

(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2004-528002

(P2004-528002A)

(43) 公表日 平成16年9月16日(2004.9.16)

(51) Int. Cl. <sup>7</sup>	F I	テーマコード (参考)
<b>C 1 2 N 15/09</b>	C 1 2 N 15/00	Z N A A
<b>A 6 1 K 38/00</b>	A 6 1 K 45/00	2 G O 4 5
<b>A 6 1 K 45/00</b>	A 6 1 P 25/00	4 B O 2 4
<b>A 6 1 P 25/00</b>	A 6 1 P 29/00	4 B O 6 3
<b>A 6 1 P 29/00</b>	A 6 1 P 35/00	4 B O 6 4
		4 B O 6 5
	審査請求 未請求 予備審査請求 有	(全 233 頁) 最終頁に続く

(21) 出願番号	特願2002-507862 (P2002-507862)	(71) 出願人	301005050
(86) (22) 出願日	平成13年6月28日 (2001.6.28)		インサイト・ゲノミックス・インコーポレイテッド
(85) 翻訳文提出日	平成14年12月26日 (2002.12.26)		アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160
(86) 国際出願番号	PCT/US2001/020704	(74) 代理人	100089266
(87) 国際公開番号	W02002/002610		弁理士 大島 陽一
(87) 国際公開日	平成14年1月10日 (2002.1.10)	(72) 発明者	リー、アーンステーション・エイ
(31) 優先権主張番号	60/215,465		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
(32) 優先日	平成12年6月29日 (2000.6.29)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/239,384		
(32) 優先日	平成12年10月10日 (2000.10.10)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/253,639		
(32) 優先日	平成12年11月28日 (2000.11.28)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 分泌分子および輸送分子

## (57) 【要約】

本発明はヒト分泌分子および輸送分子 (S A T)、並びに S A T 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、S A T 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 - 9 (配列番号 1 乃至 9) 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 - 9 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アミノ酸配列を含む天然のポリペプチドを有するポリペプチド

(c) SEQ ID NO: 1 - 9 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

(d) SEQ ID NO: 1 - 9 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抗原性断片 10

## 【請求項 2】

SEQ ID NO: 1 - 9 を有する群から選択した請求項 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: 10 - 18 (配列番号 7 乃至 12) からなる群から選択した請求項 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結合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 のポリペプチドを生産する方法であって、以下の過程を含む方法。 30

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

(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。

## 【請求項 10】

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 11】

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ヌクレオチド。

(a) SEQ ID NO: 10 - 18 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 40

(b) SEQ ID NO: 10 - 18 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% が同一であるような天然のポリヌクレオチド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

(e) (a) ~ (d) の RNA 等価物

## 【請求項 12】

請求項 11 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。

## 【請求項 13】

請求項 11 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 50

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含む少なくとも 20 の連続したヌクレオチドを含む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 (ただし、前記プローブと前記標的ポリヌクレオチドまたは断片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、前記プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズする) と、

( b ) 前記ハイブリダイゼーション複合体の存在・不存在を検出し、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からなり、前記プローブと前記標的ポリヌクレオチドまたは断片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、前記プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズすることを特徴とする方法。

10

【請求項 14】

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 13 に記載の方法。

【請求項 15】

請求項 11 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 16】

請求項 1 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 17】

前記ポリペプチドが、SEQ ID NO: 1 - 9 を有する群から選択したアミノ酸配列を含むことを特徴とする請求項 16 に記載の成分。

【請求項 18】

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

30

【請求項 19】

請求項 1 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露する過程と、

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

【請求項 20】

請求項 19 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組成物。

【請求項 21】

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

40

【請求項 22】

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露する過程と、

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

【請求項 23】

請求項 22 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

50

る賦形剤とを含む組成物。

【請求項 24】

機能的な S A T の過剰な発現に関連する疾患や病態の治療方法であって、そのような治療が必要な患者に請求項 23 の組成物を投与することを含むことを特徴とする治療方法。

【請求項 25】

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 ) 適切な条件下で請求項 1 に記載のポリペプチドを少なくとも 1 つの試験化合物に結合させるステップと、

( b ) 請求項 1 に記載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 に記載の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ことを特徴とする方法。 10

【請求項 26】

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 に記載のポリペプチドの活性が許容された条件下で、請求項 1 に記載のポリペプチドを少なくとも 1 つの試験化合物に結合させる過程と、

( 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とを含み、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 27】

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

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

( b ) 前記標的ポリヌクレオチドの変異発現を検出する過程と、

( 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 28】

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 ) 処理した前記生体サンプルの核酸と、請求項 11 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含むプローブをハイブリダイズさせるステップであって、このハイブリダイゼーションゼーションが、前記プローブと前記生体サンプルの標的ポリヌクレオチドとの間で特異的なハイブリダイゼーション複合体が形成される条件下で行われ、前記標的ポリヌクレオチドが、請求項 11 のポリヌクレオチドのポリヌクレオチド配列またはその断片を含むポリヌクレオチドである、前記ステップと、

( c ) ハイブリダイゼーション複合体の収量を定量するステップと、 40

( d ) 前記処理された生物学的サンプル中の前記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前記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前記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ことを特徴とする方法。

【請求項 29】

生物学的サンプル中の S A T 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が形成されるのに適した条件下で、前記生物学的サンプルを請求項 10 に記載の抗体と結合する過程と、

(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30】

前記抗体が、

(a) キメラ抗体

(b) 単鎖抗体

(c) Fab断片

(d) F(ab')<sub>2</sub>断片

(e) ヒト化抗体のいずれかであることを特徴とする請求項10に記載の抗体。

【請求項31】

請求項10に記載の抗体と、許容できる賦形剤とを含む化合物。

【請求項32】

被検者のSAT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31に記載の化合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33】

前記抗体が標識されることを特徴とする請求項31に記載の化合物。

【請求項34】

被検者のSAT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33に記載の化合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35】

請求項10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、(a) 抗体反応を誘発する条件下で、SEQ ID NO: 1-9を有する群から選択したアミノ酸配列含むポリペプチドまたはその免疫抗原性断片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: 1-9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に特異結合するようなポリクローナル抗体を同定する過程とを含むことを特徴とする方法。

【請求項36】

請求項35に記載の方法で産出した抗体。

【請求項37】

請求項36に記載の抗体及び適切なキャリアを含む化合物。

【請求項38】

請求項10に記載の抗体の特異性を有するモノクローナル抗体を製造する方法であって、(a) 抗体反応を誘発する条件下で、SEQ ID NO: 1-9を有する群から選択したアミノ酸配列含むポリペプチドまたはその免疫抗原性断片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) 前記抗体産出細胞を不死化の細胞と融合して、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: 1-9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に特異結合するような前記培養モノクローナル抗体から単離する過程とを含むことを特徴とする方法。

【請求項39】

請求項38に記載の方法で産出したモノクローナル抗体。

【請求項40】

請求項39に記載の抗体及び適切なキャリアを含む化合物。

【請求項41】

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

10

20

30

40

50

項 10 に記載の抗体。

【請求項 42】

組換え免疫グロブリンライブラリのスクリーニングにより前記抗体を産出することを特徴とする請求項 10 に記載の抗体。

【請求項 43】

SEQ ID NO: 1 - 9 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を検出する方法であって、

(a) 前記抗体と前記ポリペプチドの特異結合を許容する条件下で、サンプルを用いて請求項 10 に記載の抗体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 - 9 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。

10

【請求項 44】

SEQ ID NO: 1 - 9 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を精製する方法であって、

(a) 前記抗体と前記ポリペプチドの特異結合を許容する条件下で、サンプルを用いて請求項 10 に記載の抗体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 - 9 を有する群から選択したアミノ酸配列を有する精製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

20

【請求項 45】

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

【請求項 46】

SEQ ID NO: 2 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 47】

配SEQ ID NO: 3 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 48】

SEQ ID NO: 4 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 49】

SEQ ID NO: 5 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

30

【請求項 50】

SEQ ID NO: 6 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 51】

SEQ ID NO: 7 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 52】

SEQ ID NO: 8 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 53】

SEQ ID NO: 9 のアミノ酸配列を含む請求項 1 に記載のポリペプチド。

【請求項 54】

SEQ ID NO: 10 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

40

【請求項 55】

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

【請求項 56】

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

【請求項 57】

SEQ ID NO: 13 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

50

## 【請求項 58】

SEQ ID NO : 14 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

## 【請求項 59】

SEQ ID NO : 15 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

## 【請求項 60】

SEQ ID NO : 16 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

## 【請求項 61】

SEQ ID NO : 17 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

## 【請求項 62】

SEQ ID NO : 18 のポリヌクレオチド配列を有する請求項 11 に記載のポリヌクレオチド。

## 【発明の詳細な説明】

## 【0001】

## (技術分野)

本発明は、分泌分子および輸送分子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小胞輸送障害、輸送障害、神経系疾患、自己免疫/炎症疾患、および細胞の異常増殖の診断・治療・予防に関する。本発明はさらに、分泌分子および輸送分子の核酸配列及びアミノ酸配列の発現における外来性化合物の効果についての評価に関する。本発明さらに、新規のヒト分泌及び輸送分子 (SAT) 及び SAT をコードするポリヌクレオチドの発見に基づき、これらの組成物を利用した小胞輸送障害、神経系疾患、自己免疫/炎症性の疾患、および細胞増殖異常の診断、治療、及び予防に関する。

## 【0002】

## (発明の背景)

真核細胞は脂質二分子膜細胞膜によって結合され、機能的に別個の、膜で結合された区画に細分される。膜は、細胞質、細胞外環境および細胞質内の各オルガネラの腔空間に本質的な差を維持する。膜内在性タンパク質、分泌タンパク質、およびオルガネラのルーメン用のタンパク質を含む真核タンパク質は、小胞体 (ER) 内で合成され、翻訳後の処理およびソートするためゴルジ複合体に送られ、次に特定の細胞内と細胞外の目標地に輸送される。物質は、エンドサイトーシスにより細胞外の環境から取り込まれる。これは神経シグナル、代謝シグナル、増殖シグナルの伝達にとって必須のプロセスである。また多くの必須の栄養素の取り込みを行う。さらに進入有機体に対し防御する。タンパク質分子のこの細胞内・細胞外の移動は小胞輸送と名付けられる。輸送は、供与体オルガネラ膜から開始し目標膜で融合する特殊小胞へタンパク質分子をパッケージングして遂行される (Rothman, J. E.、Wieland, F. T. (1996) Science 272: 2272-34)。

ER 膜を横切るタンパク質の輸送は、バクテリア、イースト、哺乳動物におけるプロセスと類似している (Gorlich, D. 他 (1992) Cell 71: 489-503 を参照)。哺乳類システムでは、輸送は細胞質内シグナル認識粒子 (SRP) の作用により開始する。SRP は、伸長している、合成されたばかりのポリペプチド上のシグナル配列を認識し、ER 膜上のシグナル認識粒子 (SRP) 受容体を経てポリペプチドとそのリボソーム複合体を ER 膜に結合させる。シグナルペプチドは切断され、付着しているポリペプチドと共にリボソーム複合体は膜に囲まれる。ポリペプチドは、その後 ER 膜を横切り、小胞へ転移させられる (Blobel, G. 及び B. Dobberstein (1975) J. Cell Biol. 67: 852-862)。

イースト中の ER 膜を横切ってポリペプチドの転置に係わされるタンパク質には、SEC61p、SEC62p および SEC63p がある。これらのタンパク質をコードする遺伝

10

20

30

40

50

子の突然変異は、移行プロセスに欠陥をもたらす。SEC61は、このタンパク質の遺伝子中にある突然変異が起こると多くのタンパク質の移行を阻害するため、特に重要である可能性がある(前出Gorlich)。

#### 【0003】

イーストSEC61(mSEC61)の哺乳類相同体はイヌおよびネズミの中で識別されている(前出Gorlich)。また、哺乳類SEC61はSECYp、バクテリアの細胞質膜移行タンパク質に構造が似ている。mSEC61は膜に囲まれたリボソームと強い結合をすることが分かっている。この結合は、膜を標的とする新生ポリペプチド鎖により引き起こされ、リボソームがそれらを構成するサブユニットに解離することによって弱められる。新生ポリペプチドはリボソームにより翻訳の終了後ER膜に転送され、mSEC61は、ER膜にある推定タンパク質伝導チャンネルの成分であると仮定される(前出Gorlich)。

10

#### 【0004】

分泌とエンドソーム経路に沿った物質の通過における数ステップは、輸送小胞の形成を必要とする。特に、小胞は、転移小胞体(tER)、ゴルジ槽の縁、トランスゴルジネットワーク(TGN)の表面、原形質膜(PM)およびエンドソームの管状伸張部で形成される。膜の一部領域が供与体オルガネラから離れるとき、小胞形成が起こる。膜で囲まれた小胞は、輸送されるタンパク質を含んでおり、タンパク質のコートにより取り囲まれる。それらの成分は細胞質から動員される。最初の分離プロセスとコーティング・プロセスは、細胞質のras様のGTP結合タンパク質、ADP-リボシル化要因(Arf)、およびアダプタタンパク質(AP)によりコントロールされる。小胞形成時に、細胞質内のGTPを結合したArfも、小胞に組み入れられる。ArfとAPの両方の異なるアイソフォームは異なる部位の小胞分離に関与する。例えば、Arf5、3および1はゴルジからの小胞分離に、Arf4はエンドソームからの小胞分離に、またArf6は原形質膜からの小胞分離に必要である。2種類のクラスのコートタンパク質も同定されている。クラスリンコートはTGNとPMに由来した小胞を形成し、他方コートマ(coatmer)(COP)のコートはERとゴルジから由来した小胞を形成する(Mellman, I. (1996) Annu. Rev. Cell Dev. Biol. 12: 575-625)。

20

#### 【0005】

クラスリンを基にした小胞形成では、APは分離しようとする膜の表面に小胞に入れる内容物とコートタンパク質を集積する。APは2つの大きな鎖から形成されたヘテロ四量体複合体である。一つの鎖は、a、g、d又はe鎖から成り、b鎖、中鎖(m)、小鎖(s)を伴う。クラスリンはb-アダプチンサブユニットのカルボキシ末端の付随領域を経てAPに結合する(LeBourgne, R., Hoflack, B. (1998) Curr. Opin. Cell Biol. 10: 499-503)。AP-1は、TGNとエンドソームからエンドソーム/リソソームシステムの区画へのタンパク質選別を行う。AP-2は原形質膜でクラスリンを媒介としたエンドサイトーシスで機能し、他方AP-3はエンドソーム及び/又はTGNに関係し、リソソームとリゾソーム関連オルガネラへの輸送のため膜内在性タンパク質を動員する。最近単離されたAP-4複合体は、TGNあるいは隣接区分に局在化し、ポスト-ゴルジ区分内で起こるとおもわれる選別イベントにおいて役割を果たしている可能性がある(Dell'Angelica, E. C. 他(1999) J. Biol. Chem. 274: 7278-7285 参照)。小胞の形成とともに、細胞質内のGTP結合Arfも、小胞に組み入れられる。別のGTP結合タンパク質(ダイナミン)は、形成しつつある小胞の頸部まわりにリング複合体を形成し、供与体膜から小胞の解放に必要な機械化学的の力を与える。つぎに、コーティングされた小胞複合体は細胞質内を通過して輸送される。輸送プロセス中に、Arfに結合したGTPはGDPに加水分解され、コートは輸送小胞から解離する(West, M. A. 他(1997) J. Mol. Biol. 138: 1239-1254 参照)。

30

40

50

## 【0006】

コートマ (COP) のコート (コートタンパク質の別のクラス) は、ER とゴルジに由来した小胞に形成する。COP コートはさらに COPI と COPII して分類され、COPI はゴルジを経て、ゴルジから ER まで逆行輸送に関与し、COPII は ER からゴルジまで順行輸送に関与する (前出 Mellman)。COP コートは GTP 結合タンパク質 (Arf または Sar) とコートプロトマ (コートマ) の 2 つの主成分から成る。コートマは 7 種類のタンパク質 ( - , - , ' - , - , - , - , - COP) の等モル複合体である。コートマ複合体は膜内在性タンパク質の細胞質テイル上に含まれるディリシン (dilysine) モチーフに結合する。それらのモチーフには、ER の膜タンパク質のディリシンを含んでいる検索 (retrieval)モチーフと、p24 ファミリーメンバーの二塩基ノジフェニルアミンモチーフがある。タイプ I 膜タンパク質の p24 ファミリーは、COPI 小胞の主要膜タンパク質である (Harter, C., Wieland, F.T. (1998) Proc. Natl. Acad. Sci. USA 95: 11649 - 11654 を参照)。

10

## 【0007】

小胞は同型融合 (同種の小胞との融合) あるいは異型融合 (異種の小胞との融合) をでき得る。適切な標的と小胞の融合に必要な分子には、小胞膜内のタンパク質、目標の膜、および細胞質から補充されたタンパク質が含まれる。供与体区分からの小胞の分離中に、複合内在性タンパク質である VAMP (小胞結合膜タンパク質) が小胞に組み入れられる。小胞の脱コートの直後に、細胞質内のプレニレートされた GTP 結合タンパク質 (Rab) は小胞膜に挿入される。Rab タンパク質のアミノ酸配列は、Ras スーパーファミリのメンバに特有の保存された GTP 結合領域を明らかにする。小胞膜では、GTP 結合 Rab は VAMP と相互作用する。小胞が目標の膜に達すると、目標の膜中の GTPase 活性化タンパク質 (GAP) は Rab タンパク質を GDP 結合形式に変換する。細胞質のタンパク質である、グアニンヌクレオチド解離阻害因子 (GDI) は小胞膜から GDP 結合 Rab を取り除く。Rab アイソフォームのいくつかは同定されており、細胞内の特定区分と結合しているらしい。例えば、Rab 4、5 および 11 は初期のエンドソームに関連し、他方 Rab 7 および 9 は後期エンドソームと関連する。これらの違いで、小胞とそれらの目標の膜の間の結合の選択性を提供する可能性がある (Novick, P., 及び Zerial, M. (1997) Cur. Opin. Cell Biol. 9: 496 - 504)。

