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly,

WO 02/090530

PCT/US02/01369

inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.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.

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 KPP.

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 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 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 KPP. 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.

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 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.

WO 02/090530

PCT/US02/01369

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

At least one, and up to a plurality, of test compounds may be screened for effectiveness in 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 library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding KPP 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 KPP are assayed by 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 KPP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) *Nucleic Acids Res.* 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) *Biochem. Biophys. Res. Commun.* 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide

WO 02/090530

PCT/US02/01369

sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

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

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

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

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

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

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

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of

WO 02/090530

PCT/US02/01369

the macromolecule. Alternatively, KPP 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).

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

10 A therapeutically effective dose refers to that amount of active ingredient, for example KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose
15 lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD₅₀/ED₅₀ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity.
20 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
25 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
30 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.

WO 02/090530

PCT/US02/01369

DIAGNOSTICS

In another embodiment, antibodies which specifically bind KPP may be used for the diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP 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.

5
10
15
A variety of protocols for measuring KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding KPP 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 KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of KPP, and to monitor regulation of KPP levels during therapeutic intervention.

20
25
30
In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding KPP or closely related molecules may be used to identify nucleic acid sequences which encode KPP. 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 KPP, 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 KPP 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:9-16 or from genomic sequences including promoters, enhancers, and introns of the KPP gene.

35
Means for producing specific hybridization probes for DNAs encoding KPP include the

WO 02/090530

PCT/US02/01369

cloning of polynucleotide sequences encoding KPP or KPP 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 KPP may be used for the diagnosis of disorders associated with expression of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose 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, 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, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasma pneumoniae, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, 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,

WO 02/090530

PCT/US02/01369

glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hyper eosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, cephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, 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, cranioretardation, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine

WO 02/090530

PCT/US02/01369

palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and 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 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, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. The polynucleotide sequences encoding KPP 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 KPP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding KPP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding KPP 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 KPP 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 KPP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification.

WO 02/090530

PCT/US02/01369

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. 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.

5 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.

10 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.

15 Additional diagnostic uses for oligonucleotides designed from the sequences encoding KPP 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 KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding KPP, 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.

20 In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding KPP 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 KPP 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 using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis

25
30
35

WO 02/090530

PCT/US02/01369

methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry
5 singling, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis,
10 sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the
15 anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu
20 (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of KPP 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
25 accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray
30 can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the
35 activities of therapeutic agents in the treatment of disease. In particular, this information may be used

WO 02/090530

PCT/US02/01369

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.

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

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of
10 gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 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
15 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,
20 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
25 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
30 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
35 comparison of expression data after treatment with different compounds. While the assignment of

WO 02/090530

PCT/US02/01369

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
5 <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
10 the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present
15 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
20 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
25 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
30 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.

35 A proteomic profile may also be generated using antibodies specific for KPP to quantify the

WO 02/090530

PCT/US02/01369

levels of KPP 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; Mendozze, 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 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

WO 02/090530

PCT/US02/01369

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 KPP 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.I. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding KPP 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, KPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug

WO 02/090530

PCT/US02/01369

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 KPP 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 KPP, or fragments thereof, and washed. Bound KPP is then detected by methods well known in the art. Purified KPP 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 KPP specifically compete with a test compound for binding KPP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with KPP.

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

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

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/263,083, U.S. Ser. No. 60/271,205, U.S. Ser. No. 60/271,117, U.S. Ser. No. 60/276,859, U.S. Ser. No. 60/278,504, U.S. Ser. No. 60/278,522, U.S. Ser. No. 60/280,510 and U.S. Ser. No. 60/280,266 are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). 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

WO 02/090530

PCT/US02/01369

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 SUPERSCRIP^T plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the 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), PCR2-TOPO^TA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pLGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), 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

WO 02/090530

PCT/US02/01369

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.

5 Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABICATALYST 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
10 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 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
15 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
20 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; PROTEOME databases with sequences from Homo sapiens,
Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae,
25 Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); 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
30 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
35 polynucleotide sequences were translated to derive the corresponding full length polypeptide

WO 02/090530

PCT/US02/01369

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, the PROTEOME databases, 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 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).

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:9-16. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative kinases and phosphatases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides were analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases were also identified by homology to Incyte cDNA sequences that had been annotated as kinases and phosphatases. These selected Genscan-predicted sequences were then compared by

WO 02/090530

PCT/US02/01369

BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public 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 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

WO 02/090530

PCT/US02/01369

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

10 VI. Chromosomal Mapping of KPP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:9-16 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:9-16 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.

30 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.)

35 Analogous computer techniques applying BLAST were used to search for identical or related

WO 02/090530

PCT/US02/01369

molecules in cDNA databases such as GenBank or LIFESEQ (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:

5

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

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and 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 KPP 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 KPP. cDNA sequences and cDNA library/tissue

WO 02/090530

PCT/US02/01369

information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of KPP 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.

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 Mg²⁺, (NH₄)₂SO₄, 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 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones

WO 02/090530

PCT/US02/01369

were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:9-16 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46

WO 02/090530

PCT/US02/01369

female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:9-16 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. 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 technologies should be uniform and solid with a non-porous surface (Schna (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g.,

WO 02/090530

PCT/US02/01369

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 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 fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia

WO 02/090530

PCT/US02/01369

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 U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, 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 μ l of sample mixture consisting of 0.2 μ g 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 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.

WO 02/090530

PCT/US02/01369

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).

XII. Complementary Polynucleotides

Sequences complementary to the KPP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring KPP. 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 KPP. 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 KPP-encoding transcript.

XIII. Expression of KPP

WO 02/090530

PCT/US02/01369

Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (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 KPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of KPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding KPP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (SF9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, KPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from KPP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified KPP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, XX and XXI where applicable.

XIV. Functional Assays

KPP function is assessed by expressing the sequences encoding KPP 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 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

WO 02/090530

PCT/US02/01369

formulations or electroporation. 1-2 μg 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 KPP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding KPP 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 KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of KPP Specific Antibodies

KPP 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 KPP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

WO 02/090530

PCT/US02/01369

increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-KPP activity by, for example, binding the peptide or KPP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring KPP Using Specific Antibodies

Naturally occurring or recombinant KPP is substantially purified by immunoaffinity chromatography using antibodies specific for KPP. An immunoaffinity column is constructed by covalently coupling anti-KPP 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 KPP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of KPP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/KPP 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 KPP is collected.

XVII. Identification of Molecules Which Interact with KPP

KPP, 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 KPP, washed, and any wells with labeled KPP complex are assayed. Data obtained using different concentrations of KPP are used to calculate values for the number, affinity, and association of KPP with the candidate molecules.

Alternatively, molecules interacting with KPP 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).

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by KPP in the presence of [γ -³²P]ATP. KPP is incubated with the protein substrate, ³²P-ATP, and an appropriate kinase buffer. The ³²P incorporated into the substrate is separated from free ³²P-ATP by electrophoresis and the incorporated ³²P is counted using a radioisotope counter. The

WO 02/090530

PCT/US02/01369

amount of incorporated ^{32}P is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P -ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated ^{32}P -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P -ATP. The reservoir of the centrifuged unit containing the ^{32}P -peptide product as retentate is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and p34^{cdc2}kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) *Methods Enzymol.* 200:62-81).

In another alternative, protein kinase activity of KPP is demonstrated in vitro in an assay containing KPP, 50 μl of kinase buffer, 1 μg substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 μg ATP, and 0.5 μCi [γ - ^{32}P]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ - ^{32}P]ATP is removed by washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a 12% SDS polyacrylamide gel followed by autoradiography. Incorporated radioactivity is corrected for reactions carried out in the absence of PKIN or in the presence of the inactive kinase, K38A. The amount of incorporated ^{32}P is proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be measured by the incorporation of ^{32}P from [γ - ^{32}P]ATP into ADP or GDP using a gamma radioisotope counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and ^{32}P -labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is proportional to the activity of KPP.

In yet another alternative, other assays for KPP include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant

WO 02/090530

PCT/US02/01369

proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β -mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100 μ l of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μ l of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Safitig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of KPP in the assay.

In the alternative, KPP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final volume of 30 μ l containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β -mercaptoethanol and 10 μ M substrate, 32 P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30°C for 10-15 min. Reactions are quenched with 450 μ l of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $\text{Na}_4\text{P}_2\text{O}_7$, and 2 mM NaH_2PO_4 , then centrifuged at 12,000 \times g for 5 min. Acid-soluble 32 Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XIX. Kinase Binding Assay

Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of 125 I-labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated 32 P is proportional to the amount of bound KPP.

XX. Identification of KPP Inhibitors and Activators

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVIII. KPP activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules

WO 02/090530

PCT/US02/01369

which enhance KPP activity.

Agonists or antagonists of KPP activation or inhibition may be tested using assays described in section XVIII. Agonists cause an increase in KPP activity and antagonists cause a decrease in KPP activity.

5 **XXI. Identification of KPP Substrates**

A KPP "substrate-trapping" assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of KPP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in *Escherichia coli*, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in *E. coli* and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated transfection with 20 µg of CsCl-purified DNA per 10-cm dish of cells or 8 µg per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris-HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 µg/ml leupeptin/5 µg/ml aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 µg of mAb or 10 µl of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

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 02/090530

PCT/US02/01369

Table 1

Lucyte Project ID	Polypeptide SEQ ID NO:	Lucyte Polypeptide ID	Polynucleotide SEQ ID NO:	Lucyte Polynucleotide ID
55074884	1	55074884CD1	9	55074884CB1
7480588	2	7480588CD1	10	7480588CB1
7482931	3	7482931CD1	11	7482931CB1
2080788	4	2080788CD1	12	2080788CB1
71918969	5	71918969CD1	13	71918969CB1
8187571	6	8187571CD1	14	8187571CB1
7494145	7	7494145CD1	15	7494145CB1
5807954	8	5807954CD1	16	5807954CB1

WO 02/090530

PCT/US02/01369

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO. or PROTEOME ID NO.	Probability Score	Annotation
1	55074884CD1	g6646994	1.3E-19	[Homo sapiens] MAP kinase phosphatase 6 Martí, F. et al. (2001) J Immunol 166:197-206
2	7480388CD1	g3300096	0	[Rattus norvegicus] glomerular mesangial cell receptor protein-tyrosine phosphatase precursor Wright, M.B. et al. (1998) J. Biol. Chem. 273:23929-23937
3	7482931CD1	g15341198	1.0E-179	[Mus musculus] Tauubulin kinase Tomizawa, K., et al. (2001) FEBS letters. 492:221-227
4	2080788CD1	g3015538	0	[Homo sapiens] nuclear dual-specificity phosphatase Cui, X. et al. (1998) Nucleic Acids Res. 26:331-337
5	7191896CD1	g14148952	1.0E-163	[Mus musculus] Serine/Threonine kinase 33 Mujica, A.O., et al. (2001) Gene 280:175-181
6	8187571CD1	g2804429	5.4E-41	[Caenorhabditis elegans] similar to the protein phosphatase 2c family Lazzano, M.A. et al. (1997) J. Neurochem. 69:348-364.
7	7494145CD1	g2392814	1.3E-103	[Mus musculus] PPTAIRE kinase Lazzano, M.A. et al. (1997) J. Neurochem. 69:348-364.
8	5807954CD1	g11245474	1.1E-90	[Homo sapiens] pyrimidine 5'-nucleotidase Amici, A. et al. (2000) Blood 96:1596-1598

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	5507484CD1	176	S30 S135	S50	Signal_cleavage: M1-A36 Signal Peptide: M1-G35 P18-A156 Dual specificity phosphatase, catalytic domain: VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DMG3823 P25562169-314; V20-A159 I38890129-320; Y20-L157 A5611511-336; Y20-A159 Q0225611-174; G97-E155	SPSCAN HMMER HMMER_PPAM BLAST_DOMO
2	7480588CD1	2299	S104 S182 S186 S200 S298 S332 S350 S490 S557 S567 S649 S661 S668 S687 S715 S733 S751 S761 S772 S862 S898 S912 S927 S1277 S1312 S1338 S1381 S1385 S1411 S1530 S1558 S1590 S1628 S1836 S1900 S1941 S1975 S1979 S1984 S2018 S2237 S2261	N155 N162 N311 N349 N384 N575 N713 N731 N765 N770 N809 N895 N903 N959 N988 N996 N1008 N1038 N1059 N1071 N1161 N1194 N1248 N1253 N1292 N1380 N1399 N1524 N1584 N1594 N1618 N1647 N1694 N1719 N1802 N2059 N2162 N2273	Signal Peptide: M1-T17, M1-V19	HMMER

WO 02/090530

PCT/US02/01369

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2 (cont.)			T117 T28 T120 T208 T249 T313 T328 T365 T389 T413 T511 T607 T614 T623 T633 T660 T678 T757 T767 T824 T834 T849 T870 T907 T998 T1013 T1048 T1095 T1129 T1145 T1158 T1223 T1334 T1424 T1635 T1759 T1770 T1884 T1894 T1959 T2020 T2082 T2145 T2157 Y710 Y1940		Protein-tyrosine phosphatase: N027-D258	HMMER_PPAM
					Fibronectin type III domain: P51-S138, P665-V746, P854-R956, P759-S842, P1340-S1418, P1150-S1231, L1243-S1298, P1430-S1526, P948-S1040, P505-S651, P1538-S1620, P309-S383, P1641-V1734, P1053-S1134, P150-G286, F394-T534 Predicted transmembrane segments: T1604-A1622, R1902-I1930	HMMER_PPAM
					Tyrosine specific protein phosphatases: proteins BL00383, K2030-V2044, S2055-I2063, Q2160, P2172, V2198-G2208, R2236-F2251	TMAP BLIMP8_BLOCKS

WO 02/090530

PCT/US02/01369

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	(cont.)				Receptor tyrosine kinase class V proteins BL00790: G66-Y91, N115-T145	BLIMPS_BLOCKS
					Tyrosine specific protein phosphatases signature and profiles Ivc_phosphatase.prf: L2178-R2236	PROFILESCAN
					Fibronectin type III repeat signature PR00014: T1165-P1174, Y1402-D1416	BLIMPS_PRINTS
					Protein tyrosine phosphatase signature PR00700: D2056-I2063, F2072-E2092, M2156-E2173, P2195-L2213, V2226-C2241, M2242-L2252	BLIMPS_PRINTS
					GLOMERULAR MESSAGESIAL CELL RECEPTOR PROTEINTYROSINE PHOSPHATASE PRECURSOR PD178547: C1913-W2026 PD174021: W1528-K1642 PD184208: T145-D205	BLAST_PRODOM
					HYDROLASE PHOSPHATASE PROTEIN TYROSINE PRECURSOR PD000167: N2027-N2221	BLAST_PRODOM
					PROTEIN-TYROSINE-PHOSPHATASE DM000089 A57064896-1170: H1991-I2256 S60613924-1198: H1991-I2256 I40372113-387: H1991-I2256 I4937411-269: L1995-I2256	BLAST_DOMO
					ATP/GTP-binding site motif A (p-loop) A1621-S1628	MOTIFS

WO 02/090530

PCT/US02/01369

Table 3

SFO ID NO:	Incrv Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	(cont.)				Tyrosine specific protein phosphatases active site: V2198-E2210	MOTIFS
3	7482931CD1	478	S173 S237 S263 S278 S283 S353 T118 T155 T189 T269 T300 T402 T431 T456 T463 Y471		Eukaryotic protein kinase domain: W21-S256	HMMER_PPAM
					SIMILAR TO CASEIN KINASES PD115501: F213-E298, L11-T114 PROTEIN KINASE DOMAIN DM00004P4216974-330, I27-E240 DM00004P4873011-265, R25-Y273 DM00004P35506119-273, V23-Y273 DM00004P4023313-267, R25-Y273 Protein kinases ATP-binding region signature I27- K50	BLAST_PRODOM BLAST_DOMO
						MOTIFS

WO 02/090530

PCT/US02/01369

Table 3

SEQ ID NO: Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	2080788CD1 1867	S75 S137 S182 S339 S430 S559 S637 S659 S716 S722 S773 S803 S812 S873 S968 S1047 S1094 S1113 S1120 S1174 S1199 S1232 S1257 S1259 S1280 S1284 S1362 S1387 S1465 S1517 S1527 S1532 S1561 S1630 T84 T95 T96 T102 T113 T538 T636 T933 T971 T1069 T1599 T1747 T1827 T1838 T1839 Y1769	N1585 N1761	DENN (AEX-3) domain: L171-G310	HMMER_PPAM
				PH domain: R1762-S1865 Transmembrane Signatures: A212-S240, L2146-P268, L274-P292, N629-K657, L1394-L1410 N-terminus non-cytosolic	HMMER_PPAM TMAP

WO 02/090530

PCT/US02/01369

Table 3

SEQ ID NO	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	(cont.)				HYDROLASE PROTEIN MYOTUBULARIN DISEASE MUTATION F53A2.8 PROTEIN TYROSINE PHOSPHATASE C19A8.03 CPA2NNF1 P0014611: P1379-H1470, D1129-G1266, G1474- Y1556, G927-F938	BLAST_PRODOM
5	71918969CD1	514	S4, S10, S20, S29, S71, S80, S87, S169, S211, S249, S274, S296, S326, S349, S388, S407, S420, S441, S473, T11, T70, T134, T184, T371, T412, T440, T450, T459, T510	N95, N213, N411, N438	PROTEIN REGULATOR OF PRESYNAPTIC ACTIVITY, SERINE PROTEASE INHIBITOR RAB3 GDP/GTP P0008900: P177-D340, R3-H115_D364_R408 Leucine zipper pattern L239-L260 Eukaryotic protein kinase domain: Y116-L381	BLAST_PRODOM MOTIFS HMMER-PFAM
					Protein Kinases signatures and profile (protein_kinase_1yr.p0): E214_A269 Tyrosine kinase catalytic domain signature: PR00109: M192-D205, Y228-V246, C304-S326, T184-R226	PROFILESCAN
					PROTEIN KINASE DOMAIN: DM000046574721-BLAST-DOMO 266: F118-T371	BLIMPS PRINTS