20

30

輸送小胞の目標膜とのドッキングは、小胞 SNARE 受容体 (v-SNARE)、目標膜 (t-) SNARE および他の膜タンパク質と、細胞質のタンパク質間の複合体を伴う。ドッキング複合体のそれらの正確な機能は不確かなままであるが、これらの他のタンパク質の多くが同定された (Tellam, J.T. 他 (1995) J. Biol. Chem. 270: 58575863; Hata, Y., Sudhof, T.C. (1995) J. Biol. Chem. 270: 1302213028 を参照)。N-エチルマレイミド感受性要因 (NSF) と可溶性 NSF-アッタチメントタンパク質 (-SNAP と -SNAP) は、イーストからヒトまでで保存されており、ほとんどの細胞内膜融合反応で機能する、2 つのタンパク質である。Sec1 は、膜融合を含む分泌経路の種々の段階で機能するイーストタンパク質のファミリーを表わす。最近、Munc-18 タンパク質と呼ばれる Sec1 の哺乳類相同体が同定された (Katagiri, H. 他 (1995) J. Biol. Chem. 270: 49634966; Hata 他を参照)。

40

## 【0008】

SNARE 複合体には 3 種類の SNARE 分子があり、1 つは小胞膜に、2 つは目標膜にある。それらがいっしょになって、4 つの -ヘリックスのコイルドコイルの棒状複合体を形成する。3 つの SNARE すべての膜固着ドメインは棒の一端から突き出る。この複合体は、ミクソウイルス、インフルエンザ、フィロウイルス (エボラ)、HIV、SIV のレトロウイルスのような、エンベロープウイルスの特徴である、融合タンパク質によっ

50

て形成された棒状の構造に似ている (Skehel, J. J., Wiley, D. C. (1998) Cell 95: 871-874)。SNARE複合体は膜融合に十分であると提案されており、複合体と結合するタンパク質が融合イベントの制御を提供することを示唆している (Weber, T. 他 (1998) Cell 92: 759-772 参照)。例えば、制御されたエキソサイトーシスをするニューロンでは、ドッキングされた小胞は、カルシウムの流入を引き起こす脱分極が起こるまで、シナプス前膜と融合しない (Bennett, M. K., Scheller, R. H. (1994) Annu. Rev. Biochem. 63: 63-100)。シナプトタグミン (シナプス小胞中の膜内在性タンパク質) はドッキング複合体中の t-SNARE シンタキシンと結合する。シナプトタグミンは、マイナスに荷電したリン脂質の付いた複合体中でカルシウムと結合する。このマイナスに荷電したリン脂質によって、細胞質内 SNAP タンパク質はシンタキシンをシナプトタグミンに置き換え、融合を起こさせる。したがって、シナプトタグミンはニューロン中の融合に負の調節を行う (Littleton, J. T. 他 (1993) Cell 22: 817-1134 を参照)。

10

## 【0009】

シナプス小胞で最も豊富にある膜タンパク質は、糖タンパク質シナプトフィシン (4つの膜貫通ドメインと2つの小胞内ループを備えた38kDaタンパク質) であると考えられている。シナプトフィシンモノマーは、シナプス小胞膜中にチャンネルを形成するホモポリマー内に結合する。神経組織内のシナプトフィシンのカルシウム結合力、チロシン燐酸化、および広範な配分は、神経分泌において潜在的な役割をしていることを示唆する (Bennett, 前出)。

20

## 【0010】

タンパク質輸送の小胞内からの入出は、細胞膜と、スペクトリンとその他のタンパク質から成る膜細胞骨格の間の相互作用に依存する。アンキリンと呼ばれる関連タンパク質の大ファミリーは、膜骨格タンパク質スペクトリンと、バンド3と呼ばれる細胞膜中のタンパク質 (細胞膜中の陰イオンチャンネルの構成成分) に結合することにより、輸送プロセスに参加する。したがって、アンキリンは、細胞骨格と細胞膜間で重要なリンクとして機能する。

## 【0011】

アンキリンは、元々赤血球系細胞との関連で発見されたのであるが、他の組織にも存在する (Birkenmeier, C. S. 他 (1993) J. Biol. Chem. 268: 9533-9540 参照)。アンキリンは大きなタンパク質 (約1800個のアミノ酸) であり、これに含まれるものとして、N末端89kDaドメイン (細胞膜タンパク質バンド3およびチューブリンを結合する)、中枢部の62kDaドメイン (細胞骨格タンパク質スペクトリンとピメンチンを結合する)、C末端 (他の2つのドメインの結合活性力の修飾体として機能する55kDa調節ドメイン) がある。アンキリンの個々の遺伝子は、種々の挿入および削除により多数のアンキリンアイソフォームの生成が可能である。それらのアイソフォームはほとんど同一のサイズであるが、異なる機能がある場合がある。また、N末端 (バンド3結合) からのコード化配列の大きな領域と、中枢部 (スペクトリン結合) ドメインが欠落した、小さな転写物が生成される。アンキリンタンパク質のそのような大ファミリーの存在と、1種類以上のアンキリンが同一細胞タイプ内に発現されることのある観察は、アンキリンは膜骨格を原形質膜に単に結合するより多くの特殊機能を持っている場合があることを、示唆する (Birkenmeier, 前出)。

30

40

## 【0012】

ヒトでは、アンキリンの2つのアイソフォームが交互に、発達中の赤血球系と成熟赤血球系にそれぞれ発現される (Lambert, S. 他 (1990) Proc. Natl. Acad. Sci. USA 87: 1730-1734)。赤血球のスペクトリンおよびアンキリン中の欠乏は溶血性貧血、遺伝性球状赤血球症に関連していた (Coetzer, T. L. 他 (1988) New Engl. J. Med. 318: 23

50

0234 を参照)。

【0013】

タンパク質の正確な輸送には、上皮細胞の適切な機能が特に重要である。上皮細胞は、脂質と膜関連タンパク質のような異なる細胞膜成分を含んでいる、特有の表層側と基底側のドメインへ極性化される。タンパク質には柔軟なものがあり、細胞の種類または成長条件により、表層側が基底側に選別され得る。例えば、細胞が高密度に培養されている場合、腎臓陰イオン交換体 (kAE1) は、表層から基底ドメインへ標的を変えることが可能である。タンパク質カナダプチンは、kAE1の細胞質ドメインに結合するタンパク質として分離される。また、それは膜内ではなく、小胞中で kAE1 と共に局在化する。これは、カナダプチン機能が基底の標的膜に kAE1 包含小胞をガイドしていることを示唆している (Chen, J. 他 (1998) J. Biol. Chem. 273:1038-1043 参照)。

【0014】

小胞輸送は神経伝達の過程で重要である。シナプス小胞は、ニューロンの細胞質からシナプスまで神経伝達物質分子を搬送する。Rab3 はシナプス小胞上にある GTP 結合タンパク質のファミリーである。タンパク質の RIMファミリーは、Rab3 のためのエフェクターであると思われる (Wang, Y. 他 (2000) J. Biol. Chem. 275:20033-20044 参照)。Rabphilin-3 はシナプス小胞タンパク質である。グラニューフィリン (granophilin) はラプフィリン (rabphilin) と相同性のあるタンパク質であり、エキソサイトーシスにおいて独特の役割をもつことがある (Wang, J. 他 (1999) J. Biol. Chem. 274:28542-28548 参照)。

【0015】

無数のヒトの疾患及び障害の病因は、タンパク質のオルガネラまたは細胞膜への輸送における欠陥に起因し得る。膜結合受容体およびイオンチャネルの輸送の欠陥は、嚢胞性繊維症 (嚢胞性繊維症の膜内外のコンダクタンス調節因子、CFTR)、グルコース-ガラクトース吸収不良症候群 (Na<sup>+</sup>/グルコース共同輸送体)、コレステロール過剰血 (低密度リポタンパク質 (LDL) 受容体)、および糖尿病 (インシュリン受容体) の形式に関連がある。異常なホルモンの分泌は、尿崩症 (バソプレッシン)、高血糖症および低血糖症 (インシュリン、グルカゴン)、グレーヴス疾病および甲状腺腫 (甲状腺ホルモン)、およびクッシング疾病およびアディソン疾病 (副腎皮質刺激ホルモン; ACTH)、などの障害に起因している。

【0016】

癌細胞は過量のホルモンあるいは他の生物活性ペプチドを分泌する。腫瘍細胞により生物活性ペプチドの過度の分泌と関係する障害には次のものがある。インスリノーマアイレ (insulinoma islet) 細胞腫瘍からの高インシュリンに起因する絶食性低血糖症; 副腎髄質および感神経性パラングリア (paraganglia) のクロム親和性細胞腫 (pheochromocytomas) から分泌されるエピネフリンおよびノルエピネフリン増大に起因する高血圧症; 癌腫のような症候群 (腹部の激痛、下痢および心臓弁膜症を含み、腸の腫瘍から分泌された過量の血管作用物質 (セロトニン、ブラジキニン、ヒスタミン、プロスタグランジンおよびポリペプチドホルモン) に起因する)。生物学上活発なペプチド (腫瘍から期待されないペプチド) の転位の合成および分泌には、肺と膵臓癌に ACTH とバソプレッシン、肺と膀胱癌の中の副甲状腺ホルモン、肺と乳癌の中のカルシトニン、骨髄の甲状腺癌中の乾燥甲状腺刺激ホルモンが含まれる。ヒトの種々の病原体は宿主細胞タンパク質を自己の利点のため輸送経路を変更する。例えば、HIV タンパク質 Nef は、クラスリン コーティングされたピットを介するエンドサイトーシスを加速することにより CD4 分子の細胞表面発現を下方調節する。Nef のこの機能は感染細胞からの HIV の拡散にとって重要である (Harris, M. (1999) Curr. Biol. 9:R449-R461)。ヒトの最近識別されたタンパク質に、Nef 結合因子 1 (Nef1) があり、これは4つの伸長したコイルドコイ

ルのドメインを有するタンパク質であり、Nefと結合することが発見された。Nef1の過剰発現は、Nefによって抑制され得るCD4の細胞表面発現を増加させる。(Fukushi, M. 他(1999) FEBS Lett. 442: 83-88)。  
 新たな分泌分子及び輸送分子及びこれらをコードするポリヌクレオチドの発見は、小胞輸送障害、輸送障害、神経障害、自己免疫/炎症疾患、および細胞増殖異常の診断、治療並びに予防において、また、分泌分子及び輸送分子の核酸及びアミノ酸配列の発現に対する外因性化合物の効果の評価において有用であるような新たな組成物を提供することにより、当分野における必要性を満たす。

## 【0017】

(発明の概要)

本発明の特徴は、精製ポリペプチド、および分泌・輸送分子にあり、これは集合的に「SAT」、個別には「SAT-1」、「SAT-2」、「SAT-3」、「SAT-4」、「SAT-5」、「SAT-6」、「SAT-7」、「SAT-8」、「SAT-9」と呼ばれている。或る実施態様において本発明は、(a)SEQ ID NO: 1-9を有す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: 1-9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有するポリペプチド、(c)SEQ ID NO: 1-9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: 1-9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抗原性断片を含む群から選択した実質上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: 1-9のアミノ酸配列を含む実質上単離されたポリペプチドを提供する。

10

20

## 【0018】

また、本発明は(a)SEQ ID NO: 1-9を有す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: 1-9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有するポリペプチド、(c)SEQ ID NO: 1-9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: 1-9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抗原性断片、を含む群から選択されたポリペプチドをコードする実質上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SEQ ID NO: 1-9を有する群から選択した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: 10-18を有する群から選択される。

30

## 【0019】

本発明は更に、(a)SEQ ID NO: 1-9を有す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: 1-9を有する群から選択したアミノ酸配列と少なくとも90%の相同性の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: 1-9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: 1-9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抗原性断片を有す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結合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。

40

## 【0020】

更に本発明は、(a)SEQ ID NO: 1-9からなる一群から選択されたアミノ酸配列からなるポリペプチドと、(b)SEQ ID NO: 1-9からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列を有するポリペプチドと、(c)SEQ ID NO: 1-9からなる一群から選択されたアミノ酸配列を有するポリペプチドの生物学的に活性な断片と、(d)SEQ ID NO: 1-9とからなる一群から選択されたアミノ酸配列を有するポリペプチドの免疫抗原性断片とで構成される

50

一群から選択され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受容する過程とを有し、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結合したプロモーター配列を有する。

**【0021】**

本発明は更に、(a)SEQ ID NO: 1 - 9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、(b)SEQ ID NO: 1 - 9からなる群から選択したアミノ酸配列と少なくとも90%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 - 9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 - 9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抗原性断片から構成される群から選択されたポリペプチドに特異結合するような実質上単離された抗体を提供する。

10

**【0022】**

本発明は更に、(a)SEQ ID NO: 10 - 18からなる群から選択したポリヌクレオチド配列、(b)SEQ ID NO: 10 - 18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)に相補的なポリヌクレオチド配列、(d)(b)に相補的なポリヌクレオチド配列、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。

20

**【0023】**

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: 10 - 18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: 10 - 18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)に相補的なポリヌクレオチド、(d)(b)に相補的なポリヌクレオチド配列、および(e)(a)~(d)のRNA等価物からなる群から選択されたポリヌクレオチド配列を含む。検出方法は、(a)サンプル中の標的ポリヌクレオチドに相補的な配列を含む少なくとも20の連続したヌクレオチドを含む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)ハイブリダイゼーション複合体の存在・不存在を検出し、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からなり、プローブと標的ポリヌクレオチドあるいはその断片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、プローブは標的ポリヌクレオチドに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。

30

**【0024】**

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: 10 - 18を有する群から選択したポリヌクレオチド配列からなるポリヌクレオチド、(b)SEQ ID NO: 10 - 18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%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、(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)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)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。

40

**【0025】**

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成分を提供し

50

、有効量のポリペプチドは、(a) SEQ ID NO: 1 - 9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 - 9を有する群から選択したアミノ酸配列と少なくとも90%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 - 9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 - 9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抗原性断片を含む。一実施態様では、成分は配列番号1乃至6を有する群から選択したアミノ酸配列を含む。更に本発明は、機能性SAT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成分を投与する過程を有する方法を提供する。

**【0026】**

本発明はまた、(a) SEQ ID NO: 1 - 9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 - 9を有する群から選択したアミノ酸配列と少なくとも90%の相同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 - 9を有する群から選択したアミノ酸配列のポリペプチドの生物学的活性断片、または(d) SEQ ID NO: 1 - 9を有する群から選択したアミノ酸配列のポリペプチドの免疫抗原性断片を含む群から選択されたアゴニストとしてのポリペプチド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) ポリペプチドを有する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。一実施態様では、本発明は機能性SAT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成分を投与する過程を含む方法を提供する。

10

20

30

**【0027】**

本発明は更に、(a) SEQ ID NO: 1 - 9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 - 9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有するポリペプチド、(c) SEQ ID NO: 1 - 9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 - 9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抗原性断片を含む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す過程と、(b) サンプル中のア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成分を提供する。別の実施態様では、機能性SATの過剰発現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成分を投与する過程を含む方法を提供する。

**【0028】**

本発明は更に、(a) SEQ ID NO: 1 - 9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 - 9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 - 9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 - 9からなる群から選択したアミノ酸配列の免疫抗原性断片を含む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つの試験化合物に結合させる過程と、(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。

40

**【0029】**

この方法は、(a) このポリペプチドを好適な条件下で少なくとも1つの化合物と結合させるステップと、(b) このポリペプチドとこの試験化合物との結合を検出して、このポ

50

リペプチドと特異的に結合する化合物を同定するステップとを含む。

( a ) S E Q I D N O : 1 - 9 を有する群から選択したアミノ酸配列、( b ) S E Q I D N O : 1 - 9 を有する群から選択したアミノ酸配列と少なくとも90%の相同性を有する天然のアミノ酸配列を含むポリペプチド、( c ) S E Q I D N O : 1 - 9 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 ) S E Q I D N O : 1 - 9 を有する群から選択したアミノ酸配列の免疫抗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。

10

#### 【0030】

更に本発明は、S E Q I D N O : 10 - 18 からなる一群から選択された配列を含む標的ポリヌクレオチドの発現を変化させるのに効果的な化合物をスクリーニングする方法であって、( a ) この標的ポリヌクレオチドを含むサンプルを化合物に曝露するステップと、( b ) この標的ポリヌクレオチドの発現の変化を検出するステップとを含む、該スクリーニング方法を提供する。

#### 【0031】

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 ) S E Q I D N O : 10 - 18 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、( i i ) S E Q I D N O : 10 - 18 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、( i i i ) ( i ) に相補的な配列を有するポリヌクレオチド、( i v ) ( i i ) のポリヌクレオチドに相補的なポリヌクレオチド、( v ) ( i ) ~ ( i v ) のRNA等価物からなる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を含むプローブを用いて、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 ) S E Q I D N O : 10 - 18 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、( i i ) S E Q I D N O : 10 - 18 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、( i i i ) ( i ) のポリヌクレオチドに相補的な配列を有するポリヌクレオチド、( i v ) ( i i ) 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( v ) ( i ) ~ ( i v ) のRNA等価物からなる群から選択する。或いは、標的ポリヌクレオチドは、上記( i ) ~ ( v ) からなる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。

20

30

40

#### 【0032】

( 発明の実施の形態 )

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ものではないことも併せて理解されたい。

#### 【0033】

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、

50

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の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」と記されている場合にはそのような宿主細胞が複数あることもあり、「或る抗体」と記されている場合には単数または複数の抗体、及び、当業者に公知の抗体の等価物等についても言及しているのである。

【0034】

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。

10

【0035】

(定義)

「SAT」は、実質上精製されたSATのアミノ酸配列であって、任意の種、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の種から得たもので、任意の天然物、合成物、半合成物或いは組換え物を起源とするものを指す。

【0036】

「アゴニスト」の語は、SATの生物学的活性を強化または擬態する分子を指す。このアゴニストは、SATに直接相互作用するか、或いはSATが関与する生物学的経路の成分と作用して、SAT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。

20

【0037】

用語「対立遺伝子変異配列」は、SATをコードする遺伝子の別の形を指す。対立遺伝子変異体は、核酸配列における少なくとも1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からも作製し得る。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。

30

【0038】

SATをコードする「変異」核酸配列は、様々なヌクレオチドの欠失、挿入、或いは置換が起こっても、SATと同じポリペプチド或いはSATの機能特性の少なくとも1つを備えるポリペプチドを指す。この定義には、SATをコードするポリヌクレオチド配列の正常な染色体の遺伝子座ではない位置でのアレル変異配列との不適當或いは予期しないハイブリダイゼーション、並びにSATをコードするポリヌクレオチドの特定のオリゴヌクレオチドプローブを用いて容易に検出可能な或いは検出困難な多形性を含む。コードされたタンパク質も「変異」し得るものであり、サイレント変化を生ぜしめて結果的に機能的に等価なSATとなるようなアミノ酸残基の欠失、挿入または置換を含み得る。意図的なアミノ酸置換は、生物学的或いは免疫学的にSAT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、セリンとスレオニンがある。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。

40

【0039】

「アミノ酸」または「アミノ酸配列」の語は、オリゴペプチド、ペプチド、ポリペプチド

50

若しくはタンパク質の配列またはその断片を指し、天然または合成分子を指す。ここで、「アミノ酸配列」は天然のタンパク質分子のアミノ酸配列を指すものであり、「アミノ酸配列」及び類似の語は、アミノ酸配列を、列挙したタンパク質分子に関連する完全な本来のアミノ酸配列に限定しようとするものではない。

## 【0040】

「増幅」は、核酸配列の追加複製に関連する。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。

## 【0041】

用語「アンタゴニスト」は、SATの生物学的活性を阻害或いは減弱する分子である。アンタゴニストは、SATに直接相互作用するか、或いはSATが関与する生物学的経路の成分と作用して、SAT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。

10

## 【0042】

用語「抗体」は、エピトープの決定基と結合することができる、そのままの免疫グロブリンやその断片、例えばFab、F(ab')<sub>2</sub>及びFv断片を指す。SATポリペプチドと結合する抗体は、免疫抗原として、そのままのポリペプチド、または関心のある小ペプチドを含む断片を用いて作製可能である。動物（マウス、ラット、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、好みに応じて担体タンパク質に抱合することも可能である。通常用いられる担体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カスイガイのヘモシアニン（KLH）等がある。結合ペプチドは、動物を免疫化するために用いる。

20

## 【0043】

「抗原決定基」の語は、特定の抗体と接触している分子の領域（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、抗体への結合において無損傷抗原（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。

## 【0044】

「アンチセンス」の語は、特定の核酸配列の「センス」（コーディング）鎖と塩基対を形成することが可能な任意の成分を指す。アンチセンス成分には、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連鎖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含まれる。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」若しくは「マイナス（-）」の語が参照DNA分子のアンチセンス鎖を、「正」若しくは「プラス（+）」がセンス鎖を指すことがある。

30

40

## 【0045】

「生物学的に活性」の語は、天然分子の構造的機能、調節機能または生化学的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SAT、合成のSAT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。

## 【0046】

「相補（的）」または「相補性」の語は、塩基対形成によってアニーリングする2つの一本鎖核酸の間の関係を指す。例えば、配列「5' A - G - T 3'」は、相補配列「3' T - C - A 5'」と対を形成する。

50

## 【0047】

「所定のポリヌクレオチド配列を含む成分」及び「所定のアミノ酸配列を含む成分」は、所定のポリヌクレオチド配列またはアミノ酸配列を含む広範囲の任意の成分を指す。この成分には、乾燥製剤または水溶液が含まれ得る。SAT 若しくはSAT 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）及びその他の構成エレメント（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。

10

## 【0048】

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

## 【0049】

「保存的なアミノ酸置換」は、置換がなされた時に元のタンパク質の特性を殆ど損なわないと予測されるような置換、即ちタンパク質の構造と特に機能が保存され、そのような置換による大きな変化がない置換を指す。下表は、タンパク質中で元のアミノ酸と置換可能で、保存アミノ酸置換と認められるアミノ酸を示している。

20

元残基	保的置換	
Ala	Gly, Ser	
Arg	His, Lys	
Asn	Asp, Gln, His	
Asp	Asn, Glu	
Gly	Ala, Ser	10
Gln	Asn, Glu, His	
Glu	Asp, Gln, His	
Gly	Ala	
His	Asn, Arg, Gln, Glu	
Ile	Leu, Val	
Leu	Ile, Val	20
Lys	Arg, Gln, Glu	
Met	Leu, Ile	
Phe	His, Met, Leu, Trp, Tyr	
Ser	Gly, Thr	
Thr	Ser, Val	
Trp	Phe, Tyr	30
Tyr	His, Phe, Trp	
Val	Ile, Leu, Thr	

## 【 0 0 5 0 】

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 ヘリックス構造、( 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。

## 【 0 0 5 1 】

「欠失」は、結果的に 1 個若しくは数個のアミノ酸またはヌクレオチドが失われてなくなるようなアミノ酸またはヌクレオチド配列における変化を指す。

## 【 0 0 5 2 】

「誘導体」の語は、ポリペプチド配列またはポリヌクレオチド配列の化学修飾を指す。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 ( p e g y l a t i o n )、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から少なくとも 1 つの生物学的若しくは免疫学的機能を保持しているプロセスによって、修飾されたポリペプチ

ドである。

【0053】

「検出可能な標識」は、測定可能な信号を生成することができ、ポリヌクレオチドまたはポリペプチドに共有結合または非共有結合するようなレポーター分子または酵素を指す。

【0054】

「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、または非調節、あるいは減少、下方調節、または欠損遺伝子またはタンパク発現を指す。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。

【0055】

用語「断片」は、S A TまたはS A Tをコードするポリヌクレオチドの固有の部分であって、その親配列 ( p a r e n t s e q u e n c e ) と同一であるがその配列より長さが短いものを指す。断片は、画定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、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長さであり得る。断片は、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最初の250または500アミノ酸（またはポリペプチドの最初の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。

10

20

【0056】

SEQ ID NO : 10 - 18の断片は、例えば、この断片を得たゲノム内の他の配列とは異なる、SEQ ID NO : 10 - 18を明確に同定する固有のポリヌクレオチド配列の領域を含む。SEQ ID NO : 10 - 18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 : 10 - 18を関連ポリヌクレオチド配列から区別する類似の方法に有用である。ある断片と一致するSEQ ID NO : 10 - 18の正確な断片の長さや領域は、その断片の目的に基づいて当分野で一般的な技術によって日常的に測定できる。

30

【0057】

SEQ ID NO : 1 - 9の断片は、SEQ ID NO : 10 - 18の断片によってコードされる。SEQ ID NO : 1 - 9の断片には、SEQ ID NO : 1 - 9を特異的に同定する固有のアミノ酸配列領域が含まれている。例えば、SEQ ID NO : 1 - 9の断片は、SEQ ID NO : 1 - 9を特異認識する抗体を産出するための免疫抗原性ペプチドとして有用である。SEQ ID NO : 1 - 9の断片及び断片に対応するSEQ ID NO : 1 - 9の領域の正確な長さは、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

【0058】

「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。

【0059】

「相同性」の語は、配列類似性即ち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間で互換可能な配列同一性である。

【0060】

ポリヌクレオチド配列に適用される「一致率」または「一致%」の語は、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た少なくとも2つ以上のポリヌクレオチド配列間で一致する残基の割合を意味する。このようなアルゴリズムは、2配列間のアラインメン

50

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。

#### 【0061】

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. 及び P.M. Sharp (1989) CABIOS 5:151-153、Higgins, D.G. 他(1992) CABIOS 8:189-191に記載されている。ポリペプチド配列を2つ1組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。残基「重み付け」表をデフォルトとして選択する。CLUSTAL Vは、アラインメントされたポリヌクレオチド配列対間の「類似率」として一致率を報告する。

#### 【0062】

或いは、米国国立バイオテクノロジー情報センター(NCBI)のBasic Local Alignment Search Tool (BLAST)が一般的に用いられ、且つ、無料で入手可能な配列比較アルゴリズム一式を提供している(Altschul, S.F. (1990) J. Mol. Biol. 215:403-410)。このアルゴリズムは、幾つかの情報源から入手可能であり、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/>)からも入手可能である。BLASTソフトウェア一式には様々な配列分析プログラムが含まれており、既知のポリヌクレオチド配列を種々のデータベースから得た別のポリヌクレオチド配列とアラインメントする「blastn」もその1つである。その他にも、2つのヌクレオチド配列を対で直接比較するために用いる「BLAST 2 Sequences」と称されるツールも利用可能である。「BLAST 2 Sequences」は、<http://www.ncbi.nlm.nih.gov/gorf/bl2.html>にアクセスして対話形式で利用することが可能である。「BLAST 2 Sequences」ツールは、blastnとblastp(後述)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。

#### 【0063】

```

MatrixHCSLM2
Rewardfrmatch1
Penaltyfrmismatch-2
GapOp5及ExtensionCapペナルティ2
Gapxdropff:50
Expect 10
WordSize 11
Filteron

```

一致率は、完全に画定された(例えば特定の配列番号で画定された)配列長さと比較して測定し得る。或いは、より短い長さ、例えばより大きな画定された配列から得られた断片(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。

【0064】

高度の相同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。

【0065】

ポリペプチド配列に適用される「一致率」または「一致%」の語は、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た少なくとも2以上のポリペプチド配列間で一致する残基の割合を意味する。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、置換部位の電荷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。

【0066】

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2つ1組で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と設定する。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。

【0067】

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対で比較をする場合、ある者は、デフォルトパラメータで設定された「BLAST 2 Sequences」ツールVersion 2.0.12 (Apr - 21 - 2000)でblastpを使用するであろう。デフォルトパラメータの設定例を以下に示す。

【0068】

```
Matrix: BLOSUM62
Gap: 11 及び Extension Gap: ペナルティ 1
Gap: 50
Expect: 10
WordSize: 3
Filter: on
```

一致率は、完全に画定された（例えば特定の配列番号で画定された）ポリペプチド配列の長さと比較して測定し得る。一致率は、配列或いは、より短い長さ、例えばより大きな画定された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。

【0069】

「ヒト人工染色体（HAC）」は、約6 kb（キロベース）～10 MbのサイズのDNA配列を含み得る、染色体の複製、分離及び維持に必要な全てのエレメントを含む直鎖状の小染色体である。

## 【0070】

「ヒト化抗体」の語は、非抗体結合領域におけるアミノ酸配列はヒト抗体により近づくように変異させた抗体分子であって、本来の結合能力はそのまま保持しているような抗体分子を指す。

## 【0071】

「ハイブリダイゼーション」は、所定のハイブリダイゼーション条件下で塩基対を形成することによって、一本鎖ポリヌクレオチドが相補的鎖とアニーリングするプロセスを指す。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同性を共有することを示すものである。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する。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100μg/mlのせん断して変性したサケ精子DNAが含まれる。

10

## 【0072】

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における特定の配列の融点（ $T_m$ ）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ら（1989）Molecular Cloning: 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

## 【0073】

本発明のポリヌクレオチド間の高いストリンジェンシーのハイブリダイゼーションでは、約0.2×SSC及び約1%のSDSの存在の下、約68℃で1時間の洗浄過程を含む。Alternatively, temperatures of about 65, 60, 55, or 42℃ may be used. 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。このようなブロッキング剤には、例えば、約100～200μg/mlの変性サケ精子DNAがある。特定条件下で、例えばRNAとDNAのハイブリダイゼーションに有機溶剤、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

40

## 【0074】

「ハイブリダイゼーション複合体」の語は、相補的塩基対間の水素結合の形成力によって2つの核酸配列間に形成された複合体を指す。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<sub>0</sub>tまたはR<sub>0</sub>t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、ピンまたはガラスライド、或いは他の適切な基質であって細胞若しくはその核酸が固定される基質）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。

## 【0075】

「挿入」及び「付加」の語は、1個若しくは数個のアミノ酸残基またはヌクレオチド配列

50

を各々付加するようなアミノ酸またはヌクレオチド配列における変化を指す。

【0076】

「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。

【0077】

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SAT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SATのポリペプチド断片またはオリゴペプチド断片を含む。

10

【0078】

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

【0079】

「エレメント」または「アレイエレメント」の語は、マイクロアレイ上に定義された固有の位置にあるハイブリダイズ可能なポリヌクレオチド、ポリペプチドその他の化合物を指す。

【0080】

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

20

【0081】

「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。

【0082】

「機能的に結合した」は、第1核酸配列が第2核酸配列と機能的な関係があるように配置された状態を指す。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結合している。機能的に結合したDNA配列は非常に近接するか、或いは連続し得る。そして、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は、同一のリーディングフレーム内にある。

30

【0083】

「ペプチド核酸」(PNA)は、アンチセンス分子または抗遺伝子物質であって、リジン末端とするアミノ酸残基のペプチド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からなるものを指す。末端のリジンは、成分に溶解性を与える。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拡張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。

40

【0084】

SAT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、SATの酵素環境に依存し、細胞の種類によって異なり得る。

【0085】

「プローブ」は、SAT、SATの相補配列またはこれらの断片をコードする核酸配列を指し、同一核酸配列、対立遺伝子核酸配列または関連する核酸配列の検出に用いられる。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

50

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」は、短い核酸、通常はDNAオリゴヌクレオチドであり、相補的塩基対を形成することで標的ポリヌクレオチドにアニーリングされ得る。プライマーは次に、DNAポリメラーゼ酵素によって標的DNA鎖に延在し得る。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。

#### 【0086】

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。

#### 【0087】

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. ら(1989) Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NY、Ausubel, F.M. ら、(1987) Current Protocols in Molecular Biology, Greene Publi. Assoc. & Wiley-Intersciences, New York NY、Innis ら(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。

#### 【0088】

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、オリゴヌクレオチド及び最大5,000までの大きめのポリヌクレオチドであって32キロボースまでのインプットポリヌクレオチド配列から得たもの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(マサチューセッツ州ケンブリッジのWhitehead Institute/MITゲノム研究センターから一般向けに入手可能)ではユーザーが「ミスプライミング・ライブラリ」をインプットすることができ、ここでプライマー結合部位として避けたい配列はユーザーが指定する。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、各自のソースから得てユーザー固有のニーズを満たすように変更してもよい。)PrimerGenプログラム(英国ケンブリッジ市の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、固有な、および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。

【0089】

「組換え核酸」は天然配列ではない配列であるか或いは人為的に組み合わせなければ離隔しているような配列の2以上のセグメントを人為的に組み合わせで産出した配列を有する配列である。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、例えばのSambrookらの文献(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。組換え核酸の語は、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結合した核酸配列が含まれる。このような組換え核酸は、ベクターの不可欠なエレメントであって例えばある細胞を形質転換するために用いられるようなものであり得る。

10

【0090】

或いはこのような組換え核酸は、ウイルスベクターの不可欠なエレメント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。ワクシニアウイルスは哺乳動物のワクチン接種に用いることが可能で、その際に組換え核酸が発現して哺乳動物の防御免疫応答を誘導する。

【0091】

「調節エレメント」は、通常は遺伝子の未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'の未翻訳領域(UTR)を含む。調節エレメントは、転写、翻訳またはRNA安定性を調節する宿主またはウイルスタンパク質と相互作用する。

20

【0092】

「レポーター分子」とは、核酸、アミノ酸または抗体を標識するのに用いられる化学的または生化学的成分である。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。

【0093】

DNA配列に関する「RNA等価物」は、窒素塩基チミンが全てウラシルに置換されていることと、糖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構成されていることを除いて、参照DNA配列と同一のヌクレオチド線形配列から構成されている。

30

【0094】

「サンプル」の語は、その最も広い意味で用いられる。SAT、SAT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。

【0095】

「特異結合」または「特異的に結合する」の語は、タンパク質またはペプチドと、アゴニスト、抗体、アンタゴニスト、小分子、任意の天然成分または合成結合成分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。例えば、抗体がエピトープ「A」に対して特異的である場合、標識された遊離したA及びその抗体を含む反応において、エピトープA(つまり遊離し、標識されていないA)を含むポリペプチドの存在が、抗体に結合している標識されたAの量を低減させる。

40

【0096】

「実質上精製された」の語は、自然環境から取り除かれ、或いは単離または分離された核酸またはアミノ酸配列であって、自然に会合するその他の構成エレメントの少なくとも約60%、好ましくは少なくとも約75%、最も好ましいのは少なくとも約90%が遊離し

50

ているものを指す。

【0097】

「置換」は、1個若しくは数個のアミノ酸またはヌクレオチドを各々別のアミノ酸またはヌクレオチドに置換することを意味する。

【0098】

「基質」は、任意の好適な固体または半固体の支持体を指すものであって、膜、フィルタ、チップ、スライド、ウェハ、ファイバー、磁性または非磁性ビーズ、ゲル、管、プレート、ポリマー、微細粒子、毛管が含まれる。基質は、凹み、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、基質表面にはポリヌクレオチドやポリペプチドが結合する。

10

【0099】

「転写イメージ」は、所与の時間、条件での固有の細胞タイプまたは組織による遺伝子発現の集合的パターンを指す。

【0100】

「形質転換」は、外来性のDNAが受入細胞に入り込むプロセスを表す。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またはウイルス感染、電気穿孔法（エレクトロポレーション）、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、限られた時間に一時的に導入DNA若しくは導入RNAを発現する細胞も含まれる。

20

【0101】

ここで用いる「遺伝形質転換体」とは任意の有機体であり、限定するものではないが動植物を含み、有機体の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 (1989) の参考文献に記載されている。

30

【0102】

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40%の相同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%またはそれ以上の相同性を示し得る。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「多形性」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相同性を有し得るが、mRNAプロセッシング中のエキソンの交互スプライシングによって通常多数の或いは僅かな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、種相互に異なるポリヌクレオチド配

40

50

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相同性を有する。多形性変異体は、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が異なる。また、多形性変異体は、1つのヌクレオチド塩基によってポリヌクレオチド配列が変化する「1塩基ヌクレオチド多形性」(SNP)を含み得る。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。

#### 【0103】

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40%の相同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  
10  
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。

#### 【0104】

(発明)

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  
20  
、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ポリヌクレオチド配列番号(IncyteポリヌクレオチドID)によって表示した。