WO 02/090530

PCT/US02/01369

Table 3

SFO ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	(cont.)				Protein Kinases ATP-binding region signature: L122-K145 Serine/Threonine protein kinases active-site signature: I234-V246 Protein phosphatase 2C: S63-R128, V212-A282	MOTIFS MOTIFS HMMER_PPAM
6	8187571CD1	138	S55 S220 S281 S320 T134 T155 T246 T268 T305 Y248	N178	Transmembrane domain: V89-I117 N-terminus is non-cytosolic. Protein phosphatase 2C proteins BL01032: L64-A81, N109-E148, R205-D218, D253-D265, S320-I329 PROTEIN PHOSPHATASE 2C DM00377P49596 L295: S55-G143, R205-S281 DM00377 Q09173 L296: V66-L141, E188-A282 DM00377 S62462 L297: V66-L141, E188-A282 Eukaryotic protein kinase domain: Y52-P275	TMAP BLIMPS_BLOCKS BLAST_DOMO
7	7494143CD1	321	S19 S24 S228 Y63		Transmembrane domain: D231-L248; N-terminus is cytosolic. Protein Kinases signatures and profile (protein_kinase_379.pf): N149-A197 KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATP-BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN: PD002608: Q203-P231 PROTEIN KINASE DOMAIN: DM00004 Q00536 166-436: L53-L305	HMMER_PPAM TMAP PROFLESCAN BLAST-PRODOM BLAST_DOMO

WO 02/090530

PCT/US02/01369

Table 3

SFO ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	(cont.)				KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATP-BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN: PD002608: Q203-P281 PROTEIN KINASE DOMAIN: DM00004(000536)[66-436: L53-L305 Protein Kinases ATF binding region signature: L58-K81	BLAST-PRODOM BLAST-DOMO MOTIFS
8	38079-34CD1	292	S32 S200 S223 T86 T228	N198	Serine/Threonine protein kinases active-site signature: V169-I181 Transmembrane domain: N135-K163 N-terminus is cytosolic	MOTIFS TMAP

Table 4

Polynucleotide SEQ ID NO./ Instyle ID/ Sequence Length	Sequence Fragments
9/55074884CBI/ 1605	1-361, 1-830, 4-517, 188-830, 346-731, 437-731, 440-493, 440-495, 440-516, 440-517, 440-528, 440-731, 440-1205, 440-1285, 443-731, 450-731, 480-731, 520-774, 526-774, 540-774, 544-731, 544-731, 596-731, 616-731, 645-731, 651-731, 672-731, 683-731, 777-1605, 1026-1191, 1052-1191, 1102-1377, 1337-1440
10/7480588CBI/ 7225	1-148, 1-460, 159-377, 376-894, 646-1584, 1331-2448, 1517-1882, 1700-2082, 1883-1953, 1883-2556, 1883-3153, 2242-2678, 2272-2680, 2719-3351, 2986-3351, 3154-3444, 3171-3283, 3171-3477, 3171-3513, 3171-3522, 3171-3551, 3171-3573, 3171-3604, 3171-3630, 3171-3646, 3171-3680, 3171-3689, 3179-3689, 3209-3689, 3220-3682, 3242-3689, 3244-3689, 3246-3691, 3446-4283, 3459-4284, 4234-4925, 4234-4925, 4235-4920, 4237-4925, 4245-4925, 4271-4925, 4288-5687, 4291-4925, 4297-4925, 4303-4925, 4307-4925, 4324-4925, 4330-4309, 4330-4678, 4330-4685, 4382-4677, 4382-4681, 4382-4682, 4382-4684, 4382-4685, 4384-4684, 4405-4683, 4433-4684, 4467-4680, 5110-5794, 5110-5832, 5110-5834, 5110-5862, 5110-5908, 5110-5916, 5111-5854, 5230-6114, 5444-6104, 5686-6861, 6047-6781, 6047-6784, 6047-6834, 6047-6840, 6047-6854, 6047-6863, 6047-6885, 6047-6885, 6047-6891, 6470-7225, 6495-7225, 6507-7225, 6510-7225, 6541-7225, 6567-7225, 6582-7225, 6738-6864, 6861-6883
11/7482931CBI/ 1691	1-243, 1-576, 32-235, 93-230, 93-378, 186-377, 307-703, 307-763, 307-803, 307-836, 307-845, 307-1049, 307-1054, 307-1086, 307-1093, 307-1076, 307-1078, 307-1080, 307-1112, 307-1127, 308-845, 311-372, 377-787, 377-797, 427-797, 465-1132, 577-1082, 757-1630, 816-1618, 822-1608, 826-1624, 878-1567, 887-1314, 907-1614, 933-1614, 935-1167, 940-1608, 982-1085, 982-1543, 1020-1691, 1021-1691, 1038-1691, 1052-1604, 1053-1431, 1053-1614, 1053-1642, 1053-1667, 1053-1680, 1053-1691, 1054-1691, 1055-1680, 1055-1691, 1063-1691, 1072-1691, 1163-1691
12/2080788CBI/ 6146	1-540, 4-335, 4-580, 11-216, 11-464, 19-359, 19-647, 240-412, 265-595, 271-719, 343-931, 446-1064, 522-1140, 539-821, 617-1206, 661-1241, 664-977, 696-1261, 696-1283, 700-1230, 756-1259, 756-1364, 801-1374, 814-1393, 848-1376, 953-1448, 956-1526, 974-1304, 1005-1465, 1005-1655, 1192-1838, 1206-1743, 1234-1784, 1244-1439, 1258-1819, 1259-1534, 1263-1524, 1271-1543, 1302-1734, 1307-1906, 1319-1672, 1362-1863, 1363-1784, 1368-1749, 1377-1610, 1378-1589, 1378-1640, 1378-1643, 1388-2021, 1415-1689, 1415-1736, 1415-1816, 1415-1986, 1429-1972, 1466-1973, 1475-1756, 1501-1769, 1525-1926, 1527-2112, 1539-1844, 1559-2049, 1584-1782, 1584-2191, 1584-2213, 1586-1828, 1586-2154, 1590-2332, 1603-2222, 1616-2168, 1643-2102, 1671-2212, 1692-2343, 1836-2130, 1837-2140, 1840-2110, 1845-2434, 1846-2188, 1848-2387, 1863-2420, 1878-2204, 1883-2411, 1889-2475, 1890-2426, 1892-2508, 1901-2401, 1906-2401, 1909-2415, 1919-2169, 1919-2436, 1921-2533, 1927-2271, 1945-2307, 1951-2454, 1972-2417, 1987-2515, 2028-2611, 2036-2562, 2036-2641, 2041-2292, 2041-2337, 2043-2658, 2045-2620, 2068-2509, 2100-

Table 4

Polynucleotide SEQ ID NO./ Inset ID/ Sequence Length	Sequence Fragments
137/1918909CBI/ 2362	1-341, 1-450, 1-488, 1-546, 1-590, 1-753, 25-325, 25-443, 30-341, 66-338, 249-938, 251-597, 457-951, 457-969, 457-1091, 520-785, 523-1161, 580-1032, 606-1177, 625-1147, 630-883, 631-1368, 634-1297, 755-1422, 762-1323, 816-1493, 818-1529, 822-1520, 829-1113, 829-1339, 842-1520, 847-1120, 865-1395, 915-1584, 920-1638, 946-1575, 951-1623, 995-1622, 1003-1594, 1021-1735, 1030-1620, 1082-1820, 1086-1769, 1111-1785, 1135-1309, 1145-1309, 1169-1309, 1169-1769, 1192-1620, 1198-1897, 1205-1899, 1209-1309, 1218-1785, 1220-1920, 1222-1460, 1222-1505, 1223-1309, 1224-1309, 1235-1307, 1235-1754, 1243-1831, 1247-1309, 1248-1882, 1250-1309, 1253-1880, 1276-1961, 1277-1908, 1291-1866, 1301-1703, 1311-1929, 1358-1948, 1386-1926, 1454-1969, 1455-1698, 1468-1871, 1477-2276, 1481-1677, 1482-2128, 1498-2010, 1504-2130, 1504-2152, 1511-2169, 1511-2245, 1544-2319, 1557-2310, 1566-2154, 1580-2240, 1581-2323, 1596-2152, 1597-2132, 1602-2323, 1620-2266, 1624-2171, 1632-2277, 1643-2325, 1650-1903, 1660-2151, 1666-2054, 1685-2323, 1686-1906, 1686-1920, 1686-2163, 1686-2266, 1699-1925, 1718-2174, 1722-2171, 1748-2020, 1749-2170, 1756-2338, 1761-2323, 1783-2042, 1804-2168, 1805-2166, 1812-2212, 1827-2362, 1828-2173, 1890-2325, 1900-2325, 1901-2077, 1903-2325, 1920-2166
14818737CBI/ 1535	1-617, 142-828, 175-330, 178-348, 181-459, 239-814, 238-487, 270-795, 311-855, 400-640, 400-983, 557-812, 711-938, 715-988, 783-1079, 803-1371, 810-1509, 854-1431, 886-1528, 1022-1519, 1089-1533, 1124-1535, 1131-1527, 1133-1379, 1145-1373, 1182-1531, 1188-1531, 1215-1336, 1215-1518, 1215-1528, 1215-1535, 1216-1531, 1242-1533, 1254-1528, 1259-1325, 1277-1532, 1294-1526, 1339-1531, 1331-1535, 1343-1531
157494145CBI/ 1376	1-469, 149-469, 153-207, 161-213, 189-521, 208-417, 268-778, 268-812, 268-951, 268-982, 556-1142, 556-1145, 556-1147, 556-1149, 568-1149, 842-1147, 864-1154, 904-1376
165807954CBI/ 1482	1-41, 1-298, 14-485, 34-381, 69-264, 144-642, 191-891, 196-438, 238-482, 259-460, 259-470, 259-473, 271-575, 283-655, 293-381, 293-961, 359-570, 359-639, 378-639, 382-1008, 438-1020, 483-762, 562-891, 635-879, 640-895, 653-1172, 654-941, 654-1003, 691-938, 886-1117, 906-1155, 909-1156, 918-1146, 929-1191, 932-1122, 932-1178, 932-1186, 932-1238, 934-1173, 945-1156, 995-1429, 1007-1203, 1007-1272, 1009-1227, 1009-1236, 1009-1381, 1009-1408, 1009-1447, 1138-1477, 1170-1333, 1257-1482, 1284-1477

WO 02/090530

PCT/US02/01369

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
9	55074884CB1	HEARNON03
10	7480568CB1	LUNGFER04
11	7482931CB1	BRAQNOT01
12	2080788CB1	BRAUNOR01
13	71918969CB1	TESTNOT03
14	8187571CB1	UTRSNOT01
15	7494145CB1	BRSTTMC01
16	5807954CB1	LNODNOT03

WO 02/090530

PCT/US02/01369

Table 6

Library	Vector	Library Description
BRAQNOT01	pFNCY	Library was constructed using RNA isolated from midbrain tissue removed from a 35-year-old Caucasian male. No neuropathology was found. Patient history included dilated cardiomyopathy, congestive heart failure, and an enlarged spleen and liver.
BRAUNOR01	pFNCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81 year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative aortic aneurysm, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolephrosclerosis, nodular colloid goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRSTMC01	pFNCY	This large size-fractionated library was constructed using pooled cDNA from four donors. cDNA was generated using mRNA isolated from diseased breast tissue removed from a 40-year-old Caucasian female (donor A) during a bilateral reduction mammoplasty; from breast tissue removed from a 46-year-old Caucasian female (donor B) during unilateral extended simple mastectomy with breast reconstruction; from breast tissue removed from a 56-year-old Caucasian female (donor C) during unilateral extended simple mastectomy with open breast biopsy; and from breast tissue removed from a 57-year-old Caucasian female (donor D) during a unilateral extended simple mastectomy. Pathology indicated bilateral mild fibrocystic and proliferative changes (A), deep fascia was negative for tumor (B), non-proliferative fibrocystic change (C); and benign fat replaced breast parenchyma (D). Pathology for the matched tumor tissue (B) indicated invasive grade 3 adenocarcinoma, ductal type, with apocrine features. Pathology for the matched tumor tissue (C) indicated invasive grade 3 ductal adenocarcinoma. Pathology for the matched tumor tissue (D) indicated residual microscopic infiltrating grade 3

WO 02/090530

PCT/US02/01369

Table 6

Library	Vector	Library Description
BRSTM01 (cont.)		ductal adenocarcinoma and extensive grade 2 intraductal carcinoma. Patient history included breast hypertrophy and pure hypercholesterolemia (A), breast cancer (B), chronic airway obstruction and emphysema (C), and benign hyperplasia, hyperlipidemia, cardiac dysrhythmia, a benign colon neoplasm, a solitary breast cyst, and a breast neoplasm of uncertain behavior (D). Previous surgeries included open breast biopsy (B). Donor B's medications included Cytoxan and Adriamycin.
HEARN003	pINCY	This normalized heart tissue library was constructed from 8.4 million independent clones from a heart tissue library. Starting RNA was made from heart tissue removed from a 44-year-old Caucasian male, who died from intracranial hemorrhage. Serology was positive for anti-CMV (cytomegalovirus). Patient history included back and neck pain, hypertension, pneumonia, sinus infection, alcohol use, and daily pipe tobacco use (x3 years). Patient medications included Percocet. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Donalio et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LNOOD003	pINCY	Library was constructed using RNA isolated from lymph node tissue obtained from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. On microscopic exam, this tissue was found to be extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, congestive heart failure, atherosclerotic coronary artery disease.
LUNGFER04	PCDNA2.1	This random primed library was constructed using RNA isolated from lung tissue removed from a Caucasian male fetus who died from fetal demise.
TESTNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 27-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
UTRSNOT01	PSFORT1	Library was constructed using RNA isolated from the uterine tissue of a 59-year-old female who died of a myocardial infarction. Patient history included cardiomyopathy, coronary artery disease, previous myocardial infarctions, hypercholesterolemia, hypertension, and arthritis.

WO 02/090530

PCT/US02/01369

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.	
ABUPARACEL FDF	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, ffasta, fastx, tfastx, and ssearch.	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=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E. value=1.0E-8 or less Full Length sequences: fastx scores=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. 266:88-105; and Arwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMIMER	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. 26:370-372; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Scores= 0 or greater

WO 02/090530

PCT/US02/01369

Table 7 (cont.)

Program	Description	Reference	Parameter/Threshold
ProfilesScan	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; Baroch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores/GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Pured	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.	
Purup	A Phil's 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
Consed	A graphical tool for viewing and editing Plurip assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SFSscan	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.	Bainoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

WO 02/090530

PCT/US02/01369

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 SEQ ID NO:1-8,
 - 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:1-8,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and
 - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of 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

WO 02/090530

PCT/US02/01369

- b) recovering the polypeptide so expressed.
10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
- 5
11. An isolated antibody which specifically binds to a polypeptide of claim 1.
12. An isolated polynucleotide selected from the group consisting of:
- 10 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16,
- 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:9-16,
- 15 c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).
13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.
- 20
14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, 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 complex is formed between said probe and said target polynucleotide or fragments thereof, and
- 25 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
- 30
15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.
16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- 35 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.
- 5 17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.
18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
- 10 19. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition of claim 17.
- 15 20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.
- 20 21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.
22. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment a composition of
- 25 claim 21.
23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- 30 b) detecting antagonist activity in the sample.
24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.
- 35 25. A method for treating a disease or condition associated with overexpression of functional

KPP, comprising administering to a patient in need of such treatment a composition of claim 24.

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

- 5
- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
 - 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.

10 27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- 15
- 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.
- 20

28. A method of 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:

- 25
- 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.
- 30

29. A method of assessing toxicity of a test compound, the method comprising:

- 35
- 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 12 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 12 or fragment thereof.
- 5 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.
- 10 30. A diagnostic test for a condition or disease associated with the expression of KPP in a biological sample, the method comprising:
- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
15 b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.
31. The antibody of claim 11, wherein the antibody is:
- 20 a) a chimeric antibody,
b) a single chain antibody,
c) a Fab fragment,
d) a F(ab')₂ fragment, or
e) a humanized antibody.
- 25 32. A composition comprising an antibody of claim 11 and an acceptable excipient.
33. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.
- 30 34. A composition of claim 32, wherein the antibody is labeled.
- 35 35. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:
- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
 - b) isolating antibodies from said animal, and
 - c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
37. A polyclonal antibody produced by a method of claim 36.
38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.
39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:
- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
 - b) isolating antibody producing cells from the animal,
 - 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 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
40. A monoclonal antibody produced by a method of claim 39.
41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.
42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.
43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant

immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in a sample, the method comprising:

- 5
- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
 - b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in the sample.
- 10

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 15 b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 20 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- 25 b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

30

35

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.
50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.
51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.
52. An array of claim 48, which is a microarray.
53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.
54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.
55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.
56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.
57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.
- 5 64. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:9.
65. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:10.
- 10 66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.
67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.
- 15 68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.
69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.
- 20 70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.
- 25 71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

<110> INCYTE GENOMICS, INC.
 LEE, Ernestine A.
 WALIA, Nazinder K.
 BAUGHN, Mariah R.
 ISON, Craig H.
 CURURAJAN, Rajagopal
 ARVIZU, Chandra
 YAO, Monique G.
 HILLMAN, Jennifer L.
 TANG, Y. Tom
 YUE, Henry
 TRAN, Bao
 DING, Li
 LU, Dyung Aina
 LAL, Preeti G.
 WARREN, Bridget A.