#### 【0105】

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と2は発明したポリペプチドをポリペプチド配列識別番号(ポリペプチドSEQ ID NO)と対応する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。列3は、GenBankの最も近い相同体のGenBankの識別番号(Genbank ID NO)を示す。列4は、各ポリペプチドとそのGenBank相同体との間の一致を表す確率スコアを示す。列5は、GenBank相同体のアノテーションを示し、更に該当箇所には適当な引用文も示す。これらを引用することによって本明細書の一部とする。  
30

#### 【0106】

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と2は発明した個々のポリペプチドをポリペプチド配列識別番号(SEQ ID NO)と対応する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(signature)配列、ドメイン、およびモチーフを含むアミノ酸残基を示す。列7は、タンパク質の構造/機能の分析のための分析方法を示し、該当箇所にはさらに分析方法に利用した検索可能なデータベースを示す。  
40

#### 【0107】

表2及び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、それら特性が請求の範囲に記載されたポリペプチドが分泌分子及び輸送分子であることを確立している。例えば、SEQ ID NO: 2は、Basic Local Alignment Search Tool (BLAST)により決定されるように、シナプトフィジンファミリーのメンバーであるmitsugumin29(GenBank ID g3077703)と93%が同一である。(表2参照)BLAST確率スコアは $2.9e^{-136}$ 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  
50

る。SEQ ID NO: 2 はまた、シナプトフィジン/シナプトポリン 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 及び PRODOM データベースのタンパク質シグネチャ配列への BLAST 比較そして BLIMPS 及び PROFILESCAN 分析から得たデータは、SEQ ID NO: 2 がシナプトフィジンのファミリーメンバであることを裏づける証拠を更に提供する。SEQ ID NO: 3 は、BLAST 確率スコアが 0.0 で、ラットの表層エンドソーム糖タンパク質 (GenBank ID g777776) と 72% が同一である。PRODOM データベースに対する BLAST 分析からのデータは、さらに SEQ ID NO: 3 が尖端エンドソーム糖タンパク質である確証的な証拠を示している。SEQ ID NO: 8 は、BLAST 確率スコアが 0.0 で、ドブネズミ (*Rattus norvegicus*) のシナプトタグミン III (GenBank ID g484296) と 95% が同一である。SEQ ID NO: 8 はまた、C2 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) BLIMPS、MOTIFS、及び PROFILESCAN 解析よりのデータは、SEQ ID NO: 8 がある C2 ドメインを含むタンパク質 (おそらくシナプトタグミン (synaptotagmin) ファミリーのメンバー) である、さらに実証的な証拠を提供する。SEQ ID NO: 1、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、及び SEQ ID NO: 9 については、同様の方法で分析し、注釈を付けた。SEQ ID NO: 1 - 9 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。

#### 【0108】

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ることで構築した。列 1 と 2 は本発明のポリヌクレオチドをポリヌクレオチド配列識別番号 (Polynucleotide SEQ ID NO) と対応する 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) を示す。列 3 は、各ポリヌクレオチド配列の長さ (塩基対単位) を示す。列 4 は、例えば、SEQ ID NO: 10 - 18 を同定するため、或いは SEQ ID NO: 10 - 18 と関連するポリヌクレオチド配列とを区別するためのハイブリダイゼーションまたは増幅技術に有用なポリヌクレオチド配列の断片を示す。列 5 は cDNA 配列、ゲノム DNA から予想されたコード配列 (エキソン) 及び / または cDNA 及びゲノム DNA を共に有する配列集合に対応する識別番号を示している。これらの配列は、本発明の完全長ポリヌクレオチド配列を構築するのに用いた。表 4 の列 6 および列 7 はそれぞれ、列 5 の配列に対応する cDNA 配列およびゲノム配列の開始ヌクレオチド (5') 位置および終了ヌクレオチド (3') 位置を示す。

#### 【0109】

表 4 の列 5 の識別番号は、特に例えば Incyte cDNA とそれに対応する cDNA ライブラリを指す場合もある。例えば、1438701F1 は Incyte cDNA 配列の識別番号であり、PANCNOT02 はそれが由来する cDNA ライブラリの識別番号である。cDNA ライブラリが示されていないインサイト cDNA は、プールされている cDNA ライブラリ (例えば、70767606V1) に由来する。または、列 5 の識別番号は、ポリヌクレオチド配列の組み立てに用いた GenBank の cDNA すなわち EST (例えば、g5810426) の識別番号の場合もある。さらに、列 5 の識別番号は、ENSEMBL (イギリス Cambridge の The Sanger Centre) データベース (「ENST」を含む配列) に由来した配列を識別できる場合もある。或いは、列 5 の識別番号は、NCBI RefSeq ヌクレオチド・配列・レコード・データベース (「NM」あるいは「NT」を含む配列) または NCBI RefSeq 蛋白質配列・レコード (「NP」を含む配列) に由来する場合もある。または列 5 の識別番

号は、「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の集合を指す場合もある。例えば、FL\_XXXXXX\_N1\_N2\_YYYYY\_N3\_N4 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であるような「縫合された」配列であり、またN1, 2, 3...は存在する場合、分析中に手動で編集を行える場合のある特定のエキソンを表わす(実施例5を参照)。または列5の識別番号は、「エキソンストレッチング(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、FLXXXXXXXX\_gAAAAA\_gBBBBB\_1\_N は「ストレッチ」配列の識別番号である。ここでXXXXXXXX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は特定エキソンを示す。(実施例5を参照。)RefSeq配列を「エキソンストレッチング」アルゴリズムのため蛋白質同族体として使用した場合には、RefSeq識別子(NM、NPあるいはNTにより表示)は、GenBank識別子(gBBBBB)の代わりに使用できる。

10

【0110】

或いは、プレフィックスは、手動編集、ゲノムのDNA配列からの予想、あるいは配列分析方法の組み合わせの由来による、成分配列を識別する。次の表に、プレフィックスに関連した成分配列 プレフィックスおよび対応する配列分析方法の例を掲げる(実施例4および5参照)。

20

プレフィックス	分析種別又はプログラム例
GN CG ENI	例えば GENSAN(アメリカカリフォルニア州スタンフォード大学)又はRENIS(イギリスケンブリッジのサンガー・センターのコンピュータ・ゲノミクス・グループ)を使用したゲノム配列からのエキソン予測
GH	ゲノム配列の再編成
HL	スパッチまたはストレッチされたゲノム配列の実施例5参照

30

40

【0111】

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の列5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、関連す

50

る *Incyte* cDNA 識別番号は示さなかった。

【0112】

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。

【0113】

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

10

【0114】

本発明はまた、SAT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、SATをコードするSEQ ID NO: 10 - 18 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。配列表に示したSEQ ID NO: 10 - 18 のポリヌクレオチド配列は、窒素系塩基のチミンがウラシルに置換され、糖鎖の背骨がデオキシリボースではなくリボースからなる等価RNA配列を含む。

【0115】

本発明には、SATをコードするポリヌクレオチド配列の変異配列も含まれる。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、SATをコードするポリヌクレオチド配列と少なくとも70%のポリヌクレオチド配列同一性、或いは少なくとも85%のポリヌクレオチド配列同一性、更には少なくとも95%ものポリヌクレオチド配列同一性を有する。本発明の或る実施態様では、SEQ ID NO: 10 - 18 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: 10 - 18 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、SAT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。

20

【0116】

遺伝暗号の縮重により作り出され得るSAT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。これらの組み合わせは、天然のSATのポリヌクレオチド配列に適用される標準的なトリプレット遺伝暗号を基に作られ、全ての変異が明確に開示されていると考慮する。

30

【0117】

SAT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SATのヌクレオチド配列とハイブリダイズ可能であるが、非天然のコドンを含めるなどの実質的に異なった使い方のコドンを有するSAT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ないで、SAT及びその誘導体をコードするヌクレオチド配列を実質的に変更する別の理由は、天然の配列から作られる転写物より例えば長い半減期など好ましい特性を備えるRNA転写物を作ることにある。

40

【0118】

本発明はまた、SAT及びその誘導体をコードするSAT配列またはそれらの断片を完全に合成化学によって作り出すことも含む。作製後にこの合成配列を、当分野で良く知られ

50

た試薬を用いて、種々の入手可能な発現ベクター及び細胞系の何れの中にも挿入可能である。更に、合成化学を用いて、S A Tまたはその任意の断片をコードする配列の中に突然変異を導入することも可能である。

【0119】

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: 10 - 18 及びそれらの断片とハイブリダイズ可能なポリヌクレオチド配列が含まれる（例えば、Wahl, G. M. 及び S. L. Berger (1987) *Methods Enzymol.* 152: 399 - 407; Kimmel, A. R. (1987) *Methods Enzymol.* 152: 507 - 511)。アニーリング及び洗浄条件を含むハイブリダイゼーションの条件は、「定義」に記載されている。

10

【0120】

DNAシーケンシングの方法は当分野でよく知られており、本発明の何れの実施例も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ラーゼを併用することができる。好適には、MICROLAB 2200液体転移システム (Hamilton, Reno, NV)、PTC 200サーマルサイクラー (MJ Research, Watertown 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調製を自動化する。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, pp. 856 - 853. 等を参照)。

20

30

【0121】

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、S A T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 (Sarkar, G. (1993) *PCR Methods Applic.* 2: 318322.)。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 (Triglia, T. 他 (1988) *Nucleic Acids Res.* 16: 8186.)。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増幅する方法に参与している。 (Lagerstrom, M. 他 (1991) *PCR Methods Applic.* 1: 111119.)。この方法では、PCRを行う前に複数の制限酵素の消化及び連結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

40

50

知られている。(Parker, J. D. (1991) *Nucleic Acids Res.* 19:3055-3060)。更に、PCR、ネステッドプライマー及びPromoter Finder (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。

10

**【0122】**

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、ランダムに初回抗原刺激を受けたライブラリは、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。

**【0123】**

市販されているキャピラリー電気泳動システムを用いて、シークエンシングまたは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、キャピラリーシークエンシングは、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、放出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小断片のシークエンシングに特に適している。

20

**【0124】**

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

30

**【0125】**

種々の目的でSATをコードする配列を変えるために、当分野で一般的に知られている方法を用いて、本発明のヌクレオチド配列を組換えることができる。この目的には、遺伝子産物のクローン化、プロセッシング及び/または発現の調節が含まれるが、これらに限定されるものではない。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仲介特定部位突然変異誘導を利用して、新規な制限部位の生成、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行う突然変異を導入し得る。

40

**【0126】**

本発明のヌクレオチドは、MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; 米国特許第5,837,458号; Chang, C.-C. 他(1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. 他(1999) *Nat. Biotechnol.* 17:259-264; Cramer, A. 他(1996) *Nat. Biotechnol.* 14:315-319)などのDNAシャッフリング技術の対象となり得るもので、生物学的または酵素的な活性、或いは他の分子や化合物と結合する能力な

50

どのSATの生物学的特性を変更或いは改良することができる。DNAシャッフリングは、遺伝子断片のPCR仲介再組換えを用いて遺伝子変異体のライブラリを生成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。このように、遺伝の多様性は「人為的」品種改良及び急速な分子の進化を経て創生される。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再結合し組み換えて、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再結合し組み換えて、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。

【0127】

別の実施例によれば、SATをコードする配列は、当分野で周知の化学的方法を用いて、全体或いは一部が合成可能である(例えば、Caruthers, M. H. ら(1980) *Nucl. Acids Res. Symp. Ser. 7*: 215-223; 及びHorn, T. 他(1980) *Nucl. Acids Res. Symp. Ser. 7*: 225-232を参照)。別法として、化学的方法を用いてSAT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. ら(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(Perkin Elmer)を用いて達成し得る。更にSATのアミノ酸配列または任意のその一部は、直接的な合成の際の変更、及び/または化学的方法を用いた他のタンパク質または任意のその一部からの配列との組み合わせにより、天然のポリペプチド配列を有するポリペプチドまたは変異体ポリペプチドを作製することが可能である。

【0128】

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上精製可能である(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)。

【0129】

生物学的に活性なSATを発現させるために、SATをコードするヌクレオチド配列またはその誘導体を好適な発現ベクターに挿入する。この発現ベクターは、好適な宿主に挿入されたコーディング配列の転写及び翻訳の調節に必要なエレメントを含む。これらのエレメントには、ベクター及びSAT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、SAT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。SAT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. ら(1994) *Results Probl. Cell Differ.* 20: 125-162. 等を参照)。

## 【0130】

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

10

## 【0131】

種々の発現ベクター/宿主系を利用して、S A Tをコードする配列の保持及び発現が可能である。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またはpBR322プラスミド）で形質転換させた植物細胞系、動物細胞系などの微生物等がある。（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. USA 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、Verma, I.M. 及び N. Somia (1997) Nature 389:239-242等を参照）。

20

30

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。

40

## 【0132】

細菌系では、S A T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多数のクローニングベクター及び発現ベクターを選択し得る。例えば、S A Tをコードするポリヌクレオチド配列の日常的なクローニング、サブクローニング、増殖には、PBLUESCRIPT (Stratagene, La Jolla CA) または pSPORT1 プラスミド (GIBCO BRL) などの多機能の大腸菌ベクターを用いることができる。ベクターの多数のクローニング部位にS A Tをコードする配列をライゲーションするとlacZ遺伝子が破壊され、組換え分子を含む形質転換された細菌の同定のための比色スクリーニング法が可能となる。更にこれらのベクターは、クローニングされた配列におけるin vitro転写、ジデオキシのシーケンシング、ヘルパーファージによる一本鎖の救

50

出、入れ子状態の欠失の生成にも有用であろう(例えば、Van Heeke, G. および S. M. Schuster (1989) J. Biol. Chem. 264: 55035509を参照)。例えば、抗体の産生のためなどに多量のSATが必要な場合は、SATの発現をハイレベルで誘導するベクターが使用できる。例えば、強力な誘導T5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。

【0133】

SATの発現に酵母の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、酵母菌サッカロミセス-セレビジエまたは*Pichia pastoris*に使用可能である。更に、このようなベクターは、発現したタンパク質の分泌か細胞内への保持のどちらかを誘導し、安定した増殖のために宿主ゲノムの中に外来配列を組み込む。(例えば、Ausubel, 1995, 前出、Bitter, G. A. ら (1987) Methods Enzymol. 153: 516-544、及び Scorer, C. A. ら (1994) Bio/Technology 12: 181-184. を参照)。

10

【0134】

植物系を使用してSATを発現することも可能である。SATをコードする配列の転写は、ウイルスプロモーター、例えば単独或いはTMV (Takamatsu, N. (1987) EMBO J 6: 307-311)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及び19Sプロモーターによって促進される。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. ら (1984) EMBO J. 3: 1671-1680; Broglie, R. ら (1984) Science 224: 838-843; および Winter, J. ら (1991) Results Probl. Cell Differ. 17: 85-105を参照)。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, pp. 191-196等を参照)。

20

30

【0135】

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、後発プロモーター及び3連リーダー配列からなるアデノウイルス転写物/翻訳複合体にSATをコードする配列を結合し得る。可欠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SAT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. および T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81: 36553659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。

40

【0136】

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6kb~10MbのHACsを作製し、従来の輸送方法(リポソーム、ポリカチオンアミノポリマー、またはベシクル)で供給する。(Harrington, J. J. ら (1997) Nat Genet. 15: 345-355、Price, C. M. (1993) Blood Rev. 7: 127-134、Trask, B. J. (1991) Trends Genet. 7: 149-154等を参照)。

【0137】

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SAT

50

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、S A T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。

#### 【0138】

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  
 10  
 ような選択系には、*tk* 単純細胞のために用いられる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(例えば、Wigler, M. 他 (1977) Cell 11: 223-232; Lowy, I. 他。(1980) Cell 22: 817-823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. ら (1980) Proc. Natl. Acad. Sci. USA 77: 35673570; Colbere Garapin, F.  
 20  
 ら (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アミノシアニン、緑色蛍光タンパク質(GFP; Clontech)、グルクロニダーゼ及びその基質GUS、ルシフェラーゼ及びその基質ルシフェリンなどの可視マーカーが用いられる。可視マーカー、例えばアミノシアニン、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  
 30  
 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)。

#### 【0139】

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、S A Tをコードする配列がマーカー遺伝子配列の中に挿入された場合、S A Tをコードする配列を含む形質転換された細胞は、マーカー遺伝子機能の欠落により特定可能である。または、1つのプロモーターの制御下でマーカー遺伝子がS A Tをコードする配列と一列に配置することも可能である。誘導または選択に  
 40  
 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。

#### 【0140】

一般に、S A Tをコードする核酸配列を含み且つS A Tを発現する宿主細胞は、当業者によく知られている種々の方法を用いて同定することが可能である。限定するものではないが当業者によく知られている方法には、DNA-DNA或いはDNA-RNAハイブリダイゼーション、PCR法、核酸或いはタンパク質の検出、定量、或いはその両方を行うための膜系、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の生物学的検定法または免疫学的検定法がある。

#### 【0141】

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S A T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)、  
 50

フローサイトメーター (FACS) などがある。SAT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部位モノクローナルベースの免疫アッセイ (two-site, monoclonal-based immunoassay) が好ましいが、競合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. ら (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN, Sect. IV、Coligan, J. E. ら (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY、Pound, J. D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ等を参照)。

#### 【0142】

多岐にわたる標識方法及び結合方法が、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。SATをコードするポリヌクレオチドに関連する配列を検出するための、標識されたハイブリダイゼーションプローブ或いはPCRプローブを生成する方法には、オリゴ標識化、ニックトランスレーション、末端標識化、または標識されたヌクレオチドを用いるPCR増幅が含まれる。別法として、SATをコードする配列、またはその任意の断片をmRNAプローブを生成するためのベクターにクローニングすることも可能である。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。

#### 【0143】

SATをコードするヌクレオチド配列で形質転換された宿主細胞は、細胞培地でのこのタンパク質の発現及び回収に好適な条件下で培養される。形質転換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。SAT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SAT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。

#### 【0144】

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」形を切断するような翻訳後処理を利用して、タンパク質のターゲティング、折りたたみ及び/または活性を特定することも可能である。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, Bethesda, 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。