<120> KINASES AND PHOSPHATASES

<130> PI-0344 PCT

<140> To Be Assigned
 <141> Herewith

<150> 60/263,083; 60/271,205; 60/271,117; 60/276,859; 60/278,504;
 60/278,522; 60/280,510; 60/280,266
 <151> 2001-01-18; 2001-02-23; 2001-02-23; 2001-03-16; 2001-03-23;
 2001-03-23; 2001-03-29; 2001-03-29

<160> 16
 <170> PERL Program

<210> 1
 <211> 176
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 55074884CD1

<400> 1
 Met Gly Pro Ala Glu Ala Gly Arg Arg Gly Ala Ala Ser Pro Val
 1 5 10
 Pro Pro Pro Leu Val Arg Val Ala Pro Ser Leu Phe Leu Gly Ser
 .20 25 30
 Ala Arg Ala Ala Gly Ala Glu Glu Gln Leu Ala Arg Ala Gly Val
 35 40 45
 Thr Leu Cys Val Asn Val Ser Arg Gln Gln Pro Gly Pro Arg Ala
 50 55 60
 Pro Gly Val Ala Glu Leu Arg Val Pro Val Phe Asp Asp Pro Ala
 65 70 75
 Glu Asp Leu Leu Ala His Leu Glu Pro Thr Cys Ala Ala Met Glu
 80 85 90
 Ala Ala Val Arg Ala Gly Gly Ala Cys Leu Val Tyr Cys Lys Asn
 95 100 105
 Gly Arg Ser Arg Ser Ala Ala Val Cys Thr Ala Tyr Leu Met Arg
 110 115 120
 His Arg Gly Leu Ser Leu Ala Lys Ala Phe Gln Met Val Lys Ser
 125 130 135
 Ala Arg Pro Val Ala Glu Pro Asn Pro Gly Phe Trp Ser Gln Leu
 140 145 150
 Gln Lys Tyr Glu Glu Ala Leu Gln Ala Gln Ser Cys Leu Gln Gly

```

155          160          165
Glu Pro Pro Ala Leu Gly Leu Gly Pro Glu Ala
170          175

<210> 2
<211> 2299
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480588CD1

<400> 2
Met Asp Phe Leu Ile Ile Phe Leu Leu Leu Phe Ile Gly Thr Ser
1      5      10      15
Glu Thr Gln Val Asp Val Ser Asn Val Val Pro Gly Thr Arg Tyr
20     25     30
Asp Ile Thr Ile Ser Ser Ile Ser Thr Thr Tyr Thr Ser Pro Val
35     40     45
Thr Arg Ile Gly Glu Pro Gly Pro Pro Val Phe Leu Ala Gly Glu
50     55     60
Arg Val Gly Ser Ala Gly Ile Leu Leu Ser Trp Asn Thr Pro Pro
65     70     75
Asn Pro Asn Gly Arg Ile Ile Ser Tyr Ile Val Lys Tyr Lys Glu
80     85     90
Val Cys Pro Trp Met Gln Thr Val Tyr Thr Gln Val Arg Ser Lys
95     100    105
Pro Asp Ser Leu Glu Val Leu Leu Thr Asn Leu Asn Pro Gly Thr
110    115    120
Thr Tyr Glu Ile Lys Val Ala Ala Glu Asn Ser Ala Gly Ile Gly
125    130    135
Val Phe Ser Asp Pro Phe Leu Phe Gln Thr Ala Glu Ser Ala Pro
140    145    150
Gly Lys Val Val Asn Leu Thr Val Glu Ala Tyr Asn Ala Ser Ala
155    160    165
Val Lys Leu Ile Trp Tyr Leu Pro Arg Gln Pro Asn Gly Lys Ile
170    175    180
Thr Ser Phe Lys Ile Ser Val Lys His Ala Arg Ser Gly Ile Val
185    190    195
Val Lys Asp Val Ser Ile Arg Val Glu Asp Ile Leu Thr Gly Lys
200    205    210
Leu Pro Glu Cys Asn Glu Asn Ser Glu Ser Phe Leu Trp Ser Thr
215    220    225
Ala Ser Pro Ser Pro Thr Leu Gly Arg Val Thr Pro Pro Ser Arg
230    235    240
Thr Thr His Ser Ser Ser Thr Leu Thr Gln Asn Glu Ile Ser Ser
245    250    255
Val Trp Lys Glu Pro Ile Ser Phe Val Val Thr His Leu Arg Pro
260    265    270
Tyr Thr Thr Tyr Leu Phe Glu Val Ser Ala Ala Thr Thr Glu Ala
275    280    285
Gly Tyr Ile Asp Ser Thr Ile Val Arg Thr Pro Glu Ser Val Pro
290    295    300
Glu Gly Pro Pro Gln Asn Cys Val Thr Gly Asn Ile Thr Gly Lys
305    310    315
Ser Phe Ser Ile Leu Trp Asp Pro Pro Thr Ile Val Thr Gly Lys
320    325    330
Phe Ser Tyr Arg Val Glu Leu Tyr Gly Pro Ser Gly Ala Gly Arg
335    340    345
Ile Leu Asp Asn Ser Thr Lys Asp Leu Lys Phe Ala Phe Thr Asn
350    355    360
Leu Thr Pro Phe Thr Met Tyr Asp Val Tyr Ile Ala Ala Glu Thr

```


Arg Ser Ala Pro Ile Ser Ile Leu Thr Glu Glu Asp Ala Pro Asp
 845 850 855
 Ser Pro Pro Gln Asp Phe Ser Val Lys Gln Leu Ser Gly Val Thr
 860 865 870
 Val Lys Leu Ser Trp Gln Pro Pro Leu Glu Pro Asn Gly Ile Ile
 875 880 885
 Leu Tyr Tyr Thr Val Tyr Val Trp Arg Asn Arg Ser Ser Leu Lys
 890 895 900
 Thr Ile Asn Val Thr Glu Thr Ser Leu Glu Leu Ser Asp Leu Asp
 905 910 915
 Tyr Asn Val Glu Tyr Ser Ala Tyr Val Thr Ala Ser Thr Arg Phe
 920 925 930
 Gly Asp Gly Lys Thr Arg Ser Asn Ile Ile Ser Phe Gln Thr Pro
 935 940 945
 Glu Gly Pro Ser Asp Pro Pro Lys Asp Val Tyr Tyr Ala Asn Leu
 950 955 960
 Ser Ser Ser Ser Ile Ile Leu Phe Trp Thr Pro Pro Ser Lys Pro
 965 970 975
 Asn Gly Ile Ile Gln Tyr Tyr Ser Val Tyr Tyr Arg Asn Thr Ser
 980 985 990
 Gly Thr Phe Met Gln Asn Phe Thr Leu His Glu Val Thr Asn Asp
 995 1000 1005
 Phe Asp Asn Met Thr Val Ser Thr Ile Ile Asp Lys Leu Thr Ile
 1010 1015 1020
 Phe Ser Tyr Tyr Thr Phe Trp Leu Thr Ala Ser Thr Ser Val Gly
 1025 1030 1035
 Asn Gly Asn Lys Ser Ser Asp Ile Ile Glu Val Tyr Thr Asp Gln
 1040 1045 1050
 Asp Val Pro Glu Gly Phe Val Gly Asn Leu Thr Tyr Glu Ser Ile
 1055 1060 1065
 Ser Ser Thr Ala Ile Asn Val Ser Trp Val Pro Pro Ala Gln Pro
 1070 1075 1080
 Asn Gly Leu Val Phe Tyr Tyr Val Ser Leu Ile Leu Gln Gln Thr
 1085 1090 1095
 Pro Arg His Val Arg Pro Pro Leu Val Thr Tyr Glu Arg Ser Ile
 1100 1105 1110
 Tyr Phe Asp Asn Leu Glu Lys Tyr Thr Asp Tyr Ile Leu Lys Ile
 1115 1120 1125
 Thr Pro Ser Thr Glu Lys Gly Phe Ser Asp Thr Tyr Thr Ala Gln
 1130 1135 1140
 Leu Tyr Ile Lys Thr Glu Glu Asp Val Pro Glu Thr Ser Pro Ile
 1145 1150 1155
 Ile Asn Thr Phe Lys Asn Leu Ser Ser Thr Ser Val Leu Leu Ser
 1160 1165 1170
 Trp Asp Pro Pro Val Lys Pro Asn Gly Ala Ile Ile Ser Tyr Asp
 1175 1180 1185
 Leu Thr Leu Gln Gly Pro Asn Glu Asn Tyr Ser Phe Ile Thr Ser
 1190 1195 1200
 Asp Asn Tyr Ile Ile Leu Glu Glu Leu Ser Pro Phe Thr Leu Tyr
 1205 1210 1215
 Ser Phe Phe Ala Ala Ala Arg Thr Arg Lys Gly Leu Gly Pro Ser
 1220 1225 1230
 Ser Ile Leu Phe Phe Tyr Thr Asp Glu Ser Val Pro Leu Ala Pro
 1235 1240 1245
 Pro Gln Asn Leu Thr Leu Ile Asn Cys Thr Ser Asp Phe Val Trp
 1250 1255 1260
 Leu Lys Trp Ser Pro Ser Pro Leu Pro Gly Gly Ile Val Lys Val
 1265 1270 1275
 Tyr Ser Phe Lys Ile His Glu His Glu Thr Asp Thr Ile Tyr Tyr
 1280 1285 1290
 Lys Asn Ile Ser Gly Phe Lys Thr Glu Ala Lys Leu Val Gly Leu
 1295 1300 1305
 Glu Pro Val Ser Thr Tyr Ser Ile Arg Val Ser Ala Phe Thr Lys

1310
 Val Gly Asn Gly Asn Gln Phe Ser Asn Val Val Lys Phe Thr Thr
 1325 1330 1335
 Gln Glu Ser Val Pro Asp Val Val Gln Asn Met Gln Cys Met Ala
 1340 1345 1350
 Thr Ser Trp Gln Ser Val Leu Val Lys Trp Asp Pro Pro Lys Lys
 1355 1360 1365
 Ala Asn Gly Ile Ile Thr Gln Tyr Met Val Thr Val Glu Arg Asn
 1370 1375 1380
 Ser Thr Lys Val Ser Pro Gln Asp His Met Tyr Thr Phe Ile Lys
 1385 1390 1395
 Leu Leu Ala Asn Thr Ser Tyr Val Phe Lys Val Arg Ala Ser Thr
 1400 1405 1410
 Ser Ala Gly Glu Gly Asp Glu Ser Thr Cys His Val Ser Thr Leu
 1415 1420 1425
 Pro Glu Thr Val Pro Ser Val Pro Thr Asn Ile Ala Phe Ser Asp
 1430 1435 1440
 Val Gln Ser Thr Ser Ala Thr Leu Thr Trp Ile Arg Pro Asp Thr
 1445 1450 1455
 Ile Leu Gly Tyr Phe Gln Asn Tyr Lys Ile Thr Thr Gln Leu Arg
 1460 1465 1470
 Ala Gln Lys Cys Lys Glu Trp Glu Ser Glu Glu Cys Val Glu Tyr
 1475 1480 1485
 Gln Lys Ile Gln Tyr Leu Tyr Glu Ala His Leu Thr Glu Glu Thr
 1490 1495 1500
 Val Tyr Gly Leu Lys Lys Phe Arg Trp Tyr Arg Phe Gln Val Ala
 1505 1510 1515
 Ser Ser Thr Asn Ala Gly Tyr Gly Asn Ala Ser Asn Trp Ile Ser
 1520 1525 1530
 Thr Lys Thr Leu Pro Gly Pro Pro Asp Gly Pro Pro Glu Asn Val
 1535 1540 1545
 His Val Val Ala Thr Ser Pro Phe Ser Ile Ser Ile Ser Trp Ser
 1550 1555 1560
 Glu Pro Ala Val Ile Thr Gly Pro Thr Cys Tyr Leu Ile Asp Val
 1565 1570 1575
 Lys Ser Val Asp Asn Asp Glu Phe Asn Ile Ser Phe Ile Lys Ser
 1580 1585 1590
 Asn Glu Glu Asn Lys Thr Ile Glu Ile Lys Asp Leu Glu Ile Phe
 1595 1600 1605
 Thr Arg Tyr Ser Val Val Ile Thr Ala Phe Thr Gly Asn Ile Ser
 1610 1615 1620
 Ala Ala Tyr Val Glu Gly Lys Ser Ser Ala Glu Met Ile Val Thr
 1625 1630 1635
 Thr Leu Glu Ser Ala Pro Lys Asp Pro Pro Asn Asn Met Thr Phe
 1640 1645 1650
 Gln Lys Ile Pro Asp Glu Val Thr Lys Phe Gln Leu Thr Phe Leu
 1655 1660 1665
 Pro Pro Ser Gln Pro Asn Gly Asn Ile Gln Val Tyr Gln Ala Leu
 1670 1675 1680
 Val Tyr Arg Glu Asp Asp Pro Thr Ala Val Gln Ile His Asn Leu
 1685 1690 1695
 Ser Ile Ile Gln Lys Thr Asn Thr Phe Val Ile Ala Met Leu Glu
 1700 1705 1710
 Gly Leu Lys Gly Gly His Thr Tyr Asn Ile Ser Val Tyr Ala Val
 1715 1720 1725
 Asn Ser Ala Gly Ala Gly Pro Lys Val Pro Met Arg Ile Thr Met
 1730 1735 1740
 Asp Ile Lys Ala Pro Ala Arg Pro Lys Thr Lys Pro Thr Pro Ile
 1745 1750 1755
 Tyr Asp Ala Thr Gly Lys Leu Leu Val Thr Ser Thr Thr Ile Thr
 1760 1765 1770
 Ile Arg Met Pro Ile Cys Tyr Tyr Ser Asp Asp His Gly Pro Ile
 1775 1780 1785

Lys Asn Val Gln Val Leu Val Thr Glu Thr Gly Ala Gln His Asp 1800
 1790
 Gly Asn Val Thr Lys Trp Tyr Asp Ala Tyr Phe Asn Lys Ala Arg 1805
 1810
 Pro Tyr Phe Thr Asn Glu Gly Phe Pro Asn Pro Pro Cys Thr Glu 1815
 1820
 Gly Lys Thr Lys Phe Ser Gly Asn Glu Glu Ile Tyr Ile Ile Gly 1825
 1830
 Ala Asp Asn Ala Cys Met Ile Pro Gly Asn Glu Asp Lys Ile Cys 1835
 1840
 Asn Gly Pro Leu Lys Pro Lys Lys Gln Tyr Leu Phe Lys Phe Arg 1845
 1850
 1855
 Ala Thr Asn Ile Met Gly Gln Phe Thr Asp Ser Asp Tyr Ser Asp 1860
 1865
 Pro Val Lys Thr Leu Gly Glu Gly Leu Ser Glu Arg Thr Val Glu 1870
 1875
 Ile Ile Leu Ser Val Thr Leu Cys Ile Leu Ser Ile Ile Leu Leu 1880
 1885
 Gly Thr Ala Ile Phe Ala Phe Ala Arg Ile Arg Gln Lys Gln Lys 1890
 1895
 Glu Gly Gly Thr Tyr Ser Pro Gln Asp Ala Glu Ile Ile Asp Thr 1900
 1905
 Lys Leu Lys Leu Asp Gln Leu Ile Thr Val Ala Asp Leu Glu Leu 1910
 1915
 Lys Asp Glu Arg Leu Thr Arg Leu Leu Ser Tyr Arg Lys Ser Ile 1920
 1925
 Lys Pro Ile Ser Lys Lys Ser Phe Leu Gln His Val Glu Glu Leu 1930
 1935
 Cys Thr Asn Asn Asn Leu Lys Phe Gln Glu Glu Phe Ser Glu Leu 1940
 1945
 Pro Lys Phe Leu Gln Asp Leu Ser Ser Thr Asp Ala Asp Leu Pro 1950
 1955
 Trp Asn Arg Ala Lys Asn Arg Phe Pro Asn Ile Lys Pro Tyr Asn 1960
 1965
 Asn Asn Arg Val Lys Leu Ile Ala Asp Ala Ser Val Pro Gly Ser 1970
 1975
 Asp Tyr Ile Asn Ala Ser Tyr Ile Ser Gly Tyr Leu Cys Pro Asn 1980
 1985
 Glu Phe Ile Ala Thr Gln Gly Pro Leu Pro Gly Thr Val Gly Asp 1990
 1995
 Phe Trp Arg Met Val Trp Glu Thr Arg Ala Lys Thr Leu Val Met 2000
 2005
 Leu Thr Gln Cys Phe Glu Lys Gly Arg Ile Arg Cys His Gln Tyr 2010
 2015
 Trp Pro Glu Asp Asn Lys Pro Val Thr Val Phe Gly Asp Ile Val 2020
 2025
 Ile Thr Lys Leu Met Glu Asp Val Gln Ile Asp Trp Thr Ile Arg 2030
 2035
 Asp Leu Lys Ile Glu Arg His Gly Asp Cys Met Thr Val Arg Gln 2040
 2045
 Cys Asn Phe Thr Ala Trp Pro Glu His Gly Val Pro Glu Asn Ser 2050
 2055
 Ala Pro Leu Ile His Phe Val Lys Leu Val Arg Ala Ser Arg Ala 2060
 2065
 His Asp Thr Thr Pro Met Ile Val His Cys Ser Ala Gly Val Gly 2070
 2075
 Arg Thr Gly Val Phe Ile Ala Leu Asp His Leu Thr Gln His Ile 2080
 2085
 Asn Asp His Asp Phe Val Asp Ile Tyr Gly Leu Val Ala Glu Leu 2090
 2095
 Arg Ser Glu Arg Met Cys Met Val Gln Asn Leu Ala Gln Tyr Ile 2100
 2105
 Phe Leu His Gln Cys Ile Leu Asp Leu Leu Ser Asn Lys Gly Ser 2110
 2115
 2120
 2125
 2130
 2135
 2140
 2145
 2150
 2155
 2160
 2165
 2170
 2175
 2180
 2185
 2190
 2195
 2200
 2205
 2210
 2215
 2220
 2225
 2230
 2235
 2240
 2245
 2250

```

                2255                2260                2265
Asn Gln Pro Ile Cys Phe Val Asn Tyr Ser Ala Leu Gln Lys Met
                2270                2275                2280
Asp Ser Leu Asp Ala Met Glu Gly Asp Val Glu Leu Glu Trp Glu
                2285                2290                2295
Glu Thr Thr Met

<210> 3
<211> 478
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482931CD1

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

```

```

335          340          345
Val Phe Pro Asp Glu Gln Leu Ser Asp Gly Glu Asn Gly Ile Pro
350          355          360
Val Gly Val Ser Pro Asp Lys Leu Pro Gly Ser Leu Gly His Pro
365          370          375
Arg Pro Gln Glu Lys Asp Val Trp Glu Glu Met Asp Ala Asn Lys
380          385          390
Asn Lys Ile Lys Leu Gly Ile Cys Lys Ala Ala Thr Glu Glu Glu
395          400          405
Asn Ser His Gly Gln Ala Asn Gly Leu Leu Asn Ala Pro Ser Leu
410          415          420
Gly Ser Pro Ile Arg Val Arg Ser Glu Ile Thr Gln Pro Asp Arg
425          430          435
Asp Ile Pro Leu Val Arg Lys Leu Arg Ser Ile His Ser Phe Glu
440          445          450
Leu Glu Lys Arg Leu Thr Leu Glu Pro Lys Pro Asp Thr Asp Lys
455          460          465
Phe Leu Glu Thr Trp Tyr Lys Ile Val Tyr Phe Ser Phe
470          475

```

```

<210> 4
<211> 1867
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2080788CD1

```

```

<400> 4
Met Ala Arg Leu Ala Asp Tyr Phe Val Leu Val Ala Phe Gly Pro
1          5          10          15
His Pro Arg Gly Ser Gly Glu Gly Gln Gly Gln Ile Leu Gln Arg
20          25          30
Phe Pro Glu Lys Asp Trp Glu Asp Asn Pro Phe Pro Gln Gly Ile
35          40          45
Glu Leu Phe Cys Gln Pro Ser Gly Trp Gln Leu Cys Pro Glu Arg
50          55          60
Asn Pro Pro Thr Phe Phe Val Ala Val Leu Thr Asp Ile Asn Ser
65          70          75
Glu Arg His Tyr Cys Ala Cys Leu Thr Phe Trp Glu Pro Ala Glu
80          85          90
Pro Ser Gln Glu Thr Thr Arg Val Glu Asp Ala Thr Glu Arg Glu
95          100          105
Glu Glu Gly Asp Glu Gly Gly Gln Thr His Leu Ser Pro Thr Ala
110          115          120
Pro Ala Pro Ser Ala Gln Leu Phe Ala Pro Lys Thr Leu Val Leu
125          130          135
Val Ser Arg Leu Asp His Thr Glu Val Phe Arg Asn Ser Leu Gly
140          145          150
Leu Ile Tyr Ala Ile His Val Glu Gly Leu Asn Val Cys Leu Glu
155          160          165
Asn Val Ile Gly Asn Leu Leu Thr Cys Thr Val Pro Leu Ala Gly
170          175          180
Gly Ser Gln Arg Thr Ile Ser Leu Gly Ala Gly Asp Arg Gln Val
185          190          195
Ile Gln Thr Pro Leu Ala Asp Ser Leu Pro Val Ser Arg Cys Ser
200          205          210
Val Ala Leu Leu Phe Arg Gln Leu Gly Ile Thr Asn Val Leu Ser
215          220          225
Leu Phe Cys Ala Ala Leu Thr Glu His Lys Val Leu Phe Leu Ser
230          235          240
Arg Ser Tyr Gln Arg Leu Ala Asp Ala Cys Arg Gly Leu Leu Ala

```

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

Arg Ser Ala Leu Asp Val Ala Ser Glu Gln Arg Arg Leu Trp Pro
 725 730 735
 Thr Leu Ser Arg Glu Lys Gln Gln Glu Leu Val Gln Lys Glu Glu
 740 745 750
 Ser Thr Val Phe Ser Gln Ala Ile His Tyr Ala Asn Arg Met Ser
 755 760 765
 Tyr Leu Leu Leu Pro Leu Asp Ser Ser Lys Ser Arg Leu Leu Arg
 770 775 780
 Glu Arg Ala Gly Leu Gly Asp Leu Glu Ser Ala Ser Asn Ser Leu
 785 790 795
 Val Thr Asn Ser Met Ala Gly Ser Val Ala Glu Ser Tyr Asp Thr
 800 805 810
 Glu Ser Gly Phe Glu Asp Ala Glu Thr Cys Asp Val Ala Gly Ala
 815 820 825
 Val Val Arg Phe Ile Asn Arg Phe Val Asp Lys Val Cys Thr Glu
 830 835 840
 Ser Gly Val Thr Ser Asp His Leu Lys Gly Leu His Val Met Val
 845 850 855
 Pro Asp Ile Val Gln Met His Ile Glu Thr Leu Glu Ala Val Gln
 860 865 870
 Arg Glu Ser Arg Arg Leu Pro Pro Ile Gln Lys Pro Lys Leu Leu
 875 880 885
 Arg Pro Arg Leu Leu Pro Gly Glu Glu Cys Val Leu Asp Gly Leu
 890 895 900
 Arg Val Tyr Leu Leu Pro Asp Gly Arg Glu Glu Gly Ala Gly Gly
 905 910 915
 Ser Ala Gly Gly Pro Ala Leu Leu Pro Ala Glu Gly Ala Val Phe
 920 925 930
 Leu Thr Thr Tyr Arg Val Ile Phe Thr Gly Met Pro Thr Asp Pro
 935 940 945
 Leu Val Gly Glu Gln Val Val Val Arg Ser Phe Pro Val Ala Ala
 950 955 960
 Leu Thr Lys Glu Lys Arg Ile Ser Val Gln Thr Pro Val Asp Gln
 965 970 975
 Leu Leu Gln Asp Gly Leu Gln Leu Arg Ser Cys Thr Phe Gln Leu
 980 985 990
 Leu Lys Met Ala Phe Asp Glu Glu Val Gly Ser Asp Ser Ala Glu
 995 1000 1005
 Leu Phe Arg Lys Gln Leu His Lys Leu Arg Tyr Pro Pro Asp Ile
 1010 1015 1020
 Arg Ala Thr Phe Ala Phe Thr Leu Gly Ser Ala His Thr Pro Gly
 1025 1030 1035
 Arg Pro Pro Arg Val Thr Lys Asp Lys Gly Pro Ser Leu Arg Thr
 1040 1045 1050
 Leu Ser Arg Asn Leu Val Lys Asn Ala Lys Lys Thr Ile Gly Arg
 1055 1060 1065
 Gln His Val Thr Arg Lys Lys Tyr Asn Pro Pro Ser Trp Glu His
 1070 1075 1080
 Arg Gly Gln Pro Pro Pro Glu Asp Gln Glu Asp Glu Ile Ser Val
 1085 1090 1095
 Ser Glu Glu Leu Glu Pro Ser Thr Leu Thr Pro Ser Ser Ala Leu
 1100 1105 1110
 Lys Pro Ser Asp Arg Met Thr Met Ser Ser Leu Val Glu Arg Ala
 1115 1120 1125
 Cys Cys Arg Asp Tyr Gln Arg Leu Gly Leu Gly Thr Leu Ser Ser
 1130 1135 1140
 Ser Leu Ser Arg Ala Lys Ser Glu Pro Phe Arg Ile Ser Pro Val
 1145 1150 1155
 Asn Arg Met Tyr Ala Ile Cys Arg Ser Tyr Pro Gly Leu Leu Ile
 1160 1165 1170
 Val Pro Gln Ser Val Gln Asp Asn Ala Leu Gln Arg Val Ser Arg
 1175 1180 1185
 Cys Tyr Arg Gln Asn Arg Phe Pro Val Val Cys Trp Arg Ser Gly

1190
 Arg Ser Lys Ala Val Leu Leu Arg Ser Gly Gly Leu His Gly Lys
 1205
 Gly Val Val Gly Leu Phe Lys Ala Gln Asn Ala Pro Ser Pro Gly
 1220
 Gln Ser Gln Ala Asp Ser Ser Ser Leu Glu Gln Glu Lys Tyr Leu
 1235
 Gln Ala Val Val Ser Ser Met Pro Arg Tyr Ala Asp Ala Ser Gly
 1250
 Arg Asn Thr Leu Ser Gly Phe Ser Ser Ala His Met Gly Ser His
 1265
 Gly Lys Trp Gly Ser Val Arg Thr Ser Gly Arg Ser Ser Gly Leu
 1280
 Gly Thr Asp Val Gly Ser Arg Leu Ala Gly Arg Asp Ala Leu Ala
 1295
 Pro Pro Gln Ala Asn Gly Gly Pro Pro Asp Pro Gly Phe Leu Arg
 1310
 Pro Gln Arg Ala Ala Leu Tyr Ile Leu Gly Asp Lys Ala Gln Leu
 1325
 Lys Gly Val Arg Ser Asp Pro Leu Gln Gln Trp Glu Leu Val Pro
 1340
 Ile Glu Val Phe Glu Ala Arg Gln Val Lys Ala Ser Phe Lys Lys
 1355
 Leu Leu Lys Ala Cys Val Pro Gly Cys Pro Ala Ala Glu Pro Ser
 1370
 Pro Ala Ser Phe Leu Arg Ser Leu Glu Asp Ser Glu Trp Leu Ile
 1385
 Gln Ile His Lys Leu Leu Gln Val Ser Val Leu Val Val Glu Leu
 1400
 Leu Asp Ser Gly Ser Ser Val Leu Val Gly Leu Glu Asp Gly Trp
 1415
 Asp Ile Thr Thr Gln Val Val Ser Leu Val Gln Leu Leu Ser Asp
 1430
 Pro Phe Tyr Arg Thr Leu Glu Gly Phe Arg Leu Leu Val Glu Lys
 1445
 Glu Trp Leu Ser Phe Gly His Arg Phe Ser His Arg Gly Ala His
 1460
 Thr Leu Ala Gly Gln Ser Ser Gly Phe Thr Pro Val Phe Leu Gln
 1475
 Phe Leu Asp Cys Val His Gln Val His Leu Gln Phe Pro Met Glu
 1490
 Phe Glu Phe Ser Gln Phe Tyr Leu Lys Phe Leu Gly Tyr His His
 1505
 Val Ser Arg Arg Phe Arg Thr Phe Leu Leu Asp Ser Asp Tyr Glu
 1520
 Arg Ile Glu Leu Gly Leu Leu Tyr Glu Glu Lys Gly Glu Arg Arg
 1535
 Gly Gln Val Pro Cys Arg Ser Val Trp Glu Tyr Val Asp Arg Leu
 1550
 Ser Lys Arg Thr Pro Val Phe His Asn Tyr Met Tyr Ala Pro Glu
 1565
 Asp Ala Glu Val Leu Arg Pro Tyr Ser Asn Val Ser Asn Leu Lys
 1580
 Val Trp Asp Phe Tyr Thr Glu Glu Thr Leu Ala Glu Gly Pro Pro
 1595
 Tyr Asp Trp Glu Leu Ala Gln Gly Pro Pro Glu Pro Pro Glu Glu
 1610
 Glu Arg Ser Asp Gly Gly Ala Pro Gln Ser Arg Arg Arg Val Val
 1625
 Trp Pro Cys Tyr Asp Ser Cys Pro Arg Ala Gln Pro Asp Ala Ile
 1640
 Ser Arg Leu Leu Glu Leu Gln Arg Leu Glu Thr Glu Leu Gly
 1655
 1660

Gln Pro Ala Glu Arg Trp Lys Asp Thr Trp Asp Arg Val Lys Ala
 1670 1675 1680
 Ala Gln Arg Leu Glu Gly Arg Pro Asp Gly Arg Gly Thr Pro Ser
 1685 1690 1695
 Ser Leu Leu Val Ser Thr Ala Pro His His Arg Arg Ser Leu Gly
 1700 1705 1710
 Val Tyr Leu Gln Glu Gly Pro Val Gly Ser Thr Leu Ser Leu Ser
 1715 1720 1725
 Leu Asp Ser Asp Gln Ser Ser Gly Ser Thr Thr Ser Gly Ser Arg
 1730 1735 1740
 Gln Ala Ala Arg Arg Ser Thr Ser Thr Leu Tyr Ser Gln Phe Gln
 1745 1750 1755
 Thr Ala Glu Ser Glu Asn Arg Ser Tyr Glu Gly Thr Leu Tyr Lys
 1760 1765 1770
 Lys Gly Ala Phe Met Lys Pro Trp Lys Ala Arg Trp Phe Val Leu
 1775 1780 1785
 Asp Lys Thr Lys His Gln Leu Arg Tyr Tyr Asp His Arg Val Asp
 1790 1795 1800
 Thr Glu Cys Lys Gly Val Ile Asp Leu Ala Glu Val Glu Ala Val
 1805 1810 1815
 Ala Pro Gly Thr Pro Thr Met Gly Ala Pro Lys Thr Val Asp Glu
 1820 1825 1830
 Lys Ala Phe Phe Asp Val Lys Thr Thr Arg Arg Val Tyr Asn Phe
 1835 1840 1845
 Cys Ala Gln Asp Val Pro Ser Ala Gln Gln Trp Val Asp Arg Ile
 1850 1855 1860
 Gln Ser Cys Leu Ser Asp Ala
 1865

<210> 5
 <211> 514
 <212> PRT
 <213> Homo sapiens

 <220>
 <221> misc_feature
 <223> Incyte ID No: 71918969CD1

<400> 5
 Met Ala Asp Ser Gly Leu Asp Lys Lys Ser Thr Lys Cys Pro Asp
 1 5 10 15
 Cys Ser Ser Ala Ser Gln Lys Asp Val Leu Cys Val Cys Ser Ser
 20 25 30
 Lys Thr Arg Val Pro Pro Val Leu Val Val Glu Met Ser Gln Thr
 35 40 45
 Ser Ser Ile Gly Ser Ala Glu Ser Leu Ile Ser Leu Glu Arg Lys
 50 55 60
 Lys Glu Lys Asn Ile Asn Arg Asp Ile Thr Ser Arg Lys Asp Leu
 65 70 75
 Pro Ser Arg Thr Ser Asn Val Glu Arg Lys Ala Ser Gln Gln Gln
 80 85 90
 Trp Gly Arg Gly Asn Phe Thr Glu Gly Lys Val Pro His Ile Arg
 95 100 105
 Ile Glu Asn Gly Ala Ala Ile Glu Glu Ile Tyr Thr Phe Gly Arg
 110 115 120
 Ile Leu Gly Lys Gly Ser Phe Gly Ile Val Ile Glu Ala Thr Asp
 125 130 135
 Lys Glu Thr Glu Thr Lys Trp Ala Ile Lys Lys Val Asn Lys Glu
 140 145 150
 Lys Ala Gly Ser Ser Ala Val Lys Leu Leu Glu Arg Glu Val Asn
 155 160 165
 Ile Leu Lys Ser Val Lys His Glu His Ile Ile His Leu Glu Gln
 170 175 180

Val Phe Glu Thr Pro Lys Lys Met Tyr Leu Val Met Glu Leu Cys
 185 190 195
 Glu Asp Gly Glu Leu Lys Glu Ile Leu Asp Arg Lys Gly His Phe
 200 205 210
 Ser Glu Asn Glu Thr Arg Trp Ile Ile Gln Ser Leu Ala Ser Ala
 215 220 225
 Ile Ala Tyr Leu His Asn Asn Asp Ile Val His Arg Asp Leu Lys
 230 235 240
 Leu Glu Asn Ile Met Val Lys Ser Ser Leu Ile Asp Asp Asn Asn
 245 250 255
 Glu Ile Asn Leu Asn Ile Lys Val Thr Asp Phe Gly Leu Ala Val
 260 265 270
 Lys Lys Gln Ser Arg Ser Glu Ala Met Leu Gln Ala Thr Cys Gly
 275 280 285
 Thr Pro Ile Tyr Met Ala Pro Glu Val Ile Ser Ala His Asp Tyr
 290 295 300
 Ser Gln Gln Cys Asp Ile Trp Ser Ile Gly Val Val Met Tyr Met
 305 310 315
 Leu Leu Arg Gly Glu Pro Pro Phe Leu Ala Ser Ser Glu Glu Lys
 320 325 330
 Leu Phe Glu Leu Ile Arg Lys Gly Glu Leu His Phe Glu Asn Ala
 335 340 345
 Val Trp Asn Ser Ile Ser Asp Cys Ala Lys Ser Val Leu Lys Gln
 350 355 360
 Leu Met Lys Val Asp Pro Ala His Arg Ile Thr Ala Lys Glu Leu
 365 370 375
 Leu Asp Asn Gln Trp Leu Thr Gly Asn Lys Leu Ser Ser Val Arg
 380 385 390
 Pro Thr Asn Val Leu Glu Met Met Lys Glu Trp Lys Asn Asn Pro
 395 400 405
 Glu Ser Val Glu Glu Asn Thr Thr Glu Glu Lys Asn Lys Pro Ser
 410 415 420
 Thr Glu Glu Lys Leu Lys Ser Tyr Gln Pro Trp Gly Asn Val Pro
 425 430 435
 Asp Ala Asn Tyr Thr Ser Asp Glu Glu Glu Glu Lys Gln Ser Thr
 440 445 450
 Ala Tyr Glu Lys Gln Phe Pro Ala Thr Ser Lys Asp Asn Phe Asp
 455 460 465
 Met Cys Ser Ser Ser Phe Thr Ser Ser Lys Leu Leu Pro Ala Glu
 470 475 480
 Ile Lys Gly Glu Met Glu Lys Thr Pro Val Thr Pro Ser Gln Gly
 485 490 495
 Thr Ala Thr Lys Tyr Pro Ala Lys Ser Gly Ala Leu Ser Arg Thr
 500 505 510
 Lys Lys Lys Leu

<210> 6
 <211> 338
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6187571CD1

<400> 6
 Met Ala Ser Arg Leu Leu His Arg His Ile Arg Glu Gln Leu Lys
 1 5 10 15
 Asp Leu Val Glu Ile Leu Gln Asp Pro Ser Pro Pro Pro Leu Cys
 20 25 30
 Leu Pro Thr Thr Pro Gly Thr Pro Asp Ser Ser Asp Pro Ser His
 35 40 45

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

<210> 7
 <211> 321
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 7494145CD1

<400> 7
 Met Thr Ser Phe His Pro Arg Gly Leu Gln Ala Ala Arg Ala Gln
 1 10 15
 Lys Phe Lys Ser Lys Arg Pro Arg Ser Asn Ser Asp Cys Phe Gln
 20 25 30
 Glu Glu Asp Leu Arg Gln Gly Phe Gln Trp Arg Lys Ser Leu Pro
 35 40 45
 Phe Gly Ala Ala Ser Ser Tyr Leu Asn Leu Glu Lys Leu Gly Glu
 50 55 60
 Gly Ser Tyr Ala Thr Val Tyr Lys Gly Ile Ser Arg Ile Asn Gly
 65 70 75
 Gln Leu Val Ala Leu Lys Val Ile Ser Met Asn Ala Glu Glu Gly
 80 85 90

Val Pro Phe Thr Ala Ile Arg Glu Ala Ser Leu Leu Lys Gly Leu
 95 100 105
 Lys His Ala Asn Ile Val Leu Leu His Asp Ile Ile His Thr Lys
 110 115 120
 Glu Thr Leu Thr Phe Val Phe Glu Tyr Met His Thr Asp Leu Ala
 125 130 135
 Gln Tyr Met Ser Gln His Pro Gly Gly Leu His Pro His Asn Val
 140 145 150
 Arg Leu Phe Met Phe Gln Leu Leu Arg Gly Leu Ala Tyr Ile His
 155 160 165
 His Gln His Val Leu His Arg Asp Leu Lys Pro Gln Asn Leu Leu
 170 175 180
 Ile Ser His Leu Gly Glu Leu Lys Leu Ala Asp Phe Gly Leu Ala
 185 190 195
 Arg Ala Lys Ser Ile Pro Ser Gln Thr Tyr Ser Ser Glu Val Val
 200 205 210
 Thr Leu Trp Tyr Arg Pro Pro Asp Ala Leu Leu Gly Ala Thr Glu
 215 220 225
 Tyr Ser Ser Glu Leu Asp Ile Trp Gly Ala Gly Cys Ile Phe Ile
 230 235 240
 Glu Met Phe Gln Gly Gln Pro Leu Phe Pro Gly Val Ser Asn Ile
 245 250 255
 Leu Glu Gln Leu Glu Lys Ile Trp Glu Val Leu Gly Val Pro Thr
 260 265 270
 Glu Asp Thr Trp Pro Gly Val Ser Lys Leu Pro Asn Tyr Asn Pro
 275 280 285
 Glu Glu Ser Leu Phe Thr Val Ser Gly Val Arg Leu Lys Pro Glu
 290 295 300
 Met Cys Asp Leu Leu Ala Ser Tyr Gln Lys Gly His His Pro Ala
 305 310 315
 Gln Phe Ser Lys Cys Trp
 320

<210> 8
 <211> 292
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5807954CD1