#### 【0145】

本発明の別の実施例では、SATをコードする自然或いは変更された、または組換えの核酸配列を上記した任意の宿主系の融合タンパク質の翻訳となる異種配列に結合させる。例えば、市販の抗体によって認識できる異種部分を含むキメラSATタンパク質が、SATの活性のインヒビターに対するペプチドライブラリのスクリーニングを促進し得る。また

、異種タンパク質部分及び異種ペプチド部分も、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、SATをコードする配列と異種タンパク質配列との間にあるタンパク質分解切断部位を融合タンパク質が含むように遺伝子操作すると、SATが精製の後に異種部分から切断され得る。融合タンパク質の発現及び精製方法は、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。

10

20

30

40

50

#### 【0146】

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SAT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結合したタンパク質コード配列の転写及び翻訳を結合する。翻訳は、例えば<sup>3</sup><sup>5</sup>S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。

#### 【0147】

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

#### 【0148】

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SATの天然のリガンド、または天然の基質、構造的または機能的な擬態性または自然結合パートナーに密接に関連している (Coligan, J. E. 他 (1991) Current Protocols in Immunology 1 (2) の5章等を参照)。同様に、化合物は、SATが結合する天然受容体、或いは例えばリガンド結合部位などの少なくとも受容体のある断片に密接に関連し得る。何れの場合も、既知の技術を用いてこの化合物を合理的に設計することができる。一実施例では、このような化合物に対するスクリーニングには、分泌タンパク質或いは細胞膜上のタンパク質の何れか一方としてSAT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、ショウジョウバエ、大腸菌からの細胞が含まれる。SATを発現する細胞またはSATを含有する細胞膜断片を試験化合物と接触させて、SATまたは化合物の何れかの結合、刺激または阻害を分析する。

#### 【0149】

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SATと結合させるステップと、SATとこの化合物との結合を検出するステップを含み得る。別法では、標識された競合物の存在下での試験化合物の結合の検出及び測定を行うことができる。更にこのアッセイでは、無細胞再構成系、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。

#### 【0150】

本発明のSATまたはその断片を用いて、SATの活性を調整する化合物をスクリーニングすることが可能である。このような化合物には、アゴニスト、アンタゴニスト、或るい

は部分的または逆アゴニスト等が含まれる。一実施例では、S A Tが少なくとも1つの試験化合物と結合する、S A Tの活性が許容される条件下でアッセイを実施し、試験化合物の存在下でのS A Tの活性が試験化合物不在下でのS A Tの活性と比較する。試験化合物の存在下でのS A Tの活性の変化は、S A Tの活性を調整する化合物の存在を示唆する。別法では、試験化合物をS A Tの活性に適した条件下でS A Tを含む*in vitro*または無細胞再構成系と結合させてアッセイを実施する。これらアッセイの何れかにおいて、S A Tの活性を調整する試験化合物は間接的に結合することが可能であり、試験化合物と直接接触する必要がない。少なくとも1つから複数の試験化合物をスクリーニングすることができる。

#### 【0151】

別の実施例では、胚性幹細胞（ES細胞）における相同組換えを用いて動物モデル系内で、S A Tまたはその哺乳動物相同体をコードするポリヌクレオチドを「ロックアウト」する。このような技術は当技術分野において周知であり、ヒト疾患動物モデルの作製に有用である（米国特許第5,175,383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ロックアウトする（Marth, J. D. (1996) Clin. Invest. 97:1999-2002; Wagner, K. U. 他 (1997) Nucleic Acids Res. 25:4323-4330）。形質転換したES細胞を同定し、例えばC57BL/6マウス系等から採取したマウス細胞胚盤胞に微量注入する。胚盤胞を偽妊娠メスに外科的に導入し、得られるキメラ子孫の遺伝形質を決め、これを交配させてヘテロ接合性系またはホモ接合性系を作製する。このようにして作製した遺伝子組換え動物は、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。

#### 【0152】

S A T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 282:1145-1147参照）。

#### 【0153】

S A Tをコードするポリヌクレオチドを用いて、ヒト疾患をモデルとした「ロックイン」ヒト化動物（ブタ）または遺伝子組換え動物（マウスまたはラット）を作製することが可能である。ロックイン技術を用いて、S A Tをコードするポリヌクレオチドの或る領域を動物ES細胞に注入し、注入した配列を動物細胞ゲノムに組み込ませる。形質転換細胞を胞胚に注入し、胞胚を上記のように移植する。遺伝子組換え子孫または近交系について研究し、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S A Tを乳汁内に分泌するなどS A T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）。

#### 【0154】

（治療）

S A Tのある領域と分泌性分子および輸送分子のある領域との間に、例えば配列及びモチーフの文脈における化学的及び構造的類似性が存在する。更に、S A Tの発現は、脳組織、脊髄組織、リンパ組織、および生殖組織に密接に関連する。従って、S A Tは、小胞輸送障害、輸送障害、神経の疾患、筋疾患、及び免疫疾患や細胞異常増殖においてある役割を果たすと考えられる。S A Tの発現または活性の増大に関連する疾患の治療においては

10

20

30

40

50

、SATの発現または活性を低下させることが望ましい。また、SAT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、SATの発現または活性を増大させることが望ましい。

【0155】

従って、或る実施例において、SAT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SAT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、このような疾患のうち、輸送障害には運動不能症、筋萎縮性側索硬化症、毛細血管拡張性運動失調、ベッカー筋ジストロフィー、顔面麻痺、シャルコーマリー ツース病、糖尿病、尿崩症、糖尿病性ニューロパシー、デュシェンヌ型筋ジストロフィー、高カリウム血性周期性四肢麻痺、正常カリウム血性周期性四肢麻痺、パーキンソン病、悪性高熱、多剤耐性、重症筋無力症、筋緊張性異栄養症、緊張病、錐体外路性終末欠陥症候群、ジストニー、末梢神経疾患、脳性腫瘍、前立腺癌と、口峽炎、徐脈型不整脈、頻拍性型不整脈、高血圧症、遺伝性QT延長症候群、心筋炎、心筋症、ネマリンミオパシーラネマリン筋障害、中心核ミオパシー、脂質ミオパシー、ミトコンドリアミオパシー、甲状腺中毒性ミオパシー、エタノールミオパシー、皮膚筋炎、封入体筋炎、感染性節炎、及び多発性筋炎などの輸送に関連した心臓病と、アルツハイマー病、健忘症、双極性障害、痴呆、うつ病、てんかん、トゥレット病、妄想性精神病、及び分裂病などの輸送に関連した神経障害と、神経線維腫症、帯状疱疹後神経痛、3又神経ニューロパシー、サルコイドーシス、鎌状赤血球性貧血、ウィルソン病、白内障、不妊症、肺動脈狭窄症、常染色体性感音性難聴、高/低血糖症、グレーブス病、甲状腺腫、クッシング病と、副腎機能不全、グルコース ガラクトース吸収不全症候群、高コレステロール血症、副腎性白質ジストロフィー、ツェルヴェーガー症候群、メンケス病、後角症候群、フォンギルケ症候群、シスチン尿症、イミノグリシン尿症、Hartup病、ファンコニ病が含まれ、神経の疾患の中には、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクロイツフェルト ヤコブ病、ゲルストマン症候群、Gerstmann-Straussler-Scheinker症候群を含むプリオン病 (prion disease) と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫 (cerebelloretinal hemangioblastomatosis)、脳3又神経血管症候群、ダウン症を含む中枢神経系性精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィー及び他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性精神障害、及び妄想性精神病と、季節性の感情の障害 (SAD)、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病が含まれ、筋疾患の中には、心筋症、心筋炎、デュシェンヌ型筋ジストロフィー、ベッカー型偽肥大性筋ジストロフィー、筋緊張性ジストロフィー、中心コア病、ネマリンミオパシーラネマリン筋障害、中心核ミオパシー、脂質ミオパシー、ミトコンドリアミオパシー、感染性節炎、多発性筋炎、皮膚筋炎、封入体筋炎、甲状腺中毒性ミオパシー、エタノールミオパシー (ethanol myopathy)、口峽炎、アナフィラキシー、不整脈、喘息、心血管ショック、クッシング病、高血圧症、低血糖症、心筋梗塞、片頭痛、クロム親和細胞腫、脳症、てんかん、カーンズ セイヤ症候群、乳酸アシドーシス、ミオクローヌス疾患、眼筋麻痺、および酸性マルターゼ欠損症 (AMD、ポンペ病としても知られる) を含む筋障害が含まれ、免疫疾患の中には、炎症及び日光性角化症、後天性免疫不全症候群 (AIDS) 及び副腎機能不全、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲

状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー（A P E C E D）、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎症、骨関節炎、骨粗しょう症、腓炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷、および細胞異常増殖には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病（M C T D）、骨髓線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髓腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髓、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌が含まれる。

10

## 【0156】

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S A T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、S A T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。

## 【0157】

更に別の実施例では、限定するものではないが上に列記した疾患を含むS A T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S A Tを含む組成物を好適な医薬用担体と共に患者に投与することも可能である。

20

## 【0158】

更に別の実施例では、S A T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S A T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。

## 【0159】

更なる実施例では、S A T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S A T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、上記した小胞輸送障害、輸送障害、神経疾患、自己免疫/炎症疾患および細胞の増殖異常が含まれる。一実施態様では、S A Tと特異的に結合する抗体が直接アンタゴニストとして、或いはS A Tを発現する細胞または組織に薬剤を運ぶターゲティング或いは運搬機構として間接的に用いられ得る。

30

## 【0160】

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S A T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、S A Tをコードするポリヌクレオチドの相補配列を発現するベクターを患者に投与することも可能である。

## 【0161】

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 of 医薬原理に従ってを選択し得る。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。

40

## 【0162】

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

詳しくは、精製されたS A T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S A Tと特異的に結合するものを同定が能である。S A Tの抗体も、本技術分野で一般的に知られている方法を用いて製造することが可能である。限定するものでは

50

ないが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、一本鎖抗体、F a b断片及びF a b発現ライブラリによって作られた断片が含まれ得る。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。

**【0163】**

抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ヒト及びその他のものを含む種々の宿主が、S A Tまたは任意の断片、または免疫原性の特性を備えるそのオリゴペプチドの注入によって免疫化され得る。宿主の種に応じて、種々のアジュバントを用いて免疫応答を高めることもできる。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、ブルロニックポリオール、ポリアニオン、ペプチド、油性乳剤、スカイ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、B C G（カルメット ゲラン桿菌）及びコリネバクテリウム パルヴムが特に好ましい。

10

**【0164】**

S A Tに対する抗体を誘発するために用いられるオリゴペプチド、ペプチド、または断片は、少なくとも約5個のアミノ酸からなり、一般的には約10個以上のアミノ酸からなるものが好ましい。これらのオリゴペプチド、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り且つ小さな天然の分子の全アミノ酸配列を含むことが望ましい。S A Tアミノ酸の短いストレッチは、K L Hなどの別のタンパク質の配列と融合し、キメラ分子に対する抗体が産生され得る。

20

**【0165】**

S A Tに対するモノクローナル抗体は、培地内の連続した細胞株によって、抗体分子を産生する任意の技術を用いて作製することが可能である。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E B V - ハイブリドーマ技術がある（Kohler, G. ら. (1975) Nature 256: 495 - 497、Kozbor, D. ら (1985) J. Immunol. Methods 81: 31 - 42、Cote, R. J. ら (1983) Proc. Natl. Acad. Sci. USA 80: 2026 - 2030、Cole, S. P. ら (1984) Mol. Cell Biol. 62: 109 - 120等を参照）。

**【0166】**

更に、「キメラ抗体」作製のために発達したヒト抗体遺伝子にマウス抗体遺伝子をスプライシングするなどの技術が、好適な抗原特異性及び生物学的活性を備える分子を得るために用いられる（例えば、Morrisson, S. L. 他. (1984) Proc. Natl. Acad. Sci. 81: 6851 - 6855、Neuberger, M. S. 他. (1984) Nature 312: 604 - 608; Takeda, S. ら. (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、S A T特異的一本鎖抗体を生成する。関連特異性を有するガ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）。

30

40

**【0167】**

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、或いは免疫グロブリンライブラリのスクリーニングまたは文献に開示されているような高特異結合試薬のパネルのスクリーニングによっても行い得る（Orlandi, R. ら (1989) Proc. Natl. Acad. Sci. USA 86: 3833 - 3837、Winter, G. ら (1991) Nature 349: 293 - 299等を参照）。

**【0168】**

S A Tに対する特異的な結合部位を含む抗体も得ることができる。例えば、限定するもの

50

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作製される  $F(a b')$ <sub>2</sub> 断片と、 $F(a b')$ <sub>2</sub> 断片のジスルフィド架橋を還元することによって作製される  $F a b$  断片とがある。或いは、 $F a b$  発現ライブラリを作製することによって、モノクローナル  $F a b$  断片を所望の特異性と迅速且つ容易に同定することが可能となる (H u s e , W . D . ら ( 1 9 8 9 ) *S c i e n c e* 2 4 6 : 1 2 7 5 1 2 8 1 等を参照)。

#### 【0169】

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。隔離された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる競合的な結合、または免疫放射線活性のための数々のプロトコルが、当分野では周知である。通常このようなイムノアッセイには、S A T 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 S A T ピトーブに対して反応性のモノクローナル抗体を用いる、2 部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合的結合アッセイも利用することができる (P o u n d、前出)。

10

#### 【0170】

ラジオイムノアッセイ技術と共に S c a t c h a r d 分析などの様々な方法を用いて、S A T に対する抗体の親和性を評価する。親和性を結合定数  $K a$  で表すが、この  $K a$  は、平衡状態の下で S A T 抗体複合体のモル濃度を遊離抗体と遊離抗原のモル濃度で除して得られる値である。ポリクローナル抗体は多様な S A T ピトーブに対する親和性が不均一であり、ポリクローナル抗体試薬のために決定した  $K a$  は、S A T 抗体の平均親和性または結合活性を表す。特定の S A T ピトーブに単一特異的なモノクローナル抗体医薬の  $K a$  は、親和性の真の測定値を表す。 $K a$  値が  $10^9 \sim 10^{12}$  liter / mol の高親和性抗体医薬は、S A T 抗体複合体が激しい操作に耐えなければならないイムノアッセイに用いるのが好ましい。 $K a$  値が  $10^6 \sim 10^7$  liter / mol の低親和性抗体医薬は、S A T が抗体から最終的に活性化状態で解離する必要がある免疫精製 (i m m u n o p u r i f i c a t i o n) 及び類似の処理に用いるのが好ましい。(C a t t y , D . ( 1 9 8 8 ) *Antibodies, Volume I: A Practical Approach*, I R L P r e s s , W a s h i n g t o n , D C ; L i d d e l l , J . E . 及び C r y e r , A . ( 1 9 9 1 ) *A Practical Guide to Monoclonal Antibodies*, J o h n W i l e y & S o n s

20

30

#### 【0171】

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 の特異的な抗体を含むポリクローナル抗体医薬は一般に、S A T 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 C a t t y の文献、同 C o l l i g a n らの文献等を参照)。

#### 【0172】

本発明の別の実施例では、S A T 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、S A T をコードする遺伝子のコーディング領域や調節領域に相補的な配列やアンチセンス分子 (D N A 及び R N A、修飾ヌクレオチド) を設計して遺伝子発現を変更することができる。このような技術は当分野では周知であり、センスまたはアンチセンスオリゴヌクレオチドまたは大きな断片が、S A T 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。(A g r a w a l , S . , e d . ( 1 9 9 6 ) *Antisense Therapeutics*, H u m a n a P r e s s I n c . , T o t a w a N J を参照)。

40

#### 【0173】

50

治療に用いる場合、アンチセンス配列を好適な標的細胞に導入するのに好適な任意の遺伝子送達系を用いることができる。アンチセンス配列は、転写時に標的タンパク質をコードする細胞配列の少なくとも一部に相補的な配列を発現する発現プラスミドの形で細胞内に輸送することが可能である (Slater, J. E. ら (1998) *J. Allergy Clin. Immunol.* 102 (3): 469 - 475 及び Scanlon, K. J. ら (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関連ウイルスベクター等のウイルスベクターを用いて細胞内に導入することもできる。(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. ら (1998) *J. Pharm. Sci.* 87 (11): 1308 - 1315、Morris, M. C. ら (1997) *Nucleic Acids Res.* 25 (14): 2730 - 2736. 等を参照)。

#### 【0174】

本発明の別の実施例では、SAT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. ら (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. ら (1995) *Science* 270: 475 - 480、Bordignon, C. ら (1995) *Science* 270: 470 - 475)、嚢胞性繊維症 (Zabner, J. ら (1993) *Cell* 75: 207 - 216; Crystal, R. G. ら (1995) *Hum. Gene Therapy* 6: 643 - 666、Crystal, R. G. ら (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. (1997) *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)、(iii) 細胞内の寄生虫 (例えばヒト免疫不全ウイルス (HIV) (Baltimore, D. (1988) *Nature* 335: 395 - 396、Poeschl, E. ら (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。SAT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SAT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。

#### 【0175】

本発明の更なる実施例では、SATの欠損による疾患や異常症は、SAT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SAT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. An 50

erson (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。

【0176】

SATを及ぼし得る発現ベクターには、限定するものではないが、PCDNA3.1、EPITAG、PRCCMV2、PREP、PVAXベクター(Invitrogen, Carlsbad CA)、PCMV-SCRIPT、PCMV-TAG、PEGSH/PERV(Stratagene, La Jolla CA)、PTET-OFF、PTET-ON、PTRE2、PTRE2-LUC、PTK-HYG(Clontech, Palo Alto CA)が含まれる。SAT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. 及び 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. 及び H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456)、エクジソン誘導性プロモーター(市販されているプラスミドPVGRXR及びPINDに含まれている:Invitrogen)、FK506/ラパマイシン誘導性プロモーター、またはRU486/ミフェプリストーン誘導性プロモーター(Rossi, F.M.V. 及び H.M. Blau, 前出)、または(iii)正常な個体に由来するSAT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。

【0177】

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. ら (1982) *EMBO J.* 1:841-845)を用いて形質転換を行う。(1982) *EMBO J.* 1:841-845)。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。