<400> 8
 Met Lys Ala Thr Val Leu Met Arg Gln Pro Gly Arg Val Gln Glu
 1 5 10 15
 Ile Val Gly Ala Leu Arg Lys Gly Gly Asp Arg Leu Gln Val
 20 25 30
 Ile Ser Asp Phe Asp Met Thr Leu Ser Arg Phe Ala Tyr Asn Gly
 35 40 45
 Lys Arg Cys Pro Ser Ser Tyr Asn Ile Leu Asp Asn Ser Lys Ile
 50 55 60
 Ile Ser Glu Glu Cys Arg Lys Glu Leu Thr Ala Leu Leu His His
 65 70 75
 Tyr Tyr Pro Ile Glu Ile Asp Pro His Arg Thr Val Lys Glu Lys
 80 85 90
 Leu Pro His Met Val Glu Trp Trp Thr Lys Ala His Asn Leu Leu
 95 100 105
 Cys Gln Gln Lys Ile Gln Lys Phe Gln Ile Ala Gln Val Val Arg
 110 115 120
 Glu Ser Asn Ala Met Leu Arg Glu Gly Tyr Lys Thr Phe Phe Asn
 125 130 135
 Thr Leu Tyr His Asn Asn Ile Pro Leu Phe Ile Phe Ser Ala Gly
 140 145 150

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

<210> 9
 <211> 1605
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 55074884C31

<400> 9
 ccaagcctcca gcoctggctcca cctcggaggc ctctaggacc cggggggccc cggcggcggc 60
 ccggctccca caaatagact cctgggcggg cgcctgagcc cccaaaatag atcctcaagg 120
 cccaaaagca gactcttcgg cgggggccaat gggaccgcca gaagctgggc gccgcggggc 180
 cgcctcggcc gtacctccac cgttggtgcg cgtcgcggcc tcaactctcc cgggagcgc 240
 gccagccgag ggcggggagg agcagctggc ggcgcgggga gtcacgctgt cgtcaaacgt 300
 ctcccgccag cagccgggccc cgcgcggccc cggcgtggca gagctgcgcg tgcgccgttt 360
 cgacgacccc gctgaggacc tctggcgcca cctggagccc acgtgcgccc ccatggaagg 420
 cgcggtgcgc gccggggggc cctgcctagt ctaactgcaag aacggccgca gccgctcggc 480
 cgcgctctgc accggttacc tcatggcgca ccggggcctc agcctggcga aggccttcca 540
 gatggtgaag agcgcctccc cgttagcaga accgaaccgg ggcctctggc ctcaactcca 600
 gaagtatgag gaggcctccc aggccagtc ctgcctgcag ggagagccccc cagccttagg 660
 gttggccctt gagcttgaag gcttgaaggc ctgctgctct gaggaaggat gtccttgcac 720
 tgatacagaa ggcgtggtctt tacccttctt cctcactgta atategagtt ttcttttgg 780
 tgtgtgtgtg tgaacatag tgcttttaat tttatatttc cgcctgaggt gatgcacaaa 840
 ttatcttaca gacacaaaaa gaatacatt tggtaaaaaa tcgaagacaa attaagaaaa 900
 agcaaatagg gatctccat tattttacat ccagaatctt gactgtttgg ggggtccggc 960
 gcattctacc agtctcaggg aaggacatga gtacagacta cagtaatcag aactcagtc 1020
 ataggtgaaa tcaggaaaag caactcttct tgtgtttatt tttagtatg tcataagaac 1080
 ggaactagtt ttctcgcgtg accacggatg ctctctgttg agaaaaatgca taccacgctg 1140
 ggaacttag atcaagaagg ctgttgatga agacattggt gaagggcaac ttgggtgatt 1200
 ggggaattat tctcttcttc agcccctctg caatacagct ggaagctgtcc cataggagtt 1260
 cggacaatta cggacatccc ttctcttctt ctctttcttt ttctttgctt gtttatttga 1320
 gacagagctc ctgttcgcca ggtcggagtg cagtggcgca atcttggctc actgaaacct 1380
 ccgcttcca ggttcaagtg attctctgca ctaagctccc tgagtagctg ggaactaacg 1440
 tgttgacac cagaccgggc taatttcttt agtattttta gtacagatgg gggggttcca 1500
 accatgttac ccaggatgta tcttactcct gacctcagag atccagccac ctccatcttc 1560
 aacaatcgtg atgtgaggaa aacagagtgga cgaacattcg ggtac 1605

<210> 10
 <211> 7225
 <212> DNA
 <213> Homo sapiens

```

<220>
<221> misc_feature
<223> Incyte ID No: 7480588CB1

<400> 10
atggattttc ttatcatttt tcttttactt tttattggga cttcagagac acaggttgat 60
gtttccaatg tegtccctgg tactaggtac gatataacca tctcttcaat ttctacaaca 120
taacctctac ctgttactag aataggggaa ccagggctcc cagttctctc agccggggaa 180
agagtcggat ctgctgggat tcttctgtct tggaaacac cacctaacc aaatggaag 240
attatctctt acattgtcaa atataaggaa gtttgcctg ggatgcaaac agtataata 300
caagtcagat caaagccaga cagtctgaaa gttctcttta ctaactttaa tctggaaca 360
acatatgaaa ttaaggttgc tgtgaaaac agtgctggca ttggagtggt tagtgatcca 420
tttctctcc aaactgcaga aagtgtcca ggaaaagtg tgaatctcac agttgagcc 480
tacaacgctt cagcagttaa gctgatttgg tatttaoctt ggcaacaaa tggcaaaatt 540
accagcttca agattagttc caagcatgcc aagtggggaa tagtagtgaa agatgtctca 600
atcaagtgag agaacattt gactgggaaa ttgccagaa tcaatgagaa tagtgattct 660
tttttatgga gtaacagccag cctctctcca acccttggtt gagttaacc tccatgctgt 720
accacaactt cacaagcac gttgacacag aatgagatca gctctgtgtg gaaagagct 780
atcagttttg tagtgacaca cttgagacct tatcaaat atctttttga agtttoagct 840
ctcaacactg aagcaggtta tattgatagt acgattgtca gaacaccaga atcagtgctc 900
gaagagccac cacaaaactg cgtaacaggg aacatcacag gaaagtcttt tcaatttta 960
tgggaccacc caactatagt aacagggaaa tttagttata gagtgtgaatt atatggacca 1020
tcagtgagag gtgccttttt ggataacagc acaaaagacc tcaagtttgc atcactaac 1080
ctacaacact ttacaatgta tgatgtctat attgcccctg aaaccagtgc agggactggg 1140
cccaagtcac atatttcaat atcactcca ccagatgttc caggggagct gtttgattta 1200
caacttgcag aggtagaatc cacgcaagta agaattactt ggaagaiaacc acgacaacca 1260
aatggaatta ttaaccaata ccagtgaaa gtgctagttc ccagagacgg aataattttg 1320
gaaatattt tgcctactgg aatataagag atcaatgacc ccatggctcc agaatgttg 1380
aacatagtae agccaatgtt aggatattat gagggttcag ccagatgttc gctgacctt 1440
cactcaactg ctacatttat atataacagc catccagata aaaactttcc tgcaggaaat 1500
agagctgaag accagacttc accagttgta actacaagga atcagtatat tactgacatt 1560
gcagctgaac agctgtctta lgttatcagg agacttgtac ctttactgta gcacatgatt 1620
agtgatctg ctttaccat catgggagaa ggaccacca oagttctcag tgttaggaca 1680
cgtcagcagc tgcacagctc cattaaaaat ataaactata aaaatattag ttctctcat 1740
attttgttat attgggatcc tccagaatat cccaatggaa aaataactca ctatacgatt 1800
tatgcaatag aattggatac aaacagagca ttccagataa ctaccataga taacagctt 1860
ctataaacag gtataaaaaca atggttaaag aaatacacaa aatacaaaat gagagtggca 1920
gcctcaacc ccgttggaga aagttctttg tctgaagaaa atgacatctt tgtgagaact 1980
tcagaagatg aaccggaaac atcacctcaa gatgtogaag taattgatgt tacocagat 2040
gaaaatagtt tgaagctgct accaccgaaa aagcccaatg ggatcattat tgcctatgaa 2100
gtgctatata aaaaataga tactttatat atgagagaca catcaaacac agacataata 2160
ttaaggaaat taagacctca accctctat aacatctctg taaggcttta caccagattt 2220
ggtcatggca atcaggtate ttcttbaact tctgtaagga ettcggagac tgtgctgat 2280
agtgcaccag aaaaatcac ttacaaaaat atttctctg gagagattga gctatcatt 2340
cttcccccaa gtgtcccaa tggaaatata caaaaatata caatttatct caagagaagt 2400
aatggaatg aggaaagaa tataaatata acctctttaa ccaaaaacat taaggagctg 2460
aagaaataa ccaatatac cattgaagtg tctgttagta caactcaagg tgaaggagt 2520
cggagtgtcc ccaatagtat actgacggag gaagatgctc ctgattctcc cctcaagac 2580
ttctctgtaa aacagtttgc tgggtctcag gtgaagttgt catggcaacc accctggag 2640
ccaaatggga ttatccttta ttacacagtt tatgtctgga ggaatagatc atcattaaa 2700
actattaatg tcaactgaac atcattggag ttatcagatt tggattataa tgttgaatc 2760
agtgcttatg taacagctag caccagattt ggtgatggga aaacaagaag caalateatt 2820
agcttcaaaa caccagggg accaagcagat cctcccaag atgtttatta tgcacaacct 2880
agttctcat caataattct ttctgggaca cctctcaaa aacctaatgg gatataaca 2940
tattactctg ttattacag aaatacttca ggtactttaa tgcagaattt taactctcat 3000
gaagtaccca atgactttga caatagact gtatccaaa ttatagataa actgacaata 3060
ttcagctact atacattttg gttaacgca agtactctag ttggaaatgg gaataaaaag 3120
agtgacata ttgaagata cacagatcaa gcgtactctg aaggtttgtg tggaaacctg 3180
acttaegaat ccattctgct aactgcaata aatgtaagct gggctccacc ggctcaacca 3240
aacggtctag tcttctacta tgtttcaact atcttaacag agactcctcg ccatgtgaga 3300
ccactctctg t.acatataga gagaagcata tattttgata atctggaaaa atactctgat 3360
tataatataa aaattactcc atcaacagaa aagggattct ctgataccta tactgcccag 3420
ctatacata agactgaaga agatgtccca gaaacttcaac caataatcaa cacttttaaa 3480

```

```

aacctttcct ctacctcagt tctcttatca tgggatcccc cagttaaagcc aaatgggtgca 3540
ataataagtt atgatttaac ttacaaga ccaaatgaaa attattcttt cactactctc 3550
gataattaca taatattgga agagcttca ccatttaaat tatatagctt ttttgetgoc 3560
gcaagaacta gaaaagact tggctcttcc agtattcttt tcttttccac agatgagtca 3720
gtgocgltag caccctccaca aaatttgact ttaatcaact gtaactcaga ctttgtatgg 3780
ctgaaatgga gcccaagctc tottccaggt ggtattgtta aagtatatag ttttaaaatt 3840
catgaacatg aaactgcacac tatatattat aagaatata caggatttaa aactgaagcc 3900
aaacttggtg gaactggaacc agtcagcacc tactctatcc gtgtatctgc gttcaaccaa 3960
gttggaatg gcaatcaatt tagtaatgta gtaaaattca caaccaaga atcagttcca 4020
gatgctcgtg agaattatgca gtgcatggca actagctggc agtcagtttt agtgaatgg 4080
gatccacca aaaagggaaa tggaaataa acgcagtata tggtaacagt tgaagggaa 4140
tctacaaaag tttctcccca agatccatgt tacactllca taaagcttct tgccaatacc 4200
tcataatgct ttaaagtaag agcttcaacc tcagctgggt aaggtgatga aagcaatgc 4260
catgcaagca cactacctga aacagttccc agtgttccca caaatattgc tttttctgat 4320
gtccagttcaa ctagtccaac atgacatgg ataagacctg acactatcct tggctacttt 4380
caaaattaca aaattaccac tcaacttctg gctcaaaat gcamaagatt ggaatcggaa 4440
gaatgtggtg aatatacaaa aattcaatcc tcttatgaag ctcaactaac tgaagagaca 4500
gtatatggat taaggaatt tagatgggat agattccaag tggcttccag caccaatgct 4560
ggctatggca atgtctcaaa ctggtattct acaaaaactc tgcctggccc tccagatggt 4620
ccctctgaaa atgttcatgt agtagcaaca tcacctllta gcctcagcat aagctggagt 4680
gaactcgtct tcattactgg acccaactgt tactctgatt atgtcaaatc ggtagataat 4740
gtgaaatga atatatcctt catcaagtca aatgaagaaa ataaaacct agaaattaaa 4800
gatttagaaa tattccaaag gtattctgta gtgatcactg cacttaactgg gaacattagt 4860
gctgcaatg tagaagggaa gtcaagtgt gaaatgatt ttactacttt agaalagccc 4920
ccaaagacc cccctaacaa catgacattt cagaagatc cagatgaagt tacaanaatt 4980
caatacaggt tccctctccc ttctcaacct aatggaaata tccaagtata tcaagctctg 5040
gtttaccag agatgatcc tactgctgtc cagattccca accctcagcat tatcacagaa 5100
accacactat tctctctgc aatgctagaa ggactaaaag gtggcatatc atacaatc 5160
aggttttag cagtcaatag tctgtgtgca ggtccaaagg ttccgatgag aataaocagt 5220
gatatacaag ctccagcagc accaaaacc aaaccaacc ctatttatga tgcacacaga 5280
aaactgctgt tgaactcaac aacaattaca atcagaatgc caatagtta ctacatgat 5340
gatcatggac caataaaaaa tgcacaagt ctgtgacag aaacaggagc tcagcatgat 5400
ggaaatgtaa caaagtggta tgcacatata ttaataaag caaggccata ttttcaaat 5460
gaaggtttc ctAACCCctc atgtacagaa ggaagacaa agtlttagtg caatgaagaa 5520
atctacatca taggtgctga taatgcatgc atgattcctg gcaatgaaga caaaattgc 5580
aatggaccac tgaacccaaa aaagcaatc ttatttaaat ttagagctac aatattatg 5640
ggcaaattta ctgactctga ttattctgac cctgttaaga ctttagggga aggaacttca 5700
gaaagaaccg tagagatcat tctttcctgc actttgtata tcccttcaat aattctcctt 5760
ggaaacagta ttttlgcat tgcgagaatt cgcacagaagc agaaagaagg tggcacatcc 5820
tctctcaggt atgcagaat tatgacactt aaatgaaagc tggatcagct catcacagtg 5880
gcagactgg aactgaagga cgaagatata agcggttac ttggtatag aatctcactc 5940
aagconataa gcaagaatcc ctctctgcaa catgttgaag agctttgcac aaacaacac 6000
ctaaagttc aagaagaatt ttccgaaata ccaaaatttc ttccagatct tcttcaact 6060
gtgtctgatc tgccttggaa tagagcaaaa aaccgcttc caaacataaa accatataat 6120
aaatacacag taaagctgat agctgacgct agtgttccag gttccgatta tattaatgcc 6180
agctatctct ctggttatt atgtccaaat gaatttatg ctactcaagg tccactacca 6240
ggaaacagtt gagtttttg gagaatggt tgggaaacca gagcaaaaac attagtaatg 6300
ctaacacagt gttttgaaa aggcgggac agatgccatc agtattggcc agaggaacac 6360
aagccagtta ctgtctttgg agatatagt attacaagc taatggagga tgttcaaat 6420
gattggacta tcagggatct gaaaattgaa aggcattggg attgcatgac tgttcgacag 6480
tgaacttta ctgctggccc agagcatggg gttcttgaga acagcgcacc tctaactcac 6540
tttgtgaagt tggttcgagc aagcagggca catgacacca cacctatgat tgttcaactgc 6600
agtctggag ttggagaaac tggagttttt atgtcctgg accaattaac acaacataat 6660
aatgaccagt atttttgga tatatattga ctagtactg aactgagag tgaagaaatg 6720
tgcaggtgc agaattggc acagtatct tttttaccc agtgactct ggatctctta 6780
tcaataaagg gaagtaatc gccactctgt tttgttaact atcagcact tcaagaatg 6840
gactctttgg accccatgga aggtgatgtt gagcttgaat ggaagaaac cactatgtaa 6900
atatccagac caaaggatc aatggaaaga gatttttaa tcccagggcc caaagtacc 6960
ccctcaattc tccgaattga aatgtgcaac cttaagaaa tatctatgct tetctcactg 7020
tgctttcca acgattgaa catlltaaga ctagtctgt aaaaatgcta atacagaata 7080
attattgttt ttgtacagaa taatatattat gcattttaaa tgccttaagaa aagacatccc 7140
atagttttt gaagctctcc atattttgga ataagccaaa tagaaaatta ttattatatt 7200
agcattaatg tttcaatgag aaaaa 7225

```

```

<210> 11
<211> 1691
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7482931CB1

<400> 11
cggtcccgc tgccgagggt taacccccgc cggccccggt cctgagctgg accagagccc 60
tcctccagaa acccctgcgt ccgccacggc ccaggttaaa tggaaaccac ccttgggaaac 120
tggatgcctg tgtagctggt ctaccatatic agtgattgc aatgagtggt ggaggagagc 180
agctggatat cctgagctgt ggaatccctag tgaagaaga atggaaagtg ttgagaaga 240
ttgggggtgg gggctttgga gaaatttacg atgccttggg catgctcacc agggaaaaatg 300
ttgcactgaa ggtggaatca gctcaacaac caaaacaagt tctgaaaatg gaagttgctg 360
ttttgaaaaa gctgcaaggg aaagaccatg tttgtagatt tattggctgt ggaggaaatg 420
atgalttaa ctatgtggtc atgcagttgc aggttcgaaa tctggcagat ctctccgta 480
ccagctcccg aggcacattc accatagta ccactctccg cctgggtaga cagattttgg 540
agcttattga aagcattcat tctgtgggat tcttgcatcg agacatcaa ccgtcgaact 600
tcgctatggg tegtcttctc agtacatgta ggaatgtta catgcttgat tttgctttgg 660
ctgcacaatt taccaatctc tglggtgacg tcagaccacc tcgagctgtg gcagtttttc 720
gaggagacgt tcyttatgca tcaatcaacg cacatcggaa cagggaaatg ggaagacatg 780
atgacctttg gtccttattc taacatgttg tggagtttgg ggttggctag ctgcccgtga 840
aaaaataaaa ggacaaaggg caagtaggct ctattaagga gagatagac cacaggtca 900
tgttgaaaaa tctcccctca gaattcagca tctttctaga ccatatctct tctttggatt 960
atlttcaaaa accagactac cagcttctta catccgtggt tgacaalago atcaagaact 1020
ttggagtaac tgagagtgac ccttttgact gggagaagac tggaaatgat ggctcccata 1080
caaccacac tacttctaac accctcagt tgcacactcg cttgacccct gctgcaattg 1140
gaattgcaaa tgcctactcc atccctggag acttgcttcc agaaaataca gatgaggtat 1200
ttccagatga acagcttagc gatggagaaa atggcaccct tgttggctgt tcaccagata 1260
aaltgcctgg atctctggga caccocctgc ccagaggaaa ggttggtttg gaagagatgg 1320
atgccaacaa aaacaagata aagcttggaa tttgtaaggc tgcactgaa gaggagaaca 1380
gcatgcccga ggcaaatggt ctctcctaat ctccaagcct tgggtcacca attcgtgtcc 1440
gctcagagat tactcagcca gacagagata tccaactggt gcgaaagtta cgttccattc 1500
acagctttga gctggaaaaa cgtctgacc tggagctaaa gccagacact gacaagtctc 1560
ttgagacctg gtataaaaaa gtgtattttt ctttttaagg ctctcaaggt accattatta 1620
ttgtgtcat tgtttgtatt attattgtat atttctgtta cataaagctt tcaaaaatg 1680
ccaaaaaaa a 1691

<210> 12
<211> 6146
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2080738CB1

<400> 12
ccgctcgctt gggaaagatgg oggggggagc ctggggccgc gccgcccggc ccgcccgcgc 60
ggagcgaacc aggggtgtcc ggggtgcgcy gccaaggccc ggggcccggc catgaagcgy 120
ccgtctctga gtcccagagc ccgagagccc gcccgcccct ctggggccgc cccgctccc 180
tcgcaatgga gggctctgcy gactacttgc tctgttggc gttcggcccg caaccgccc 240
ggagtgggga aggccagggc cagattctgc acgcttccc agagaaggac tgggagaca 300
accatctccc ccaggccatc gactgtttt gccagcccag cgggtggcag ctgtgtccc 360
agaggaatcc accgacttcc tttgtgtctg tctcaccga catcaactcc gagcgcact 420
actgocctcg cttagacttc tgggagccag cggagcttc acaggaaac acgocctgy 480
agatgcccac agagagggag gaagaggggg atgagggagg ccagaccacc ctgtctccca 540
cagcactctc cccatctgcc cagctgtttg caccgaagac gctggtactg gtgtcggac 600
tcgaccacac ggaggtgttc aggaacagcc ttggcctcat cactgccatc cagctggagg 660
gcctgaatgt gtcctggag acctgtattg ggaacctgct gactgcatc gtcgccctg 720
ctggggcctc gcagaggagc atctctttgg gggctggtga ccggcaggtc atccagactc 780

```

```

cactggccga ctegetgcc gtcageccgt gcagcgtggc cctgctcttc cggcagctag 840
gcataaccnaa cgtgctgtet ttgttctgtg cegccctcac gtagcacaag gttctcttcc 900
tgtcccggag ctaccagcgg ccccccagat cctgtagggg cctccgggpa ctgctgttcc 960
ctctcagata cagcttcacc tatgtgccca tccctgcggc tccagctgctc gaagctctca 1020
gtacaccac gcccttcacc attgaggta acggggcttt ccaggcagag acccaggaga 1080
tgetgatgt gattgtgct gatctgagat gaggggcgtt caccatctct gagtgtgtc 1140
acattcoacc cttgccagag coactgcaga gtcagacgca cagtgtgctg agcatggctc 1200
tggaccocgga gctggagttg gctgacccct ccttccctcc gcccaagaca tccacctcct 1260
ccctgaagat gcaggacaag gagctgcggc cgttcttctc cgggctgttc gctcagctgc 1320
tgcagggta lcgctgtgtc ctgcacgtcg tgcgcatcca cccggagcct gtcacccgt 1380
tccaatagcg agccttctct gcccagcgtg ggtggttaga ggcagatlc ctgatgaag 1440
tctgaggagg caatggccttt gctggctttg tctcagagcg tggggctcca taccgccct 1500
cggaccctgt cgatgagctg gtcggccacg aggtggcaag gatggggcg gatgagaacc 1560
accccacgag tgcctgcgt cacgtccagg aactggcaga gcagctctac aagaacgaga 1620
acccttacc agccgtggcg atgcacaagg tacagaggcc cggtagagac agccctctc 1680
gacgggtgcc ccgacccttc ccccggctgg atgagggcac cgtgcagtgg atcgtggacc 1740
agcctgagc caagatgag gctgcacccc cagctctgaa ggcggaggg aggaccacg 1800
tgcctcaag gcccccagtg actgcatac tggagcgttg cagtggctc gatgtaaca 1860
gcccaccggc gctggaggtt gtggcaact gcatctcta cgtgtttgag gggaaaatgc 1920
ttgagccaa gaagotgct ccagccgtgt tggggccctc gaagggcgca gctcccgc 1980
gctgctcgc ccagagcgtc cactgcatg tgcagcaga cctgcccgc ctggaccacc 2040
agocgttga ctttctgtc cgtatgata actgctgct gcaggactgc acttctctg 2100
accagcatcg cattgcggcg gctctgctc ctctgctcac agccttctgc cggagctga 2160
cccggggtg gacgcagttt gcatacagct gctgtagcaga gcaagtggt tggagcagc 2220
cacagtctg gaggccatg ttctatgggg atgtgcagac tcaactccgg gccctctac 2280
tggagccac ggaggactg gcccccgc ccaggagttg ggaggcact tcccaggag 2340
acgagcctc tgcctagac ttggctctg agcagcggcg cttgtggcca actctgagt 2400
gtgagaagca gcaggagctg tgcagaagg aggagagcac ggtgttcagc caggccatcc 2460
actatgcca ccgatgagc tccctctcc tgcctctgga cagcagcaag agccgctac 2520
ttcggagcg tgcgggctg gctgacctg agagcgcag caaacactc gtaaccaaca 2580
gcatsctgg cagtgtgccc gagagctatg acacggagag cggctcgag gatgcagaga 2640
cctgcagact agtggggct gtggctccct tcatcaacc ctttctggac aagctctga 2700
cggagagtg gttcaccagc gaccacctca agggctgca tgtcatggt ccagacatg 2760
tcoagatgca catcgagacc ctggaggccc tgcagcggga gacccggag ctccgccca 2820
tccagaagcc caagctgct cggccgccc tctgcccgg tggagagtg gctgctgag 2880
gctctgctg ctacctgct cgggatggc gtgaggaggg cgggggggg agtctgggg 2940
gaccagcatt gctcccagct gaggcgccc tcttctcac caactccgg gtcacttca 3000
cgggatgccc caaggcccc ctggttgggg agcagtggt ggtccgctc tcccgggtg 3060
ctgctctca caaggagaag cgcatacag tccagacccc tctggaccag ctctgcagg 3120
acggctcca gctgctccc tgcacatccc agctgctgaa aatggcctt gacggagag 3180
tgggtctga cagggcag ctctctccta agcagctgca laagctggg laccggcgg 3240
actcaaggc caccctggc tccacttgg gctctgcca caaacctgg aggcacccg 3300
gagtaacca ggacaaggct ccttccctca gaaacctgtc ccggaacctg gtaagaacg 3360
ccaagaagc catcggggcg cagcatgca ctgcagaaga gtaaacccc cccagctgg 3420
agcaccgggg ccagcccgc cctgaggacc aggagagca gatctcagt tggaggagc 3480
tggagccccc caccctgacc cctctctcag cctcgaacc ctccgaccg atgaccatga 3540
gcagctggt ggaagggtc tctgtctcag actaccagc cctcgtctg ggcaccctga 3600
gcagcagct ggcggggcc aagtctgagc cctccgcat tctccgctc aaccgcatg 3660
atgcatctg ccgacagctc ccaggctgc tgatctgccc caagagttc caggacaacg 3720
ccctcagcg cgtgtcccgc tctaccgcc agaaccgctt cccctggtc tctgggcca 3780
gagggcgtc caaggcggg ctgctgcgct ctggaggcct gcatggcaaa ggtgtgctg 3840
gctcttcaa ggcccagaac gcaactctc caggccagt ccaggcggac tggatagcc 3900
tggagcagga gaagtacct caggctgtgg tcaactccat gccccgctac gccagcgtc 3960
cgggaccca cagcttagc ggtctctcc cagccacat ggcagctcac ggtcaagtgg 4020
gcaagtctcg gaccagtgga cgcagcagtg gctctgacc cgaagtgggc tcccgtgag 4080
ctggcagaga cgcctgctgc ccaccacag ccaacgggg cctcccgac cgggcttcc 4140
tgcctccga cggagcagcc ctctatctc ttggggacaa agcccagctc aaggtgtgc 4200
ggtcagacc cctgcagcag tgggagctg tgcacctga ggtattcgg gacggcagg 4260
tgaaggctag ctcaagaag ctgctgaaag catgtctcc aggtgcccc gctgctgagc 4320
ccagcccagc ctccctctg cgtcaactg aggaactcaga gctgctgatc cagatccaca 4380
agctgctga ggtgtctgt ctgggtgtg agctcctgga ttcaggctc tccgtgctg 4440
tgggctgga ggaatggctg gacataacca cccagtggt atcctgtgt cagctgctc 4500
cagaccctt ctaccgacg ctggaggctc ttgcctgct ggtggagaag gagtgtgct 4560

```

```

cctteggcca teqctteagc caccgtggag ctcacacct ggccggggcag agcagcggct 4620
tcacaaccgt ctccctgcag ttccctggact gegtacacca ggtccacetg cagttcccca 4680
tggagtttga gttcagccag ttctaactca agttcctcgg ctaccaccat gtttcccgcc 4740
gtttccggac ctctctctct gactctgact atgagcgaat tgagctgggg ctgtgtatg 4800
aggaagaagg gaaacgcagg ggccaagtgc cgtgcaggtc tttatgggag tatgtggacc 4860
ggctzagcaa gaggacgct gtgttccaca attacatgta tgcaccagac gacgcagagg 4920
tctctcgccc ctacagaac gtgtccaacc tgaagtggtt ggaactctac actgaggaga 4980
cgtctggccg gggccctccc tatgactggg aactggccca gggcccccct gaacccccag 5040
aggaagaacg gttctgagga ggcgctcccc agagcagcgg ccgctgtgtg tggcctctgt 5100
acgacagctg cccgctggcc cagcctgacg ccatctcaag cctgctggag gagctgcaga 5160
ggctggagac agagtggggc caaccocgtg agcgtggtaa ggacacctgg gaccgggtga 5220
aggctgcaca ggcctcgag ggcggccag acggcctgta cccccctagc tccctctctg 5280
tgtccaccgc accccaacc cgtcctctgc tgggtgtgta cctgcaggag gggccctggt 5340
gctccacctc gagcctcagc ctggacagcg accagagtag tggctcaacc acatccgctt 5400
cccccaagc tgcccgcgcg agcaaccagc cctgtacag ccagttccag acagcagaga 5460
gtgagaacag gtccacagag ggcactctgt acaagaaggg ggccttcata aagccttggg 5520
agggccgctg gttctgtctg gacaagacca agcaccagct gcgctactac gaccaccgtg 5580
tggaccacga gtgcaagggt tctatcgact tggcggaggt gtaggctgtg gcaacctgca 5640
cgccactatb ggggtccccct aagacctggg acgagaaggg cttctttgac gtgaaagcaa 5700
ogctctcgtt ttacaacttc tgtgccagg acgtgccctc ggcccagcag tgggtggacc 5760
ggatccagag ctgctctctg gacgctctag cctcccagcc ctgcccggct gctctgtctt 5820
cggctgttac cgaccactag ggggtggcag ggcgcgcgcg gccatgttta cagcccggcc 5880
cctcagacat attgagccc cgagccccca gcaacttgtt gtacagcccc cgtcccgcgc 5940
ccgcccgcgc cggccggccc taacttattt tggcgtcaca gctgagcacc gtgcccggag 6000
gtggccaagg tacagcccc aatggcctg taaatagctc ggcctcctca cgtgtctgtc 6060
gtccagccag cggctgcagg cgagttctta gaaccagagt ctatataaag agagaactaa 6120
cgccaaaaaa aaaaaaaaaa aaaaaa 6146

```

```

<210> 13
<211> 2362
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71918969CB1

```

```

<400> 13
gctctcaact gtgaatatgt gtctagtcca tctttaacct gaggacttca ccagttcgaa 60
attcacagttt tcaccatcaa ctaccittatc ctttttggcc tgggttttctt cctcaaacag 120
tggaaaacatt tttaaatgtg cttttgttgc agagtcaaac aaatggctga tagtggctla 180
gataaaaaat cccaaaaatg ccccgactgt tcactctgctt ctcaaaaaga tggactttgtg 240
gtatgttcca gaaaaacaag ggttccctca gttttgtgtg tggaaatgtc acagacaatca 300
agcatttgta gtgcagatc tttaatttca ctggagagaa aaaaagaaaa aatatacaac 360
agagataata cctccagggaa agatttgccc tcaagaacct caaatgtaga gagaagaa 420
tctcagaacac aatggggtcg ggcacaactt acagaagaaa aagtctctca cataagatt 480
ggaatggag ctgctattga ggaactctat acctttggaa gaattattggg aaaagggagc 540
tttggaaatg tcaattgagc gacagacaag gaaacagaaa cgaagtgggc aatataaaaa 600
gtgaacaagg aaaaaggctg agctctctgc gtgaagttaac ttgaacgaga ggtgaacatt 660
ctgaaaagtg taaaacatga acacatcata catctggaac aagtatttga aacgcaaaag 720
aaaatgtaac ttgtgatgga gctttgtgag gatgggaaac tcaaaagaaat tctggatagg 780
aaaagggcatt totcagagaa tgagacaagg tggateattc aaagtctcgc atcagatata 840
gcatalettc acataatgta tattgtacat agagatctga aactggaaaa tataatgglt 900
aaaagcagtc ctattgctga taacaatgaa ataaacttaa acanaaggt gactgatttt 960
ggcttagcgg tgagaagca aagttagagt gaagcctatg tgaaggcac atgtggact 1020
cctatctata tggcccctga agttatcagt gccaccagct atagcccaga gtgtgacatt 1080
tggagcatag gctctgtaat gtacatgtta ttaagtggag aaccaccctt tttggcaagc 1140
tcagaagaga agctttttga gtttaataaga aaaggagaaac tacattttga aatgcaatc 1200
tggaaatcca taagtactg tgcataaagt gttttgaaac aacttatgaa agtagactct 1260
gctccagaaa tcacagctaa ggaactacta gataaccagt ggttaacagg caataaactt 1320
totctggtag gaccacaaca tgtattagag atgatgaagg aatggaaaaa taaccagaaa 1380
agtgcttgag aaaaacacaac agaagagaaag aataagcctt ccaactgaaga aagttgaaa 1440
agttaccaac cctggggaaa tgtccctgat gccaatata cttcagatga agaggaggaa 1500

```

```

aaacagtccta ctgcttatga aaagcaatct cctgcaacca gtaaggacaa ctttgatag 1560
tgcagttcaa gtttcacatc tagcaaacct cttccagctg aaatcaaggg agaaatggag 1620
aaaccccctg tgnctccaag ccaaggaaaca gcaactcaagt accctgtcaa atccgggccc 1680
ctgtcccagaa ccaaaaagaa actctaaggt tccctccagt gttggacagt acaaaaacaa 1740
agctgctctt gttagcactt tgataggggg gtaggagggg aagaagacag cccatgctg 1800
agcttgtagc cttttagctc cacagagccc cggcatgtgt ttgcaccagg ttaaaatga 1860
agctgcttat ctccaagca gcataagctg cacgtggcat taaaggacag ccaccagtag 1920
gcttggcagt gggctgagc ggaatcaac tcaagatga cacgaaggtt ttttaggggg 1980
gcagataacc tcaatttaag gctgtgggca cactgtctca tttttacttc aaattcttat 2040
gcttaggcac agctatttat aggggaaaac aagaggccaa atattagtaat ggaggtgcca 2100
aataattatg tgcactttgc actagaagac tttgttagaa aattactaat aaacttgcca 2160
tacgtattac agcagaagtg cttcagtcac tccatgtgt tctgtagatt ttaggttgc 2220
atagattggt taagacagct tattttaaat gtgaaaaaa agggatcttt gtaactgctt 2280
gccattaac tgcctgcaaa ttcccattgt attgattaaa tcaatcaaaa aacagatggt 2340
cctcagcaca aacacacaaa aa 2362

```

```

<210> 14
<211> 1535
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<223> Incyte ID No: 8187571CB1

```

```

<400> 14
ccacacctc ccttagcctc ctgtgtggga ggaatttat ggtatgtggc tctgcccag 60
tccagtgagg ctttcacttc tactctattt cagttcctct ttcccgatct gggctggaga 120
gctttctcat tgttaaggca gcagaaactt tctctggatg gttttagatc aaggggtcat 180
caatgctggc adegctggc acaatgagga ccaggcttgc tctgaagtgg tctatgtgga 240
aggtgcgagg agtgttacag gactacatag ggagcctagc ctaggcccag gactctgctt 300
ctactactgg ggccattttg atgggcatcg agggggcgga gctgctgaaa tggcctcagc 360
agctctgcat cgcctatctc gagagcagct aaaggacctg gtagagatcc ttcaggacc 420
ttcgccacca cccctctgcc tcccaaccac tccggggacc ccagattctt ccgatccctc 480
tcaattgctt ggccctcagt cctgctggtc ttccagaaag gaagttagcc acgagagcct 540
ggtagtgagg gccattgaga atgccttcca gctcatggat gagcagatgg cccgggagcg 600
scgtggccac caagtggagg gggctgctg tgcactggtt gtgatctacc tgcaggcaca 660
gggtgactgt gccaatgagc gcgatagcag ggccatcatt gtccggaaatg gtgaaatcat 720
tccaatgtcc ogggagttta ccccgagac tgagcggcag cgtcttcagc tgccttggctt 780
cctgaaacca gagctgctag gcagtgaatt caccacctt gaggctccc gcagagttct 840
gcccaggag ctggggcaga ggaatttga cggggaccg aacatgaccg cctgggacctc 900
caaaaagatc gactggagg atctcaggtt tctctggctc tgtggggagg gcaaaaaggc 960
tccggtgatg gccaccatg gggtagcccg aggtttggga gaccacagcc ttaaggtctg 1020
cagttccaac ctgcccatca agccctttct ctctgtctc cctgagctac gactgtatga 1080
cctgacacaa tatgagcact gccagatga tgtgctagtc ctgggaacag atggcctgtg 1140
ggatgtcact actgactctg aggtagctgc cactgtggac aggggtgctg cggcctatga 1200
gctcaatgac cacagcaggt atacagctct ggcccaggct ctggtcctg 999ccgggg 1260
tacccccaga gaccgtggct ggcgtctccc caacaacaag ctgggttccc gggatgacat 1320
ctctgtcttc gtcatcccc tggggggggc aggcagttac tcttgggggg ctgaaaccca 1380
tcccctccat tagcctctcc atacttaact ctctcaacag ccaaatctg aagttgtctc 1440
cctgaccctt ctttagtggc aacttaactg aagaagggat gtcctgata tccaaaatta 1500
cagctattgg caaataaacg agatggataa aaaaa 1535