【0178】

本発明の別の実施例では、SAT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SATをコードするポリヌクレオチドと、(ii)好適なRNAパッケージングシグナルと、(iii)追加のレトロウイルス・シス作用性RNA配列及び効率的なベクターの増殖に必要なコーディング配列を伴うRev応答性エレメント(RRE)とからなるレトロウイルスベクターを作製して治療することができる。レトロウイルスベクター(例えばPFB及びPFBNEO)はStratagene社から市販されており、刊行データ(Riviere, I. ら (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92:6733-6737)に基づいている。上記データを引用することをもって本明細書の一部とする。(1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737)、参考文献に記載。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向性を有するエンベロープ遺伝子またはVSVg等の乱交雑エンベロープタンパク質を発現する(Armentano, D. ら (1987) *J. Virol.* 61:1647-1650、Bender, M.A. ら (1987

) J. Virol. 61:1639-1646、Adam, M. A. 及び A. D. Miller (1988) J. Virol. 62:3802-3806、Dull, T. ら (1998) J. Virol. 72:8463-8471、Zufferey, R. ら (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<sup>+</sup>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. ら (1997) J. Virol. 71:7020-7029、Bauer, G. ら (1997) Blood 89:2259-2267、Bonyhadi, M. L. (1997) J. Virol. 71:4707-4716、Ranga, U. ら (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206、Su, L. (1997) Blood 89:2283-2290)。

【0179】

別法では、アデノウイルス系遺伝子治療の送達系を用いて、SATの発現に関連する1或いは複数の遺伝子異常を有する細胞にSATをコードするポリヌクレオチドを送達する。アデノウイルス系ベクターの作製及びパッケージングについては、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. ら (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. ら (1999) Annu. Rev. Nutr. 19:511-544 及び 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。

【0180】

別法では、ヘルペス系遺伝子治療の送達系を用いて、SATの発現に関連する1つ或いは複数の遺伝子異常を有する標的細胞にSATをコードするポリヌクレオチドを送達する。単純疱疹ウイルス(HSV)系のベクターは、HSV親和性の中枢神経細胞にSATを導入する際に特に重要である。ヘルペス系ベクターの作製及びパッケージングは、当業者に公知である。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. ら (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")が開示されており、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. ら (1999) J. Virol. 73:519-532 及び Xu, H. ら (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

10

20

30

40

50

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。

【0181】

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SATをコードするポリヌクレオチドを標的細胞に送達する。プロトタイプのウイルスであるセムリキ森林熱ウイルス（Semliki Forest Virus, SFV）の生物学的研究が広範に行われており、遺伝子導入ベクターがSFVゲノムに基づいていることが分かった（Garoff, H. および 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、SATをコードする配列をウイルスゲノムのカプシッドをコードする領域に導入することによって、ベクター導入細胞において多数のSATをコードするRNAが産生され、高いレベルでSAT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. ら (1997) *Virology* 228: 74 - 83）。様々な宿主にウイルスを導入できることから、様々なタイプの細胞にSATを導入することができる。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。

10

20

【0182】

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

30

【0183】

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、ヌクレオチド鎖切断に先立つ相補的標的RNAへのリボザイム分子の配列特異性ハイブリダイゼーションに関与している

40

【0184】

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。GUA、GUU、GUC。一度同定すると、オリゴヌクレオチドを機能不全にするような2次構造の特徴に対して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実施容易性をテストすることによって行うことができる。

50

## 【0185】

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合物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、S A T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。

## 【0186】

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リボチオネートまたは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、内在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*) 等を加えることができる。

## 【0187】

本発明の更なる実施例は、S A T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変異発現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、S A T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、S A T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、S A T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、S A T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

## 【0188】

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。S A Tをコードするポリヌクレオチドを含むサンプルは、少なくとも1つの試験化合物に曝露して得る。サンプルには例えば、無傷細胞、透過化処理した細胞、無細胞再構成系または再構成生化学系があり得る。S A T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、S A T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つ以上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. ら (1999

10

20

30

40

50

) 米国特許第5,932,435号、Arndt, G.M. ら (2000) *Nucleic Acids Res.* 28:E15) または HeLa 細胞等のヒト細胞系 (Clarke, M.L. ら (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. ら (1997) 米国特許第5,686,242号、Bruce, T.W. ら (2000) 米国特許第6,022,691号)。

【0189】

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. ら (1997) *Nat. Biotechnol.* 15:462-466. 等を参照。(1997) *Nat. Biotechnol.* 15:462-466.)

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。

【0190】

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成分の投与に関連する。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、SAT、SATの抗体、擬態、アゴニスト、アンタゴニスト、またはSATのインヒビターなどからなる。

【0191】

本発明に用いられる成分は、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。

【0192】

肺から投与する成分は、液状または乾燥粉末状で調製し得る。このような成分は通常、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. ら, 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。

【0193】

本発明での使用に適した成分には、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成分が含まれる。有効投与量の決定は、当業者の能力の範囲内で行う。

【0194】

SAT またはその断片を含む高分子を直接細胞内に送達するべく、特殊な形態に組成物が調製されるのが好ましい。例えば、細胞不透過性高分子を含むリボソーム製剤は、細胞融合及び高分子の細胞内送達を促進し得る。別法では、SAT またはその断片を HIV Tat-1 タンパク質の短いカチオン 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. ら (1999) *Science*

10

20

30

40

50

285:1569-1572)。

【0195】

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、ウサギ、イヌ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。

【0196】

医学的に効果的な薬用量は、症状や容態を回復させる、たとえばSAT またはその断片、SATの抗体、SATのアゴニストまたはアンタゴニスト、インヒビターなどの活性処方成分の量に関連する。治療有効度及び毒性は、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ED<sub>50</sub>（集団の50%の医薬的有効量）またはLD<sub>50</sub>（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。毒性効果の薬用効果に対する投与量の比は、治療指数であり、LD<sub>50</sub>/ED<sub>50</sub>比として表すことができる。高い治療指数を示すような成分が望ましい。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、ED<sub>50</sub>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、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。

【0197】

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。作用期間が長い成分は、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。

【0198】

通常投与量は、投与の経路にもよるが約0.1～100,000μg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。

【0199】

(診断)

別の実施例では、SATに特異的に結合する抗体が、SATの発現によって特徴付けられる疾患の診断、またはSATやSAT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。SAT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SAT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。

【0200】

SATを測定するためのELISA,RIA,及びFACSを含む種々のプロトコルは、当分野では周知であり、変わった或いは異常なレベルのSATの発現を診断する元となるものを提供する。正常或いは標準的なSATの発現の値は、複合体の形成に適した条件下、正常な哺乳動物、例えばヒトなどの被験者から採取した体液または細胞とSATに対

10

20

30

40

50

する抗体とを結合させることによって決定する。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。被験者のSATの発現の量、制御及び疾患、生検組織からのサンプルが基準値と比較される。基準値と被験者との偏差が疾患を診断するパラメータとなる。

#### 【0201】

本発明の別の実施例によれば、SATをコードするポリヌクレオチドを診断のために用いることもできる。用い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SATを発現する生検組織における遺伝子の発現を検出し定量する。この診断アッセイを用いて、SATの存在の有無、更に過剰な発現を調べ、治療中のSAT値の調節を監視する。

10

#### 【0202】

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

#### 【0203】

プローブはまた、関連する配列の検出に利用され、SATをコードする任意の配列と少なくとも50%の配列同一性を有し得る。目的の本発明のハイブリダイゼーションプローブには、DNAあるいはRNAが可能であり、SEQ ID NO: 10 - 18の配列、或いはSAT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。

20

#### 【0204】

SATをコードするDNAに対して特異的なハイブリダイゼーションプローブの作製方法には、SAT及びSAT誘導体をコードするポリヌクレオチド配列をmRNAプローブの作製のためのベクターにクローニングする方法がある。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、<sup>32</sup>Pまたは<sup>35</sup>S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。

30

#### 【0205】

SATをコードするポリヌクレオチド配列を用いて、SATの発現に関連する疾患を診断することが可能である。限定するものではないが、このような疾患のうち、輸送障害には運動不能症、筋萎縮性側索硬化症、毛細血管拡張性運動失調、ベッカー筋ジストロフィー、顔面麻痺、シャルコーマリーツース病、糖尿病、尿崩症、糖尿病性ニューロパシー、デュシェンヌ型筋ジストロフィー、高カリウム血性周期性四肢麻痺、正常カリウム血性周期性四肢麻痺、パーキンソン病、悪性高熱、多剤耐性、重症筋無力症、筋緊張性異栄養症、緊張病、錐体外路性終末欠陥症候群、ジストニー、末梢神経疾患、脳性腫瘍、前立腺癌と、口峡炎、徐脈型不整脈、頻拍性型不整脈、高血圧症、遺伝性QT延長症候群、心筋炎、心筋症、ネマリンミオパシーラネマリン筋障害、中心核ミオパシー、脂質ミオパシー、ミトコンドリアミオパシー、甲状腺中毒性ミオパシー、エタノールミオパシー、皮膚筋炎、封入体筋炎、感染性節炎、及び多発性筋炎などの輸送に関連した心臓病と、アルツハイマー病、健忘症、双極性障害、痴呆、うつ病、てんかん、トゥーレット病、妄想性精神病、及び分裂病などの輸送に関連した神経障害と、神経線維腫症、帯状疱疹後神経痛、3叉神経ニューロパシー、サルコイドーシス、鎌状赤血球性貧血、ウィルソン病、白内障、

40

50

不妊症、肺動脈狭窄症、常染色体性感音性難聴、高ノ低血糖症、グレーブス病、甲状腺腫、クッシング病と、副腎機能不全、グルコース ガラクトース吸収不全症候群、高コレステロール血症、副腎性白質ジストロフィー、ツェルヴェーガー症候群、メンケス病、後角症候群、フォンギルケ症候群、シスチン尿症、イミノグリシン尿症、Hartup病、ファンコニ病が含まれ、神経の疾患の中には、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクロイツフェルト ヤコブ病、ゲルストマン症候群、Gerstmann-Straussler-Scheinker症候群を含むプリオン病(prion disease)と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫(cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、ダウン症を含む中枢神経系性精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィー及び他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性精神障害、及び妄想性精神病と、季節性の感情の障害(SAD)、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病が含まれ、筋疾患の中には、心筋症、心筋炎、デュシェンヌ型筋ジストロフィー、ベッカー型偽肥大性筋ジストロフィー、筋緊張性ジストロフィー、中心コア病、ネマリンミオパシーラネマリン筋障害、中心核ミオパシー、脂質ミオパシー、ミトコンドリアミオパシー、感染性節炎、多発性筋炎、皮膚筋炎、封入体筋炎、甲状腺中毒性ミオパシー、エタノールミオパシー(ethanol myopathy)、口峡炎、アナフィラキシー、不整脈、喘息、心血管ショック、クッシング病、高血圧症、低血糖症、心筋梗塞、片頭痛、クロム親和細胞腫、脳症、てんかん、カーンズ セイヤ症候群、乳酸アシドーシス、ミオクロノス疾患、眼筋麻痺、および酸性マルターゼ欠損症(AMD、ポンペ病としても知られる)を含む筋障害が含まれ、免疫疾患の中には、炎症及び日光性角化症、後天性免疫不全症候群(AIDS)及び副腎機能不全、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー(APECED)、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、骨粗しょう症、腓炎、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷、および細胞異常増殖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、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、子宮の癌が含まれる。S A Tをコードするポリヌクレオチド配列は、サザーン法やノーザン法、ドットプロット法、或いはその他の膜系の技術、PCR法、ディップスティック(dipstick)、ピン(pin)、ELISA式アッセイ、及び変異S A Tの発現を検出するために患者から採取した体液或いは組織を利用するマイクロアレイに使用することが可能である。このよ

10

20

30

40

50

うな定性方法または定量方法は、当分野で公知である。

【0206】

ある実施態様では、SATをコードするヌクレオチド配列は、関連する疾患、特に上記した疾患を検出するアッセイにおいて有用であろう。SATをコードするヌクレオチド配列は、標準的な方法で標識化され、ハイブリダイゼーション複合体の形成に好適な条件の下、患者から採取した体液或いは組織のサンプルに加えることができるであろう。好適な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。患者のサンプルのシグナルの量が、制御サンプルと較べて著しく変わっている場合は、サンプル内のSATをコードするヌクレオチド配列の変異レベルにより、関連する疾患の存在が明らかになる。このようなアッセイは、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。

10

【0207】

SATの発現に関連する疾患の診断の基準となるものを提供するために、発現の正常すなわち標準的なプロファイルが確立される。これは、ハイブリダイゼーション或いは増幅に好適な条件の下、動物或いはヒトの何れかの正常な被験者から抽出された体液或いは細胞と、SAT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対象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。

20

【0208】

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。

【0209】

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、実際に臨床的症状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。

30

【0210】

SATをコードする配列から設計されたオリゴヌクレオチドのさらなる診断への利用には、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SATをコードするポリヌクレオチドの断片、或いはSATをコードするポリヌクレオチドと相補的なポリヌクレオチドの断片を含み、最適な条件の下、特定の遺伝子や条件を識別するために利用される。また、オリゴマーは、やや緩いストリンジェント条件下で、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。

40

【0211】

或る実施態様において、SAT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、SATをコードするポリヌクレオチド配列由来のオリゴヌクレオチドプライマーを用いたポリメラーゼ連鎖反応（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

50

異は非変性ゲル中でのゲル電気泳動法を用いて検出可能である。f S C C P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 ( a m p l i m e r ) の検出が可能になる。更に、インシリコSNP ( i n s i l i c o S N P , i s S N P ) と呼ばれる配列データベース分析法は、一般的なコンセンサス配列に配列されるような個々の重畳するDNA断片の配列を比較することにより、多形性を同定し得る。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M A S S A R R A Yシステム ( S e q u e n o m , I n c . , S a n D i e g o C A ) を用いた質量分析によりSNPを検出し、特徴付ける。

10

**【0212】**

S A T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 ( c o a m p l i f i c a t i o n ) 及び標準曲線から得た結果の補間もある (例えば、Melby, P. C. ら (1993) J. Immunol. Methods, 159: 235 - 244; Duplaa, C. ら (1993) Anal. Biochem. 212: 229 - 236を参照)。目的のオリゴマー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。

**【0213】**

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、突然変異及び多形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有効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。

20

30

**【0214】**

別の実施例では、S A T、S A Tの断片、S A Tに特異的な抗体をマイクロアレイ上のエレメントとして用いることができる。マイクロアレイを用いて、上記のようなタンパク質-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。

**【0215】**

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及び相対存在量を定量することにより分析し得る (Seilliamer らの米国特許第5, 840, 484号 "Comparative Gene Transcript Analysis" を参照。該特許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ァイルを提供し得る。

40

**【0216】**

50

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。

【0217】

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サイン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. ら. (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同一のシグネチャを有する場合には、毒性特性を共有している可能性がある。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。自己の発現が任意の試験された化合物により変化しない遺伝子が同様に重要であっても、このような遺伝子の発現レベルを用いて残りの発現データを規準化する。規準化手順は、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が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、中毒学的スクリーニングの際に毒性シグネチャを用いて、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。

10

20

【0218】

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、非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。

30

【0219】

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、通常クーマシーブルーまたはシルバーまたは蛍光染色などの物質を用いてゲルで染色することにより、分散した、独特な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非処理のいずれかの生物学的サンプルから得られるタンパク質スポットの光学密度を比較し、処理に関連するタンパク質スポット密度の変化を同定

40

50

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が得られる。

#### 【0220】

プロテオームのプロファイルは、SATに特異的な抗体を用いてSAT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. ら (1999) Anal. Biochem. 270:103-111; Mendozze, L.G. 他(1999) Biotechniques 27:778-788を参照)。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

#### 【0221】

プロテオームレベルでの毒性サイン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サイン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. および J. Seilhamer (1997) Electrophoresis 18:533-537)、転写イメージにはそれ程影響しないがタンパク質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サインは有用たり得る。更に、体液中での転写の分析はmRNA急速な分解により困難であるので、タンパク質のプロフィール作成はこのような場合により信頼でき、情報価値がある。

20

#### 【0222】

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

#### 【0223】

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反応を示す。

#### 【0224】

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. ら(1995)の米国特許第5,474,796号、Schena, M. ら(1996) Proc. Natl. Acad. Sci. USA 93:10614-10619、Baldeschweiler らの(1995) PCT出願第WO95/251116号、Shalon, D. らの(1995) PCT出願第WO95/35505号、Heller, R.A. ら(1997) Proc. Natl. Acad. Sci. USA 94:2150-2155、Heller, M.J. らの(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周知であり、詳細については、DNA Microarrays: A Practical Approach, M. Schena, ed.

40

50

(1999) Oxford University Press, Londonに記載されている。該文献は、特別に引用することを以って本明細書の一部となす。

【0225】

本発明の別の実施例ではまた、SAT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、コード配列全体で非コード配列が好ましい。例えば、多重遺伝子ファミリーのメンバー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。核酸配列は、特定の染色体、染色体の特定領域または人工形成の染色体、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、或いは単一染色体cDNAライブラリに対してマッピングされる(Harrington, J. J. ら(1997) Nat Genet. 15:345-355、Price, C. M. (1993) Blood Rev. 7:127-134、Trask, B. J. (1991) Trends Genet. 7:149-154等を参照)。一度マッピングすると、本発明の核酸配列を用いて例えば病状の遺伝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. 及びD. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357を参照。)

10

蛍光原位置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(前出のHeinz-Ulrich, ら(1995) in Meyers, 965-968ページ、等を参照)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SATをコードする遺伝子の位置と特定の疾患との相関性、或いは特定の疾患に対する素因が、このような疾患と関連するDNA領域の決定に役立つため、更なる位置を決定するクローニングが行われる。

20

【0226】

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置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、正確な染色体の遺伝子座がわかっていない場合でも関連するマーカーを明らかにし得る。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。疾患または症候群が、血管拡張性失調症の11q22-23領域等、特定の遺伝子領域への遺伝的結合によって大まかに位置決めがなされると、該領域に対するいかなるマッピングも、更なる調査のための関連遺伝子或いは調節遺伝子を表すことができる(Gatti, R. A. ら(1988) Nature 336:577-580等を参照)転座、反転等に起因する、健常者、保有者、感染者の三者間における染色体位置の相違を発見するために本発明のヌクレオチド配列を用いてもよい。

30

【0227】

本発明の別の実施例では、SAT、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。薬剤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になる。SATと検査する薬剤との結合による複合体の形成を測定してもよい。

40

【0228】

別の薬剤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(Geyesen, らの(1984) PCT出願番号WO84/03564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質上で合成する。試験用化合物は、SAT、或いはその断片と反応してから洗浄される。次ぎに、結合されたSATが、当分野で周知の方法で

50

検出される。精製されたSATはまた、前記した薬剤をスクリーニングする技術に用いられるプレート上で直接被覆することもできる。別法では、非中和抗体を用いてペプチドを捕捉し、ペプチドを固体支持物に固定することもできる。

【0229】

別の実施例では、SATと結合可能な中和抗体がSAT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、SAT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。

【0230】

別の実施例では、発展途上の分子生物学技術にSAT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。

10

【0231】

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。従って、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。

【0232】

前述した及び以下に記載する全ての特許出願、特許、刊行物、特に米国特許出願第60/215,465号、同第60/239,384号および同第60/253,639号に言及することをもって本明細書の一部とする。

【0233】

20

1 cDNAライブラリの作製

Incyte cDNAは、LIFESEQ GOLDデータベース(Incyte Genomics, Palo Alto CA)に記載されたcDNAライブラリに由来するものであり、表4の列5に列記した。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た組織もあり、また、ホモジナイズしてフェノールまたは好適な変性剤の混合液に溶解した組織もある。変性剤の混合液は、例えばフェノールとグアニジニウムイソチオシアネートの単相溶液であるTRIZOL(Life Technologies)等である。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。

30

【0234】

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)連結磁性粒子(Promega)、OLIGOTEXラテックス粒子(QIAGEN, Valencia 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。

【0235】

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SUPERSCRIPTプラスミドシステム(Life Technologies)を用いて本技術分野で公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, unit 5.1-6.6等を参照)。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(300~1000bp)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

40

50

アガロースゲル電気泳動法を用いて行った。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。好適なプラスミドは、例えばPBLUESCRIPTプラスミド(Stratagene)、pSPORT1プラスミド(Life Technologies)またはpINCY(Incyte Pharmaceuticals, Palo Alto CA)等である。組換えプラスミドは、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。

#### 【0236】

10

##### 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。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及びQIAWELL 8 Ultra Plasmid 精製システム、R.E.A.L. Prep 96プラスミドキットの中から少なくとも1つを用いて、プラスミドを精製した。沈殿させた後、0.1mlの蒸留水に再懸濁して、凍結乾燥して或いは凍結乾燥せずに、4℃で保管した。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及びQIAWELL 8 Ultra Plasmid 精製システム、R.E.A.L. Prep 96プラスミドキット。沈殿させた後、0.1mlの蒸留水に再懸濁して、凍結乾燥して或いは凍結乾燥せずに、4℃で保管した。

20

#### 【0237】

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)。宿主細胞の溶解及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穴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。

30

#### 【0238】

##### 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Incycyte 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)に与えられた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)か、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

40

50

ム (Applied Biosystems) か、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。

【0239】

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、BLAST、FASTA 及び BLIMPS に基づくプログラムを用いて、プログラム中の注釈を得るべく、公共のデータベース、例えば GenBank の霊長類及びげっ歯類、哺乳動物、脊椎動物、真核生物のデータベースと、BLOCKS、PRINTS、DOMO、PRODOM 及び PFAM 等の Hidden Markov Model (HMM) ベースのタンパク質ファミリーデータベースの選択に対する Incyte cDNA 配列またはその翻訳を問い合わせた (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)。Eddy, S. R. (1996) Cuff. Opin. Struct. Biol. 6: 361-365 等を参照) の選択に対する Incyte cDNA 配列またはその翻訳を問い合わせた。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した。或いは、GenBank cDNA、GenBank EST、ステッチされた配列、ストレッチされた配列または Genscan 予測コード配列 (実施例 4 及び 5 を参照) を用いて Incyte cDNA の集団を完全長まで伸長させた。Phred、Phrap 及び Consed に基づくプログラムを用いて構築し、GenMark、BLAST 及び FASTA 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対してスクリーニングした。対応する完全長ポリペプチド配列を誘導するべく完全長ポリヌクレオチド配列を翻訳した。或いは、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、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) ベースの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、MACDNASIS 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) 及び 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 マルチシーケンズ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。

【0240】

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 (スコアが高ければ高いほど 2 配列間の相同性が高くなる)。

【0241】

完全長のポリヌクレオチド配列およびポリペプチド配列の組み立て及び分析に用いる上記のプログラムは、SEQ ID NO: 10 - 18 のポリヌクレオチド配列断片の同定にも

10

20

30

40

50

利用できる。ハイブリダイゼーション及び増幅技術に有用である約20～約4000ヌクレオチドの断片を表4の列4に示した。

#### 【0242】

##### 4 ゲノムDNAからのコード配列の同定及び編集

推定分泌分子と輸送分子は、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. および S. Karlin (1997) J. Mol. Biol. 268:78-94 及び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び構築されたcDNA配列を形成する。Genscanの出力は、ポリヌクレオチド及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、30kbに設定した。これらのGenscan推定cDNA配列の内、どの配列が分泌分子と輸送分子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モデルにおいて分泌分子と輸送分子について問合せて分析した。潜在的な分泌分子と輸送分子が、分泌分子と輸送分子としてアノテーションが付けられたインサイトcDNA配列に対する相同性を基に同定された。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gbpri及びgbhtgと比較した。必要であれば、genpeptからヒットしたトップのBLASTと比較することによりGenscan予測配列を編集し、余分なまたは取り除かれたエキソンなどのGenscanにより予測された配列のエラーを修正する。BLAST分析はまた、任意のIncyte cDNAまたはGenscan予測配列の公共cDNA適用範囲の発見に用いられるので、転写の証拠を提供する。Incyte cDNA適用範囲が利用できる場合には、この情報を用いてGenscan予測配列を修正または確認した。完全長ポリヌクレオチド配列は、実施例3に記載された構築プロセスを用いて、Incyte cDNA配列及び/または公共のcDNA配列でGenscan予測コード配列を構築することにより得た。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。

10

20

30

#### 【0243】

##### 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された区間を、親配列 (parent sequence) に沿って現われるようにステッチアルゴリズムで縫い合わせ、可能な最も長い配列および変異配列を作製する。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。結果として得られるステッチ配列は、BLAST分析により公共データベースgenpept及びgbpriに翻訳されて比較された。Genscanにより予測された不正確なエキソンは、genpeptからヒットしたトップのBLAST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

40

50

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。

【0244】

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関連して、キメラタンパク質内で挿入または削除が起こり得る。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決定した。

10

【0245】

6 SAT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: 10 - 18を構築するために用いた配列を、BLAST及びSmith-Watermanアルゴリズムを用いて、Incyte LIFESEQデータベース及び公共のドメインデータベースの配列と比較した。SEQ ID NO: 10 - 18と一致するこれらのデータベースの配列を、Phrap(表7)などの構築アルゴリズムを使用して、連続及び重複した配列のクラスターに組み入れた。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等の公的な情報源から入手可能な放射線ハイブリッド及び遺伝地図データを用いて、クラスタ化された配列が前もってマッピングされたかを測定した。マッピングされた配列がクラスタに含まれている結果、個々の配列番号を含めてそのクラスタの全配列が地図上の位置に割り当てられた。

20

【0246】

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、1cM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、配列が各クラスタ内に含まれるような放射線ハイブリッドマーカーに対して境界を提供するようなGenethonによってマッピングされた遺伝マーカーに基づく。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。

30

【0247】

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参与している。(前出のSambrook, 7章、同Ausubel, F.M.ら, 4章及び16章等を参照)。

40

【0248】

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 Pharmaceuticals)等のヌクレオチド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、特定の同一を厳密な或いは相同的なものとして分類するか否

50

かを決定するため、コンピュータ検索の感度を変更することができる。検索の基準はプロダクト積スコアであり、次式で定義される。

【数 1】

(BLAST スコア×配列一致率)

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

プロダクト積スコアは、2つの配列間の類似度及び配列が一致する長さの両方を考慮している。プロダクト積スコアは、0～100の規準化された値であり、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。高スコアリングセグメント対(HSP)に一致する各塩基に+5のスコアを割り当て、各不適性塩基対に-4を割り当てることにより、BLASTスコアを計算する。2つの配列は、2以上のHSPを共有し得る(ギャップにより隔離され得る)。2以上のHSPがある場合には、最高BLASTスコアの塩基対を用いてプロダクト積スコアを計算する。プロダクト積スコアは、断片的重畳とBLASTアラインメントの質とのバランスを表す。例えばプロダクト積スコア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。プロダクト積スコア70は、一端が100%一致し、70%重畳しているか、他端が88%一致し、100%重畳しているかのいずれかの場合に得られる。プロダクト積スコア50は、一端が100%一致し、50%重畳しているか、他端が79%一致し、100%重畳しているかのいずれかの場合に得られる。

【0249】

或いは、SATをコードするポリヌクレオチド配列は、由来する組織に対して分析する。例えば或る完全長配列は、Incyte cDNA配列(実施例3を参照)と少なくとも一部は重畳するように構築される。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。各cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。各ヒト組織は、以下の生物/組織カテゴリー即ち心血管系、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、血液及び免疫系、肝、筋骨格系、神経系、脾臓、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。同様に、各ヒト組織は、以下の疾患/病状カテゴリー即ち癌、細胞系、発達、炎症、神経性、外傷、心血管、鬱血、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、SATをコードするcDNAの疾患特異的な発現を反映する。cDNA配列およびcDNAライブラリ/組織の情報は、LIFESEQ GOLD データベース(Incyte Genomics, Palo Alto CA)から得ることができる。

【0250】

8 SATをコードするポリヌクレオチドの伸長  
完全長のポリヌクレオチド配列もまた、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。一方の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、他方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。開始プライマーは、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68～72の温度で標的配列にアニーリングするように、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て、cDNAから設計した。ヘアピン構造及びプライマー-プライマー二量体を生ずるようなヌクレオチドの伸長は全て回避した。

【0251】

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、付加的プライマー或いはプライマーのネステッドセット

を設計した。

【0252】

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穴プレート内で行った。反応混合液は、鋳型DNA及び200 nmolの各プライマー、 $Mg^{2+}$ と $(NH_4)_2SO_4$ とβ-メルカプトエタノールを含むバッファー、Taq DNAポリメラーゼ(Amersham Pharmacia Biotech)、ELONGASE酵素(Life Technologies)、Pfu DNAポリメラーゼ(Stratagene)を含む。プライマーの組、PCI AとPCI Bに対して以下のパラメータで増幅を行った。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、及び4を20回繰り返す。ステップ6:68で5分間、ステップ7:4で保存。別法では、プライマー対、T7とSK+に対して以下のパラメータで増幅を行った。ステップ1:94で3分間、ステップ2:94で15秒間、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、及び4を20回繰り返す。ステップ6:68で5分間、ステップ7:4で保存。

【0253】

各ウェルのDNA濃度は、1X TE及び0.5 µlの希釈していないPCR産物に溶解した100 µlのPICO GREEN定量試薬(0.25 (v/v) PICO GREEN; Molecular Probes, Eugene OR)を不透明な蛍光光度計プレート(Coming Costar, Acton MA)の各ウェルに分配してDNAが試薬と結合できるようにして測定する。サンプルの蛍光を計測してDNAの濃度を定量するべくプレートをFluoroskan II (Labsystems Oy, Helsinki, Finland)でスキャンした。反応混合物のアリコート5~10 µlを1%アガロースミニ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決定した。

【0254】

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コラウイルスエンドヌクレアーゼ(Molecular Biology Research, Madison WI)を用いて消化し、pUC18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。シヨットガン・シークエンシングのために、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pUC18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張出部(overhang)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含む培地に移し、それぞれのコロニーを切りとってLB/2Xカルベニシリン培養液の384ウェルプレートに37で一晚培養した。

【0255】

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、及び4を20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PICO GREEN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

積し、DYENAMIC energy transfer sequencing primer及びDYENAMIC DIRECT kit (Amersham Pharmacia Biotech)またはABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems)を用いてシーケンシングした。

#### 【0256】

同様に、上記手順を用いて完全長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。

#### 【0257】

9 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: 10 - 18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事実上同一の手順が用いられる。オリゴヌクレオチドは、OLIGO 4.06ソフトウェア (National Biosciences)等の最新ソフトウェアを用いて設計し、各オリゴマー50 pmolと、[ $\gamma$ - $^{32}$ P]アデノシン3リン酸 (Amersham Pharmacia Biotech) 250  $\mu$ Ciと、T4ポリヌクレオチドキナーゼ (DuPont NEN, Boston MA)を結合す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)を用いて十分に精製する。Ase I、Bgl II、Eco RI、Pst I、Xba I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<sup>7</sup>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。

#### 【0258】

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 (Nyttran Plus, Schleicher & Schuell, Durham NH)に移す。ハイブリダイゼーションは、40°Cで16時間行う。非特異的シグナルを除去するため、例えば0.1Mクエン酸ナトリウム食塩水及び0.5%ドデシル硫酸ナトリウム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。

#### 【0259】

10 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連鎖または合成は、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Baldeschweiler等を参照)、機械的マイクロスポッティング技術及びこれらから派生したものを用いて達成することが可能である。上記各技術において基質は、均一且つ非多孔性の固体とするべきである (Schena (1999), 前出)。推奨する基質には、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質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、手作業で、または利用可能な方法や機械を用いて作製でき、任意の適正数のエレメントを有し得る (Schena, M. ら (1995) Science 270: 467 - 470、Shalon, D. ら (1996) Genome Res. 6: 639 - 645、Marshall, A. 及び J. Hodgson (1998) Nat. Biotechnol. 16: 27 - 31. を参照)。

#### 【0260】

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

10

20

30

40

50

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、レーザGENE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。アレイエレメントは、生物学的サンプル中でポリヌクレオチドを用いてハイブリダイズされる。生物学的サンプル中のポリヌクレオチドは、検出を容易にするために蛍光標識またはその他の分子タグに抱合される。ハイブリダイゼーション後、生物学的サンプルから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においてハイブリダイゼーションを検出する。或いは、レーザ脱着及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及び相対存在度は、算定し得る。一実施例におけるマイクロアレイの調整及び使用について、以下に詳述する。

10

#### 【0261】

##### 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)セルソース法を用いてポリ(A)<sup>+</sup>RNAを精製する。各ポリ(A)<sup>+</sup>RNAサンプルは、MMLV逆転写酵素、0.05 pg/μlのオリゴ(dT)プライマー(21mer)、1×第1鎖緩衝液、0.03 unit/μlのRNアーゼ阻害因子、500 μMのdATP、500 μMのdGTP、500 μMのdTTP、40 μMのdCTP、40 μM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)を用いて200 ngのポリ(A)<sup>+</sup>RNA含有の25体積ml  
内で行う。特異制御ポリ(A)<sup>+</sup>RNAは、370 で2時間インキュベートした後、i  
n v i t r o転写により非コード酵母ゲノムDNAから合成する。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(CLONTE  
CH Laboratories, Inc. (CLONTECH), Palo Alto CA)を用いて精製する。結合後、2つの反応サンプルは、1 mlのグリコーゲン(  
1 mg/ml)を用いて析出させたエタノール、60 mlの酢酸ナトリウム及び300 ml  
の100%エタノールである。サンプルは次に、Speed VAC(Savant I  
nstruments Inc., Holbrook NY)を用いて乾燥して仕上げ、  
14 μlの5×SSC/0.2% SDS中で再懸濁する。

20

30

#### 【0262】

##### 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。増幅されたアレイエレメントは、SEPHACRYL-400(Amersham Pharmacia Biotech)を用いて精製される。

40

#### 【0263】

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、処理中及び処理後に0.1%のSDS及びアセトン中で超音波をかけ、蒸留水で非常に良く洗って洗浄する。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、蒸留水中で広範囲にわたって洗浄し、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ガラスは、110 のオープンで硬化させる。

#### 【0264】

50

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

#### 【0265】

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。

10

#### 【0266】

#### ハイブリダイゼーション

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0.2μg含む9μl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1.8cm<sup>2</sup>のカバーガラスで覆う。アレイは、顕微鏡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140μl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。

20

#### 【0267】

#### 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、Cy35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10Wレーザ(Coherent, Inc., Santa Clara CA)を備えた顕微鏡で検出する。20×顕微鏡対物レンズ(Nikon, Inc., Melville NY)を用いて、アレイ上に励起レーザ光の焦点を当てる。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、対物レンズを通過してラスタースキャンする。本実施例で用いた1.8cm×1.8cmのアレイは、20μmの解像度でスキャンした。

30

#### 【0268】

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。アレイと光電子増倍管間に設置された好適なフィルターを用いて、シグナルをフィルターする。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、レーザ源において好適なフィルターを用いて、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。

40

#### 【0269】

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較正する。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、その較正を、2つの蛍光色素で較正

50

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つて行う。

#### 【0270】

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起因する) を補正する。

10

#### 【0271】

グリッドが蛍光シグナルイメージ上に重ねられ、それによつ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、GEMTOOLS 遺伝子発現分析プログラム (Incycyte) である。

#### 【0272】

##### 11 相補的ポリヌクレオチド

SAT をコードする配列或いはその任意の一部に対して相補的な配列は、天然の SAT の発現を低下させるため即ち阻害するために用いられる。約 15 ~ 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、これより小さな或いは大きな配列の断片の場合でも本質的に同じ方法を用いることができる。Oligo 4.06 ソフトウェア (National Biosciences) 及び SAT 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阻害する。翻訳を阻害するためには、相補的なオリゴヌクレオチドを設計して、リボソームが SAT をコードする転写物に結合するのを阻害する。