```

```

<210> 15
<211> 1376
<212> DNA
<213> Homo sapiens
<220>
<221> misc_feature
<223> Incyte ID No: 7494145CB1

```

```

<400> 15

```

```

gggaggagga atcagcaaaag atggtttgcag agggatcccc agagaggcag gaggaaaacc 60
aaggaaagtc ccccttcaca gaatcaanga ggtcttcaaa aagggcggca tggattatggt 120
gttaaatgctt accaaagagt aaagcaagct gctctgtttt cagttgtgca catctgaatg 180
caagcaatcc ctgtctgatg tggagtttct tgcactgata aggaaaaact gctgaagtgt 240
tgaggctgct ccaggcagag ccatcatcta acagacctaa aagaagcate atgtccatg 300
acctatcttc accccagggg acttcaagct gccctgtccc agaagtcca gactaaaagg 360
ccaggagata acagtgatg ttttcaaggaa gaggatctga ggcagggttt tcagtggagg 420
aagagcctcc cttttggggc agcctcatct tacttgaact tggagaagct gggtagaagg 480
tcctatgcga cagtttacaa ggggattagc agaataaatg gacaactagt ggccttaaaa 540
gtcatcagca tgaatgcaga ggaaggagtc ccatttacag ctatcccgaga agctctctc 600
ctgaagggtt tgaaacatgc caatattgtg ctctctcatg acataatcca caccaaaagag 660
acactgacat tcttttttga atacatgcac acagacctgg cccagtatat gtctcagcat 720
ccaggagggc ttcattctca taatgtcaga cttttcatgt ttcaactttt gggggcctg 780
gctacatcc accaccaaca cgttctctcac agggacctga aacctcagaa ctllactatc 840
agtcaacctgg gagagctcaa actggctgat tttggcttg cccgggcaaa gtccatctcc 900
agccagacat actcttcaga agtcgtgacc ctctggtaac ggcctctgga tgccttgctg 960
ggagcactg aatattctct tgaactggac atatgggggt caggctgcat ctltattgaa 1020
atgttccagg gtcacacttt gtttctctgg gtttccaaca tcttgaaca gctggagaa 1080
atctggaggf tgetggggct cctacacagag gatactggc cgggagctc caagctacct 1140
aaactcaatc cagaggagtc tttgtttaca gtttcaggag tgaggctaaa gccagaaatg 1200
tgtgaccttt tggctcctca ccagaaaggt caccaccagc cccagtttag caaatgctg 1260
tgaaaagaaa gggcgagatc accaaggctc ttccagggtg gtatttctgc agtttcgggt 1320
ttcatttgc tcaacttact aagaagcttc aaatctaact ccatactgaa caaggg 1376

```

```

<210> 16
<211> 1482
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc feature
<223> Incyte ID No: 5807954CB1

```

```

<400> 16
gccgccgacc aggcctcgaa cgggatggca gatgaggtga ggagggagga agcgtccgga 60
gtggggcggg cccgagctcc gacccccggg gcgcttcgag cccccagct ggttaaccag 120
gcaccgccgc ttcaccocag ccagtagccg cccctctcgg caecccgccc ccgctcaaca 180
cgcgcgcccc agcagcccc gggctcccc ctggccccag gtggcccgag ggtcaagtgt 240
tctctcgggt ctggggacag gtgagcaacc tgatgaagge caaggctctg atggggcagc 300
ctgggcgggt gcagygatc gtgggcggcc tccgcaaggg cggcggagac cgtttacagc 360
tgatttctga ttttgacatg acctgagca ggtttgcata taalggaaag cgatgcccct 420
cttttcaaa tattctggat aatagcaaga tcatcagtga ggaagtctcg aaagagctca 480
cagcctctct tcaacctatc tacccaattg agatcgacce acaccggacc gtcagaagga 540
agctacctca tatggtggaa tgggtggacca aagcgcaaaa tctctatgt cagcaagaaga 600
ttcagaagt tccagatgcc cgggtgggta gagagccaaa tgcactctc agggagggat 660
ataagaactt cttaacacaca ctctaccata scaacattcc cttttctac ttttctcggg 720
gcattgtgta tatctcggaa gaaattatcc gacagatgaa agtgttccac cccaacatcc 780
acatcgtctc taactacatg gatttlaatg aagatggttt tctccagggg ttaagggccc 840
agctcataca cacatacaac aagaacagct ctgctgttga gaactctggt tacttccagc 900
aaactgagggt caaaaccaat gtcactctgc tggggagactc fatcggggac ctcaacctag 960
ccgatggggt tccctggtgtg cagaacattc tcaaaattgg ctctctgaat gacaaggttg 1020
aggagcggcg ggaagcgtac atgactctct atgacatcgt gctggagaag gacgagactc 1080
tggatgtggt caacgggata ctgcagcaaa tctctgtgca gggggtccag ctggagatgc 1140
aagccccctg aagcggcagg ctccagcccc gctctgaggg cgtgtgtgag aggggcgctc 1200
ccccagatc tgcctccccg tgaacacaga gcagagggcca ggggtggccag cagtgtctg 1260
gctcttccgc gccctctcgt cctctcttcc ctgagacctc tcatcaccag aggcttgaag 1320
gaaccccccc atgtggcag gcacaggcac tgttctctgt gaaccttggg ccacagcatg 1380
tcagtctctc agggattgtc tactccaggg atttcttca aaatttttaa acatggggaag 1440
ttcaaacaaa tataatgtgt gaacagaaaa aaaaaaaaaa aa 1482

```

【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
14 November 2002 (14.11.2002)

PCT

(10) International Publication Number
WO 02/090530 A3

- (51) International Patent Classification: C12N 15/54, 15/55, 9/12, 9/16, C07K 16/40, C12Q 1/68, G01N 33/573, A61K 45/00
- (21) International Application Number: PCT/US02/01369
- (22) International Filing Date: 16 January 2002 (16.01.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- | | | |
|------------|-------------------------------|----|
| 60/263,083 | 18 January 2001 (18.01.2001) | US |
| 60/271,205 | 23 February 2001 (23.02.2001) | US |
| 60/271,117 | 23 February 2001 (23.02.2001) | US |
| 60/276,859 | 16 March 2001 (16.03.2001) | US |
| 60/278,504 | 23 March 2001 (23.03.2001) | US |
| 60/278,522 | 23 March 2001 (23.03.2001) | US |
| 60/280,510 | 29 March 2001 (29.03.2001) | US |
| 60/280,266 | 29 March 2001 (29.03.2001) | US |
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Potter Drive, Palo Alto, CA 94304 (US).
- (72) Inventors; and
(75) Inventors/Applicants (for US only): LEE, Ernestine, A. [US/US]; 624 Kaïns Street, Albany, CA 94705 (US). CHAWLA, Narinder, K. [US/US]; 33 Union Square, #712, Union City, CA 94587 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). ISON, Craig, H. [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). GURURAJAN, Rajagopal [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). ARVEZU, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46633 (US). JACKSON, Jennifer, L. [US/US]; 4913 Deep
- Creek Road, Fremont, CA 94555 (US). TANG, Tom, Y. [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). TRAN, Bao [US/US]; 750 Saïberg Avenue, Santa Clara, CA 95051 (US). DING, Li [US/US]; 3353 Alamo Street #146, Palo Alto, CA 94306 (US). LU, Dyung, Anna, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). LAL, Preeti, G. [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). WARREN, Bridget, A. [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US).
- (74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Potter Drive, Palo Alto, CA 94304 (US).
- (81) Designated States (national): AI, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NA, 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, SI, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NI, SN, TD, TG).
- Published: — with international search report
- (88) Date of publication of the international search report: 20 November 2003
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/090530 A3

(54) Title: KINASIS AND PHOSPHATASES

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

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 02/01369
A. CLASSIFICATION OF SUBJECT MATTER		
IPC 7	C12N15/54 C12Q1/68	C12N15/55 G01N33/573
	C12N9/12 A61K45/00	C12N9/16 C07K16/40
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		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
EMBL, EPO-Internal, SEQUENCE SEARCH, WPI Data, PAJ, BIOSIS, EMBASE, MEDLINE		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KEYSE S M: "AN EMERGING FAMILY OF DUAL SPECIFICITY MAP KINASE PHOSPHATASES" BIOCHIMICA ET BIOPHYSICA ACTA. MOLECULAR CELL RESEARCH, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 1265, 1995, pages 152-160, XP000196716 ISSN: 0167-4889 the whole document	1-20, 23, 26-56, 64
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents:		
A document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claims) or which is cited to establish the publication date of another claim or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *Z* document member of the same patent family		
Date of the actual completion of the international search	Date of mailing of the international search report	
5 May 2003	26.06.03	
Name and mailing address of the ISA European Patent Office, P.O. Box 5818 Patentsaan 2 NL - 2200 PH Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl Fax: (+31-70) 340-3010	Authorized officer Weikl, M	

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 02/01369
G/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 GENESEQ 'Online! 18 December 2001 (2001-12-18) "Human cDNA encoding dual specific phosphatase, 18232" Database accession no. AAS12700 XP002230436 the whole document	1-20,23, 26-56,64
P, X	-& DATABASE GENESEQ 'Online! 18 December 2001 (2001-12-18) "Human dual specificity phosphatase, 18323" Database accession no. AAU08646 XP002230436 the whole document	1-20,23, 26-56,64
P, X	-& WO 01 64911 A (MILLENNIUM PHARMACEUTICALS INC) 7 September 2001 (2001-09-07) figure 1	1-20,23, 26-56,64
P, X	DATABASE EMBL 'Online! 9 March 2001 (2001-03-09) "Sequence 25 from patent wo0112819" Database accession no. AX086028 XP002230437 the whole document	1-20,23, 26-56,64
P, X	-& DATABASE EPOP 'Online! 6 March 2001 (2001-03-06) "Sequence 26 from Patent wo0112819" Database accession no. ax086029 XP002230760 the whole document	1-20,23, 26-56,64
P, X	-& WO 01 12819 A (SUGEN INC) 22 February 2001 (2001-02-22) figures 4,5	1-20,23, 26-56,64
A	MARTI F ET AL: "Negative-feedback regulation of CD28 costimulation by a novel mitogen-activated protein kinase phosphatase, MKP6." JOURNAL OF IMMUNOLOGY (BALTIMORE, MD.: 1950) UNITED STATES 1 JAN 2001, vol. 166, no. 1, 1 January 2001 (2001-01-01), pages 197-206, XP002230434 ISSN: 0022-1767 cited in the application figure 1	1-20,23, 26-56,64
A, P	-& DATABASE SWALL 'Online! 15 June 2001 (2001-06-15) "Dual specificity protein phosphatase 14 (MAP kinase phosphatase 6)" Database accession no. 095147 XP002230438 the whole document	1-20,23, 26-56,64
	---	-/--

INTERNATIONAL SEARCH REPORT		International Application No PCT/US 02/01369
C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LI L ET AL: "Form, function, and regulation of protein tyrosine phosphatases and their involvement in human diseases." SEMINARS IN IMMUNOLOGY. UNITED STATES FEB 2000, vol. 12, no. 1, February 2000 (2000-02), pages 75-84, XP002240063 ISSN: 1044-5323 the whole document	1-20, 23, 26-55, 57, 65
X	WRIGHT M B ET AL: "Proliferating and migrating mesangial cells responding to injury express a novel receptor protein-tyrosine phosphatase in experimental mesangial proliferative glomerulonephritis" JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, US, vol. 273, no. 37, 11 September 1998 (1998-09-11), pages 23929-23937, XP002229880 ISSN: 0021-9258 cited in the application figure 7	1-20, 23, 26-55, 57, 65
X	-& DATABASE EMBL 'Online!' 13 July 1998 (1998-07-13) "rattus norvegicus glomerular mesangial cell receptor protein-tyrosine phosphatase precursor (PTPRQ) mRNA, complete cds" Database accession no. AF063249 XP002240064 the whole document	1-20, 23, 26-55, 57, 65
X	-& DATABASE SWALL 'Online!' 1 November 1998 (1998-11-01) "Glomerular mesangial cell receptor protein-tyrosine phosphatase precursor" Database accession no. 088488 XP002240065 the whole document	1-20, 23, 26-55, 57, 65

	-/-	

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 02/01369
C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! 13 February 1997 (1997-02-13) "Sequence 21 from patent US5585233" Database accession no. I32039 XP002240066 the whole document	1-20, 23, 26-55, 57, 65
X	-& DATABASE GENESEQ 'Online! 7 February 1997 (1997-02-07) "Sequence 22 from patent US5585233" Database accession no. AAB44477 XP002240067 the whole document	1-20, 23, 26-55, 57, 65
X	-& US 5 585 233 A (MAX PLANCK GESELLSCHAFT) 17 December 1996 (1996-12-17) column 1-4; figure 12	1-20, 23, 26-55, 57, 65
T	DATABASE GENESEQ 'Online! 24 October 2002 (2002-10-24) "Human Nov2c coding sequence" Database accession no. ABT06283 XP002240068 the whole document	1-20, 23, 26-55, 57, 65
T	-& DATABASE GENESEQ 'Online! 24 October 2001 (2001-10-24) "Human Nov2c protein" Database accession no. AA018738 XP002240069 the whole document	1-20, 23, 26-55, 57, 65
E	-& WO 02 057450 A (CURAGEN CORPORATION) 25 July 2002 (2002-07-25) table A	1-20, 23, 26-55, 57, 65

INTERNATIONAL SEARCH REPORT	
International application No. PCT/US 02/01369	
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)	
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:	
1.	<input checked="" type="checkbox"/> Claims Nos. because they relate to subject matter not required to be searched by this Authority, namely: see FURTHER INFORMATION sheet PCT/ISA/210
2.	<input checked="" type="checkbox"/> Claims Nos. 21, 22, 24, 25, 48, 52-55 because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically: see FURTHER INFORMATION sheet PCT/ISA/210
3.	<input type="checkbox"/> Claims Nos. because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)	
This International Searching Authority found multiple inventions in this international application, as follows:	
see additional sheet	
1.	<input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2.	<input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	<input checked="" type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.: 1-55 (insofar as they relate to SEQ IDs 1, 2, 9 and 10), 56, 57, 64, 65
4.	<input type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest	
	<input type="checkbox"/> The additional search fees were accompanied by the applicant's protest.
	<input checked="" type="checkbox"/> No protest accompanied the payment of additional search fees.

International Application No. PCT/US 02 01369

FURTHER INFORMATION CONTINUED FROM PCT/SA/ 210

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

1. Claims: 1-55 (insofar as they relate to SEQ ID Nos 1 and 9), 56, 64
2. Claims: 1-55 (insofar as they relate to SEQ ID Nos 2 and 10), 57, 65
3. Claims: 1-55 (insofar as they relate to SEQ ID Nos 3 and 11), 58, 66
4. Claims: 1-55 (insofar as they relate to SEQ ID Nos 4 and 12), 59, 67
5. Claims: 1-55 (insofar as they relate to SEQ ID Nos 5 and 13), 60, 68
6. Claims: 1-55 (insofar as they relate to SEQ ID Nos 6 and 14), 61, 69
7. Claims: 1-55 (insofar as they relate to SEQ ID Nos 7 and 15), 62, 70
8. Claims: 1-55 (insofar as they relate to SEQ ID Nos 8 and 16), 63, 71

International Application No. PCT/US 02 /01369

FURTHER INFORMATION CONTINUED FROM PCT/AS/ 210

Continuation of Box I.1

Although claims 30, 33 and 35 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claim 19 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.2

Claims Nos.: 21, 22, 24, 25, 48, 52-55

Present claims 21, 22, 24 and 25 relate to a compound defined by reference to a desirable characteristic or property, namely agonistic/antagonistic activity concerning a polypeptide having SEQ ID No 1.

The claims cover all compositions and methods incorporating such compounds having this characteristic or property, whereas the application provides no support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for these compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the compound by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, no search has been carried out for claims 21, 22, 24 and 25.

Present claims 48 and dependent claims 52-55 relate to a nucleotide sequence which is only defined by reference to a desirable property namely hybridization with at least 30 contiguous nucleotides of the polynucleotide of claim 12 (the polynucleotide of the invention). Due to the lack of hybridization conditions and further structural or functional definitions of this nucleotide, basically every nucleotide could fall within the definition as set out by claim 48. Any nucleotide sequence can be imagined to hybridize under at least one particular condition to an arbitrary 30 nucleotide stretch of SEQ ID No 9. This lack of clarity is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search of claims 48 and 52-55 has been restricted to nucleotide sequences which are complementary to stretches of SEQ ID No 9.

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

International Application No. PCT/US 02 01369

FURTHER INFORMATION CONTINUED FROM PCT/SAI 210

receipt of the search report or during any Chapter II procedure.

フロントページの続き

(51)Int.Cl. ⁷	F I	テーマコード(参考)
A 6 1 P 1/16	A 6 1 P 1/16	4 B 0 6 5
A 6 1 P 1/18	A 6 1 P 1/18	4 C 0 8 4
A 6 1 P 3/00	A 6 1 P 3/00	4 C 0 8 5
A 6 1 P 3/04	A 6 1 P 3/04	
A 6 1 P 3/06	A 6 1 P 3/06	
A 6 1 P 3/10	A 6 1 P 3/10	
A 6 1 P 3/12	A 6 1 P 3/12	
A 6 1 P 5/00	A 6 1 P 5/00	
A 6 1 P 5/14	A 6 1 P 5/14	
A 6 1 P 5/38	A 6 1 P 5/38	
A 6 1 P 5/40	A 6 1 P 5/40	
A 6 1 P 7/00	A 6 1 P 7/00	
A 6 1 P 7/02	A 6 1 P 7/02	
A 6 1 P 7/04	A 6 1 P 7/04	
A 6 1 P 7/06	A 6 1 P 7/06	
A 6 1 P 7/08	A 6 1 P 7/08	
A 6 1 P 9/00	A 6 1 P 9/00	
A 6 1 P 9/04	A 6 1 P 9/04	
A 6 1 P 9/08	A 6 1 P 9/08	
A 6 1 P 9/10	A 6 1 P 9/10	
A 6 1 P 9/12	A 6 1 P 9/10	1 0 3
A 6 1 P 11/00	A 6 1 P 9/12	
A 6 1 P 11/06	A 6 1 P 11/00	
A 6 1 P 13/02	A 6 1 P 11/06	
A 6 1 P 13/12	A 6 1 P 13/02	
A 6 1 P 15/00	A 6 1 P 13/12	
A 6 1 P 17/00	A 6 1 P 15/00	
A 6 1 P 17/02	A 6 1 P 17/00	
A 6 1 P 17/06	A 6 1 P 17/02	
A 6 1 P 19/00	A 6 1 P 17/06	
A 6 1 P 19/02	A 6 1 P 19/00	
A 6 1 P 19/04	A 6 1 P 19/02	
A 6 1 P 19/06	A 6 1 P 19/04	
A 6 1 P 19/10	A 6 1 P 19/06	
A 6 1 P 21/00	A 6 1 P 19/10	
A 6 1 P 21/04	A 6 1 P 21/00	
A 6 1 P 25/00	A 6 1 P 21/04	
A 6 1 P 25/02	A 6 1 P 25/00	
A 6 1 P 25/08	A 6 1 P 25/02	
A 6 1 P 25/14	A 6 1 P 25/02	1 0 3
A 6 1 P 25/16	A 6 1 P 25/08	
A 6 1 P 25/18	A 6 1 P 25/14	
A 6 1 P 25/20	A 6 1 P 25/16	
A 6 1 P 25/22	A 6 1 P 25/18	
A 6 1 P 25/28	A 6 1 P 25/20	
A 6 1 P 27/02	A 6 1 P 25/22	
A 6 1 P 27/06	A 6 1 P 25/28	

A 6 1 P 27/12	A 6 1 P 27/02	
A 6 1 P 27/16	A 6 1 P 27/06	
A 6 1 P 29/00	A 6 1 P 27/12	
A 6 1 P 31/04	A 6 1 P 27/16	
A 6 1 P 31/06	A 6 1 P 29/00	
A 6 1 P 31/10	A 6 1 P 29/00	1 0 1
A 6 1 P 31/12	A 6 1 P 31/04	
A 6 1 P 31/18	A 6 1 P 31/06	
A 6 1 P 31/22	A 6 1 P 31/10	
A 6 1 P 33/00	A 6 1 P 31/12	
A 6 1 P 33/02	A 6 1 P 31/18	
A 6 1 P 33/10	A 6 1 P 31/22	
A 6 1 P 35/00	A 6 1 P 33/00	
A 6 1 P 35/02	A 6 1 P 33/02	
A 6 1 P 37/02	A 6 1 P 33/10	
A 6 1 P 37/04	A 6 1 P 35/00	
A 6 1 P 37/06	A 6 1 P 35/02	
A 6 1 P 37/08	A 6 1 P 37/02	
A 6 1 P 41/00	A 6 1 P 37/04	
C 0 7 K 16/40	A 6 1 P 37/06	
C 0 7 K 19/00	A 6 1 P 37/08	
C 1 2 M 1/00	A 6 1 P 41/00	
C 1 2 M 1/34	C 0 7 K 16/40	
C 1 2 N 1/15	C 0 7 K 19/00	
C 1 2 N 1/19	C 1 2 M 1/00	A
C 1 2 N 1/21	C 1 2 M 1/34	Z
C 1 2 N 5/10	C 1 2 N 1/15	
C 1 2 N 9/12	C 1 2 N 1/19	
C 1 2 N 9/16	C 1 2 N 1/21	
C 1 2 N 15/02	C 1 2 N 9/12	
C 1 2 P 21/02	C 1 2 N 9/16	B
C 1 2 P 21/08	C 1 2 P 21/02	C
C 1 2 Q 1/02	C 1 2 P 21/08	
C 1 2 Q 1/42	C 1 2 Q 1/02	
C 1 2 Q 1/48	C 1 2 Q 1/42	
C 1 2 Q 1/68	C 1 2 Q 1/48	Z
G 0 1 N 33/15	C 1 2 Q 1/68	A
G 0 1 N 33/50	C 1 2 Q 1/68	Z
G 0 1 N 33/53	G 0 1 N 33/15	Z
G 0 1 N 33/566	G 0 1 N 33/50	Z
	G 0 1 N 33/53	D
	G 0 1 N 33/53	M
	G 0 1 N 33/566	
	C 1 2 N 15/00	F
	C 1 2 N 5/00	A
	C 1 2 N 15/00	C
	A 6 1 K 37/02	

(31)優先権主張番号 60/276,859

(32)優先日 平成13年3月16日(2001.3.16)

- (33)優先権主張国 米国(US)
 (31)優先権主張番号 60/278,504
 (32)優先日 平成13年3月23日(2001.3.23)
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 60/278,522
 (32)優先日 平成13年3月23日(2001.3.23)
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 60/280,510
 (32)優先日 平成13年3月29日(2001.3.29)
 (33)優先権主張国 米国(US)
 (31)優先権主張番号 60/280,266
 (32)優先日 平成13年3月29日(2001.3.29)
 (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,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT, BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN, TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE, GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,R O,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW
- (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94587・ユニオンシティ・#712・ユニオンスクエア 3
 3
 (72)発明者 ボーゲン、マライア・アール
 アメリカ合衆国カリフォルニア州94577・サンレアンドロ・サンティアゴロード 14244
 (72)発明者 アイソン、クレイグ・エイチ
 アメリカ合衆国カリフォルニア州95118・サンノゼ・ウェザーズフィールドウェイ 1242
 (72)発明者 ガルラジャン、ラジャゴパル
 アメリカ合衆国カリフォルニア州95118・サンノゼ・デントアベニュー 5591
 (72)発明者 アービズ、チャンドラ
 アメリカ合衆国カリフォルニア州94025・メンロパーク・#1・シャーウッドウェイ 490
 (72)発明者 ヤオ、モニーク・ジー
 アメリカ合衆国インディアナ州40633・カーメル・ウッドゲートドライブ 1189
 (72)発明者 ジャクソン、ジェニファー・エル
 アメリカ合衆国カリフォルニア州94555・フレモント・ディーブクリークロード 4913
 (72)発明者 タング、トム・ワイ
 アメリカ合衆国カリフォルニア州95118・サンノゼ・ランウィックコート 4230
 (72)発明者 ユエ、ヘンリー
 アメリカ合衆国カリフォルニア州94087・サニーベイル・ルイスアベニュー 826
 (72)発明者 トラン、バオ
 アメリカ合衆国カリフォルニア州95051・サンタクララ・サルバゲアベニュー 750
 (72)発明者 ディング、リー
 アメリカ合衆国カリフォルニア州94306・パロアルト・#146・アルマストリート 335
 3
 (72)発明者 リュ、デュング・アイナ・エム
 アメリカ合衆国カリフォルニア州95123・サンノゼ・コイドライブ 233
 (72)発明者 ラル、ブリーティ・ジー
 アメリカ合衆国カリフォルニア州95056・サンタクララ・ピーオーボックス 5142
 (72)発明者 ワレン、ブリジット・エイ
 アメリカ合衆国カリフォルニア州95014・クーペルティノー・#2・パークウッドドライブ

1 0 1 3 0

Fターム(参考) 4B024 AA01 AA11 AA19 BA10 BA11 BA53 BA61 CA04 CA07 CA09
CA12 CA20 DA02 DA06 EA02 EA04 FA02 GA11 GA18 GA19
HA03 HA11 HA13 HA14
4B029 AA23 BB20 CC01 CC02 CC03 CC08 FA12 FA15
4B050 CC01 CC03 CC05 DD11 EE01 FF03E FF11E FF14E LL01 LL03
LL05
4B063 QA01 QA05 QA19 QQ02 QQ08 QQ21 QQ27 QQ33 QQ41 QQ53
QQ61 QQ89 QR08 QR32 QR35 QR40 QR42 QR48 QR56 QR62
QR77 QR80 QS16 QS25 QS33 QS34 QS36 QX01 QX02
4B064 AG26 AG27 CA01 CA02 CA10 CA19 CA20 CC01 CC24 CE02
CE03 CE12 DA01 DA13
4B065 AA01X AA26X AA58X AA72X AA87X AA90X AA91X AA93Y AB01 AB05
AC14 BA02 BA08 CA25 CA29 CA31 CA43 CA44 CA46
4C084 AA02 AA03 AA06 AA07 AA17 BA01 BA08 BA22 BA23 CA18
MA13 MA52 MA56 MA59 MA60 MA63 MA65 NA14 ZA012 ZA022
ZA032 ZA052 ZA062 ZA152 ZA162 ZA182 ZA202 ZA222 ZA242 ZA292
ZA332 ZA342 ZA362 ZA392 ZA402 ZA422 ZA452 ZA512 ZA532 ZA542
ZA552 ZA592 ZA662 ZA682 ZA702 ZA752 ZA812 ZA892 ZA942 ZA962
ZA972 ZB052 ZB072 ZB082 ZB092 ZB112 ZB132 ZB152 ZB262 ZB272
ZB332 ZB352 ZB372 ZB382 ZB392 ZC032 ZC042 ZC062 ZC082 ZC212
ZC312 ZC332 ZC352 ZC552
4C085 AA13 AA14 AA16 AA19 BB11 BB31 BB41 BB43 CC02 CC22
CC23 EE01

專利名稱(譯)	激酶和磷酸酶		
公開(公告)号	JP2004521637A	公開(公告)日	2004-07-22
申請号	JP2002587592	申請日	2002-01-16
[標]申請(專利權)人(譯)	洞察Genomics公司		
申請(專利權)人(譯)	洞察基因组公司		
[標]發明人	リーアーンステイーンエイ チョーラナリンダーケイ ボーグンマライアアール アイソクレイグエイチ ガルラジャンラジャゴバル アービズチャンドラ ヤオモニークジー ジャクソンジェニファーエル タングトムワイ ユエヘンリー トランバオ デイングリー リュデュングアイナエム ラルプリーティジー ワレンブリジットエイ		
發明人	リー、アーンステイーン・エイ チョーラ、ナリンダー・ケイ ボーグン、マライア・アール アイソン、クレイグ・エイチ ガルラジャン、ラジャゴバル アービズ、チャンドラ ヤオ、モニーク・ジー ジャクソン、ジェニファー・エル タング、トム・ワイ ユエ、ヘンリー トラン、バオ デイング、リー リュ、デュング・アイナ・エム ラル、プリーティ・ジー ワレン、ブリジット・エイ		
IPC分類号	G01N33/50 A61K38/00 A61K39/395 A61K45/00 A61P1/04 A61P1/16 A61P1/18 A61P3/00 A61P3/04 A61P3/06 A61P3/10 A61P3/12 A61P5/00 A61P5/14 A61P5/38 A61P5/40 A61P7/00 A61P7/02 A61P7 /04 A61P7/06 A61P7/08 A61P9/00 A61P9/04 A61P9/08 A61P9/10 A61P9/12 A61P11/00 A61P11/06 A61P13/02 A61P13/12 A61P15/00 A61P17/00 A61P17/02 A61P17/06 A61P19/00 A61P19/02 A61P19 /04 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/22 A61P25/28 A61P27/02 A61P27/06 A61P27/12 A61P27 /16 A61P29/00 A61P31/04 A61P31/06 A61P31/10 A61P31/12 A61P31/18 A61P31/22 A61P33/00 A61P33/02 A61P33/10 A61P35/00 A61P35/02 A61P37/02 A61P37/04 A61P37/06 A61P37/08 A61P41 /00 C07K16/40 C07K19/00 C12M1/00 C12M1/34 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/12 C12N9/16 C12N15/02 C12N15/09 C12N15/55 C12P21/02 C12P21/08 C12Q1/02 C12Q1/42 C12Q1/48 C12Q1/68 G01N33/15 G01N33/53 G01N33/566		
CPC分類号			

A01K2217/05 A61P1/04 A61P1/16 A61P1/18 A61P3/00 A61P3/04 A61P3/12 A61P11/00 A61P11/06
A61P13/02 A61P13/12 A61P15/00 A61P17/00 A61P17/02 A61P17/06 A61P19/00 A61P19/02 A61P19
/04 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14
A61P25/16 A61P25/18 A61P25/20 A61P25/22 A61P25/28 A61P27/02 A61P27/06 A61P27/12 A61P27
/16 A61P29/00 A61P31/04 A61P31/06 A61P31/10 A61P31/12 A61P31/18 A61P31/22 A61P33/00
A61P33/02 A61P33/10 A61P35/00 A61P35/02 C12N9/12 C12N9/16

F分类号

C12N15/00.ZNA.A A61K39/395.D A61K39/395.N A61K45/00 A61P1/04 A61P1/16 A61P1/18 A61P3/00
A61P3/04 A61P3/06 A61P3/10 A61P3/12 A61P5/00 A61P5/14 A61P5/38 A61P5/40 A61P7/00 A61P7
/02 A61P7/04 A61P7/06 A61P7/08 A61P9/00 A61P9/04 A61P9/08 A61P9/10 A61P9/10.103 A61P9/12
A61P11/00 A61P11/06 A61P13/02 A61P13/12 A61P15/00 A61P17/00 A61P17/02 A61P17/06 A61P19
/00 A61P19/02 A61P19/04 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02
A61P25/02.103 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/20 A61P25/22 A61P25/28
A61P27/02 A61P27/06 A61P27/12 A61P27/16 A61P29/00 A61P29/00.101 A61P31/04 A61P31/06
A61P31/10 A61P31/12 A61P31/18 A61P31/22 A61P33/00 A61P33/02 A61P33/10 A61P35/00 A61P35
/02 A61P37/02 A61P37/04 A61P37/06 A61P37/08 A61P41/00 C07K16/40 C07K19/00 C12M1/00.A
C12M1/34.Z C12N1/15 C12N1/19 C12N1/21 C12N9/12 C12N9/16.B C12P21/02.C C12P21/08 C12Q1
/02 C12Q1/42 C12Q1/48.Z C12Q1/68.A C12Q1/68.Z G01N33/15.Z G01N33/50.Z G01N33/53.D
G01N33/53.M G01N33/566 C12N15/00.F C12N5/00.A C12N15/00.C A61K37/02

F-TERM分类号

4B024/AA01 4B024/AA11 4B024/AA19 4B024/BA10 4B024/BA11 4B024/BA53 4B024/BA61 4B024
/CA04 4B024/CA07 4B024/CA09 4B024/CA12 4B024/CA20 4B024/DA02 4B024/DA06 4B024/EA02
4B024/EA04 4B024/FA02 4B024/GA11 4B024/GA18 4B024/GA19 4B024/HA03 4B024/HA11 4B024
/HA13 4B024/HA14 4B029/AA23 4B029/BB20 4B029/CC01 4B029/CC02 4B029/CC03 4B029/CC08
4B029/FA12 4B029/FA15 4B050/CC01 4B050/CC03 4B050/CC05 4B050/DD11 4B050/EE01 4B050
/FF03E 4B050/FF11E 4B050/FF14E 4B050/LL01 4B050/LL03 4B050/LL05 4B063/QA01 4B063/QA05
4B063/QA19 4B063/QQ02 4B063/QQ08 4B063/QQ21 4B063/QQ27 4B063/QQ33 4B063/QQ41 4B063
/QQ53 4B063/QQ61 4B063/QQ89 4B063/QR08 4B063/QR32 4B063/QR35 4B063/QR40 4B063/QR42
4B063/QR48 4B063/QR56 4B063/QR62 4B063/QR77 4B063/QR80 4B063/QS16 4B063/QS25 4B063
/QS33 4B063/QS34 4B063/QS36 4B063/QX01 4B063/QX02 4B064/AG26 4B064/AG27 4B064/CA01
4B064/CA02 4B064/CA10 4B064/CA19 4B064/CA20 4B064/CC01 4B064/CC24 4B064/CE02 4B064
/CE03 4B064/CE12 4B064/DA01 4B064/DA13 4B065/AA01X 4B065/AA26X 4B065/AA58X 4B065
/AA72X 4B065/AA87X 4B065/AA90X 4B065/AA91X 4B065/AA93Y 4B065/AB01 4B065/AB05 4B065
/AC14 4B065/BA02 4B065/BA08 4B065/CA25 4B065/CA29 4B065/CA31 4B065/CA43 4B065/CA44
4B065/CA46 4C084/AA02 4C084/AA03 4C084/AA06 4C084/AA07 4C084/AA17 4C084/BA01 4C084
/BA08 4C084/BA22 4C084/BA23 4C084/CA18 4C084/MA13 4C084/MA52 4C084/MA56 4C084/MA59
4C084/MA60 4C084/MA63 4C084/MA65 4C084/NA14 4C084/ZA012 4C084/ZA022 4C084/ZA032
4C084/ZA052 4C084/ZA062 4C084/ZA152 4C084/ZA162 4C084/ZA182 4C084/ZA202 4C084/ZA222
4C084/ZA242 4C084/ZA292 4C084/ZA332 4C084/ZA342 4C084/ZA362 4C084/ZA392 4C084/ZA402
4C084/ZA422 4C084/ZA452 4C084/ZA512 4C084/ZA532 4C084/ZA542 4C084/ZA552 4C084/ZA592
4C084/ZA662 4C084/ZA682 4C084/ZA702 4C084/ZA752 4C084/ZA812 4C084/ZA892 4C084/ZA942
4C084/ZA962 4C084/ZA972 4C084/ZB052 4C084/ZB072 4C084/ZB082 4C084/ZB092 4C084/ZB112
4C084/ZB132 4C084/ZB152 4C084/ZB262 4C084/ZB272 4C084/ZB332 4C084/ZB352 4C084/ZB372
4C084/ZB382 4C084/ZB392 4C084/ZC032 4C084/ZC042 4C084/ZC062 4C084/ZC082 4C084/ZC212
4C084/ZC312 4C084/ZC332 4C084/ZC352 4C084/ZC552 4C085/AA13 4C085/AA14 4C085/AA16
4C085/AA19 4C085/BB11 4C085/BB31 4C085/BB41 4C085/BB43 4C085/CC02 4C085/CC22 4C085
/CC23 4C085/EE01

优先权

60/263083 2001-01-18 US
60/271205 2001-02-23 US
60/271117 2001-02-23 US
60/276859 2001-03-16 US
60/278504 2001-03-23 US
60/278522 2001-03-23 US
60/280510 2001-03-29 US
60/280266 2001-03-29 US

外部链接

[Espacenet](#)

摘要(译)

本发明提供了鉴定和编码KPP的人激酶和磷酸酶 (KPP) 以及多核苷酸。 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KPP异常表达有关的疾病的方法。

と5を参照)。

接頭コード	解析タイプやプログラムの例
GNN 、 GFG 、 ENST	例えば、GENSCAN (Stanford University, CA, USA) または FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)を用いたゲノム配列からのエクソン予測
GBI	手動で編集されたゲノム配列の解析
FL	ステイッチまたはストレッチゲノム配列(実施例5参照)
INCY	EST 配列のゲノムへのマッピングからの全長転写とエクソンの予想エクソンと結果の転写を予想するために、ゲノム位置と EST 構成データが組み合わされる。