20

#### 【0273】

##### 12 SAT の発現

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

30

40

#### 【0274】

50

殆どの発現系では、S A Tが、例えばグルタチオンSトランスフェラーゼ (G S T)、またはF L A Gや6 - H i sなどのペプチドエピトープ標識で合成された融合タンパク質となるため、未精製の細胞溶解物からの組換え融合タンパク質の親和性ベースの精製が素早く1回で行うことができる。G S Tは日本住血吸虫からの26 k D aの酵素であり、タンパク質の活性及び抗原性を維持した状態で、固定化グルタチオン上で融合タンパク質の精製を可能とする (A m e r s h a m P h a r m a c i a B i o t e c h)。精製の後、G S T部分を特定の操作部位でS A Tからタンパク質分解的に切断できる。F L A G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F L A G抗体 (E a s t m a n K o d a k) 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 - H i sは、金属キレート樹脂 (Q I A G E N) 上での精製を可能にする。タンパク質の発現及び精製の方法は、前出のA u s u b e l (1995) 10章、16章に記載されている。これらの方法で精製したS A Tを直接用いて以下の実施例16および17のアッセイを行うことができる。

10

## 【0275】

## 13 機能的アッセイ

S A Tの機能は、哺乳動物細胞培養系において生理学的に高められたレベルでのS A Tをコードする配列の発現によって評価する。c D N Aを、c D N Aを高いレベルで発現する強いプロモーターを含む哺乳動物発現ベクターにサブクローニングする。選り抜きのベクターには、p C M V S P O R Tプラスミド (L i f e T e c h n o l o g i e s) 及びp C R 3 . 1プラスミド (I n v i t r o g e n) 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。リポソーム製剤或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時的に形質移入する。更に、標識タンパク質をコードする配列を含む1 ~ 2 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、c D N A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 (G F P ; C l o n t e c h)、C D 6 4またはC D 6 4 - G F P融合タンパク質から選択できる。自動化された、レーザー光学に基づく技術であるフローサイトメトリー (F C M) を用いて、G F PまたはC D 6 4 - G F Pを発現する形質移入された細胞を同定し、その細胞のアポトーシス状態や他の細胞特性を評価する。F C M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を検出して計量する。このような現象として挙げられるのは、プロピジウムヨウ化物によるD N A染色によって計測される核D N A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量の低下によって計測されるD N A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、O r m e r o d , M . G . (1994) Flow Cytometry Oxford, New York, N Y. に記述がある。

20

30

## 【0276】

遺伝子発現におけるS A Tの影響は、S A Tをコードする配列とC D 6 4またはC D 6 4 - G F P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。C D 6 4またはC D 6 4 - G F P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 (I g G) の保存領域と結合する。形質転換された細胞と形質転換されない細胞とは、ヒトI g GがC D 6 4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 (D Y N A L . L a k e S u c c e s s . N Y)。m R N A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。S A T及び目的の他の遺伝子をコードするm R N A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。

40

## 【0277】

50

#### 14 SAT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SAT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。

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

10

##### 【0278】

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、免疫抗原性を高める (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-KLM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SAT活性を検査するには、ペプチドまたはSATを基板に結合し、1%BSAを用いてブロッキング処理し、ウサギ抗血清と反応させて洗浄し、さらに放射性ヨウ素標識されたヤギ抗ウサギIgGと反応させる。

20

##### 【0279】

#### 15 特異的な抗体を用いる天然SATの精製

天然SAT或いは組換えSATを、SATに特異的な抗体を用いるイムノアフィニティークロマトグラフィにより実質的に精製する。イムノアフィニティークラムは、CNBr-活性化SEPHAROSE (Amersham Pharmacia Biotech) のような活性化クロマトグラフィー用レジンと抗SAT抗体とを共有結合させることにより形成する。結合後に、製造者の使用説明書に従って樹脂をブロックし、洗浄する。

##### 【0280】

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

30

##### 【0281】

#### 16 SATと相互作用する分子の同定

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

40

##### 【0282】

別法では、SAT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。

##### 【0283】

SATはまた、高処理な方法で酵母2ハイブリッドシステムを使用するPATHCALL

50

ING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. ら (2000) 米国特許第6,057,101号)。(2000) 米国特許第6,057,101号)。

#### 【0284】

##### 17 SAT 活性の実証

SAT 活動は被覆小胞中でその包含により測定される。SAT は、SAT をコードする真核生物発現ベクターを備えた COS7、HeLa、若しくは CHO のようなほ乳類株化細胞を形質転換することで発現されうる。真核生物発現ベクターは市販されており、それらを細胞内に導入する技術は当業者には周知である。 - ガラクトシダーゼのような複数の標識遺伝子いずれか一つを発現させる少量の第二のプラスミドは、細胞へと同時形質転換され、異質な DNA を吸収し発現させるそれら細胞の迅速な同定を可能とする。細胞は、形質転換の後、株化細胞が SAT 及び 13 - ガラクトシダーゼを発現し蓄積するのに適した条件下で、48 - 72 時間に渡ってインキュベートされる。形質転換細胞は集められ、細胞溶解物は小胞構成のため分析される。加水分解できない形式の GTP である GTP S、ATP を再生成するシステムが、溶解物に加えられ、混合液は 10 分間 37 °C でインキュベートされる。これらの条件の下では、小胞の 90% 以上はコートしたままである (Orci, L. 他 (1989) Cell 56:357-368 参照)。輸送小胞は、ゴルジ体膜から塩によって放出され、ショ糖密度勾配の下にかけられ、分別物は SDS-PAGE により回収され分析される。SAT がクラスリン又は COP コーティマと同じ局在性を示すと、小胞形成中の SAT 活性を意味する。小胞形成の SAT の寄与は、GTP S 追加の前に溶解産物と SAT 特定の抗体をインキュベートすることにより確認できる。抗体は SAT に結合し、その活性を妨害し、これにより小胞形成を妨げる。或いは、SAT 活性は、小胞輸送経路を変更するその能力により測定できる。SAT で形質転換された細胞の小胞輸送を、蛍光顕微鏡を使用して調べる。小胞コートタンパク質、またはトランスフェリン又はマンノース - 6 - リン酸塩受容体などの典型的な小胞輸送基質に特異的な抗体は市販されている。ER、ゴルジ体、ペルオキシゾーム、エンドソーム、リゾソームおよび原形質膜などの種々の細胞成分が検査される。対照細胞と比較して、SAT で形質転換された細胞内の小胞の数および位置の変化は、SAT 活性の特性である。

#### 【0285】

当業者は、本発明の範囲及び精神から逸脱することなく本発明の記載した方法及びシステムの種々の改変を行い得る。本発明について説明するにあたり特定の好適実施例に関連して説明を行ったが、本発明の範囲が、そのような特定の実施例に不当に制限されるべきではないことを理解されたい。実際に、分子生物学または関連分野の専門家には明らかな、本明細書に記載されている本発明の実施方法の様々な改変は、特許請求の範囲内にあるものとする。

#### 【0286】

(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。

#### 【0287】

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 GenBank 相同体が一致する確率スコアも併せて示す。

#### 【0288】

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ヌクレオ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。

#### 【0289】

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

10

20

30

40

50

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。

【0290】

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。

【0291】

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。

【0292】

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。

【表1】

表 1

Incyte プロジェクト ID	ポリペプチド SEQ ID NO:	Incyte ポリペプチド ID	ポリスクレオチド SEQ ID NO:	Incyte ポリスクレオチド ID
1577952	1	1577952CD1	10	1577952CB1
4983705	2	4983705CD1	11	4983705CB1
1310465	3	1310465CD1	12	1310465CB1
4291779	4	4291779CD1	13	4291779CB1
4728247	5	4728247CD1	14	4728247CB1
7472259	6	7472259CD1	15	7472259CB1
7476740	7	7476740CD1	16	7476740CB1
7473774	8	7473774CD1	17	7473774CB1
7946329	9	7946329CD1	18	7946329CB1

10

20

30

40

【表 2】

表 2

ポリペプチド SBQ ID NO.	Incyte ポリペプチド ID	GenBank ID NO.	確率スコア	GenBank 相同体
2	4983705CDI	g3077703	2.90E-136	[カイウサギ] mitsugumin29 akeshima, H. 他 (1998) Biochem. J. 331:317-322.
3	1310465CDI	g777776	0	[ドブネズミ] 表皮エンドソーム糖タンパク質 Speelman, B.A. 他(1995) J. Biol. Chem. 270:1583-1588
4	4291779CDI	g3378150	2.30E-45	[トリパノソマ(Trypanosoma brucei rhodesiense)] リソソーム/エンドソーム膜タンパク質 P67 Kelley, R.J. 他 (1999) Mol. Biochem. Parasitol. 98:17-28
5	4728247CDI	g5926736	0	[ハツカネズミ] グラニニューフィリン-a (granophilin-a) ang, J. 他(1999) J. Biol. Chem. 274:28542-28548
6	7472259CDI	g8925888	3.70E-101	[ドブネズミ] RIM 結合タンパク質 1A Wang, Y. 他(2000) J. Biol. Chem. 275:20033-20044
7	7476740CDI	g8977950	5.60E-43	[分裂酵母 (Schizosaccharomyces pombe)] 推定 v-snare 結合タンパク質; UBA ドメイン
8	7473774CDI	g484296	0	[ドブネズミ] シナプトタグミン III Mizuta, M. 他(1994) J. Biol. Chem. 269:11675-11678
9	7946329CDI	g6136794	1.30E-221	[ハツカネズミ] シナプトタグミン XI Fukuda, M. 他(1999) J. Biol. Chem. 274:31421-31427

10

20

30

40

表 3 - 1

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
1	1577952CD1	315			signal_peptide:M1-A33 膜貫通ドメイン: S101-L121 Rhomboid family: Rhomboid:C59-M214	HMMER HMMER HMMER_PFAM
2	4983705CD1	272	S101 S14 S2 S262	N213	シナプトアイジン/シナプトボリン: DM02999 P20488 21-148:G36-V166 シナプトアイジン PD005837:R27-A268 シナプトアイジン/シナプトボリン BL00604A:L41-L95 BL00604B:T104-L133 BL00604C:V134-L165 BL00604E:C201-Q242 シナプトアイジン/シナプトボリン PR00220A:I38-T60 PR00220B:A62-H87 PR00220C:F117-N141 PR00220D:F149-G172 PR00220E:V216-F234 シナプトアイジン/シナプトボリン シグネチャ synaptop.prf:Q42-I89 膜貫通ドメイン: P114-Y136, I214-F234 シナプトアイジン/シナプトボリン: シナプトアイジン:R27-Q272	BLAST_DOMO BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_PRINTS PROFILESCAN HMMER HMMER_PFAM

10

20

30

40

【表 4】

表 3 - 2

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
3	1310465CD1	1217	S1103 S1183 S1188 S140 S318 S329 S350 S451 S457 S518 S815 S897 T118 T133 T188 T206 T272 T494 T639 T649 T739 T863 T877 T903	N204 N282 N340 N584 N637 N836	シグナルペプチド:M1-A23 シグナルペプチド:M1-G21 膜貫通ドメイン:V1156-G1174 MAM ドメイン C67-L223, T272-P426, T494-P645, C657-A810, C814-Q970, C974-Q1139 低密度リポタンパク質受容体ドメイン: A228-R268, D456-T493 LDL 受容体クラス A BL01209: C249-E261 低密度リポタンパク質受容体ドメイン PR00261: K240-E261 MAM ドメインシグネチャ PR000020: D662-E680, Y113-L129, E724-P735, V1099-A1113, G791-R804 MAM ドメイン DM01344 A55620 961-1128: L958-L1126, L798-G950, P64-L213, S656-G791, D257-G414 MAM ドメイン DM01344 A55620 618-796: R616-G794, C814-G950, Q970-G1117, A66-T187, E486-D624 MAM ドメイン DM01344 A55620 798-959: T795-A957, V945-G1117, G634-A797, C67-A212, D273-L402	SPScan HMMER HMME HMME-PFAM HMME-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-DOMO BLAST-DOMO

10

20

30

40

表 3 - 3

SEQ ID NO:	4	Incyte ポリペプチド ID 4291779CD1	アミノ酸残基数 589	潜在的リン酸化部位 S157 S172 S308 S456 S506 T208 T361 T399 Y196	潜在的グリコシル化部位 N110 N231 N436 N465 N515 N88	MAM ドメイン DM01344 A55620 300-418: P299-L418, A842-L961, P684-V801, P100-L216 尖端インドソーム糖タンパク質 PD140389:H269-V655, C230-V427 PD108201:E1130-P1217 PD108200:M1-F65 シグネチヤ配列、 ドメインおよびモチーフ	BLAST-DOMO  BLAST-PRODOM  分析方法及びデータベース  SPSCAN BLAST-PRODOM BLAST-PRODOM
SEQ ID NO:	5	4728247CD1	671	S11 S139 S147 S190 S193 S213 S217 S231 S235 S274 S289 S363 S393 S414 S467 S494 S502 S520 S55 S563 S59 S625 S635 S652 S74 T115 T131 T151 T212 T262 T357 T413 T419	N537 N579 N971	C2 domains: L373-E462, L528-R617 C2 ドメイン シグネチヤ/プロファイル: E498-M570, I360-K416 シナプトガミン シグネチヤ PR00399: I360-V375, P433-H448 C2 ドメイン シグネチヤ PR00360: A543-L555, K572-N585 C2 ドメイン DM00150 P40748 294-423: G358-E460, E527-L618	MOTIFS HMMER-PFAM PROFILESCAN BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO

10

20

30

40

【表 6】

表 3 - 4

			T427 T567 T600 T636 T79 T99 Y120		C2 F-XIノ DM00150 JC2473 249-373: G358-G459, L528-D634	BLAST-DOMO
					C2 F-XIノ DM00150 P46097 140-266: G358-G482, E524-E639	BLAST-DOMO
					C2 F-XIノ DM00150 P24507 233-362: G358-E460, E527-L618	BLAST-DOMO

10

20

30

40

【表 7】

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
6	7472259CD1	1519	S104 S1043 T843 S1104 T685 T740 T442 S1114 T390 T442 S1135 S1142 T57 S1182 T273 S897 S1312 S832 S813 S1443 S152 S163 T1450 S178 S186 T1405 S193 S202 T1273 S206 S225 T1199 S412 S568 T1086 S600 S628 T1080 S656 S660 T1026 S665 S687	N664 N883 N890	SH3 ドメイン: K715-I777 ロイシン ジッパー モチーフ: L253-L274, L260-L281, L297-L318 Src 相同性 3 (SH3) ドメイン: D763-Q776 ウイルスス腫瘍タンパク質シグネチャ: PR00049:R203-P217	HMMER-PFAM MOTIFS BLOCKS-BLIMPS BLIMPS-PRINTS
7	7476740CD1	396	S106 S319 S374 S380 S387 S391 T295		レトロウイルス アスハルチルプロテアーゼ Y245-A271, Q315-L344 DDI1 (DNA 複製誘導性タンパク質):PD017951:A203-H386	HMMER-PFAM BLAST-PRODOM

10

20

30

40

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
8	7473774CD1	590	S160 S166 S240 S245 S385 S393 S427 S488 S584 S75 T184 T223 T487 T493 T497	N365 N436	プロテインキナーゼ C 2 領域: DM03932 P40748 1-292:M1-G290 C2-ドメイン DM00150  P40748 294-423:G297-G426, A430-L554  P40748 425-552:S427-A555, A298-V423  P24507 364-491:S427-N556, A298-V423 シナプトタグミン 膜貫通リピート シナプス III SYTIII VI C シナプス小胞体 PD022173:F142-Q315 シナプトタグミン膜貫通リピート シナプス III SYTIII VI C シナプス小胞体タンパク質 PD012608:M1-P102 PD007398:R536-S580 タンパク質 C リポート シナプトタグミン ホスホリパーゼ 膜貫通 シナプス結合 ホルボールエステル キナーゼ PD000136:L316-Q400, L448-C535 C2 ドメインタンパク質 PF00168:L458-D468, L475-E500 シナプトタグミン シグネチャ PF00399:P373-D388, S393-L403, I303-V318, V318-S331 C2 ドメイン シグネチャ PR00360:S331-L343, K358-S371, L382-D390	BLAST_DOMO BLAST_DOMO BLAST_PRODUM BLAST_PRODUM BLAST_PRODUM BLIMPS_PFBAM BLIMPS_PRINTS BLIMPS_PRINTS

10

20

30

40

【表 9】

表 3 - 7

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		HMMER
9	7946329CDI	431	S117 S150 S151 S253 S291 S406 S417 S427 S70 T140 T177 T272 T348 T354 T387 T388	N360	膜貫通ドメイン: I51-V74 C2 ドメイン C2:L316-V402, L448-R536 C2 ドメイン シグネチャ gcg_motif: A29-Y338 A455-Y470 C2 ドメイン シグネチャ及びプロファイル c2_domain.prf: L435-K490, I303-K358 シグネチャ配列、ドメインおよびモチーフ	HMMER HMMER.PFAM MOTIFS PROFILESCAN 分析方法及びデータベース
					C2-ドメイン DM00150 I59355 151-280:L158-N285, G293-S406  P40749 151-280:L158-N285, G293-S406 I59355 282-412:Q287-R419  P40749 282-412:Q287-R419 シナプトタグミン IV 膜貫通リポドット XI PD022115:M1-V175 PD151917:E260-Q302 シナプトタグミン XI PD163942:L396-Y431 タンパク質 C リポドット シナプトタグミン ホスホリパーゼ膜貫通 シナプス 結合 ホルモンレスチルキナーゼ PD000136:L174-E260, M308-E391 シナプトタグミン シグネチャ PR00399:V310-I323, P233-S248, S253-V263, L161-V176 C2 ドメイン シグネチャ PR00360:Q190-I202, K217-Y230, L242-D250	BLAST_DOMO BLAST_PRODUM BLAST_PRODUM BLIMPS_PRINTS BLIMPS_PRINTS

10

20

30

40

表 3 - 8

膜貫通ドメイン: V16-W37	HMMER
C2ドメイン C2	HMMER_PFAM
L174-M262, M308-I397	
C2ドメイン シグネチャ及びプロファイル c2_domain.prf:	PROFILESCAN
L295-K351, L161-K217	
SpSCAN signal_cleavage: M1-S38	SPSCAN

10

20

30

40

【表 1 1】

表 4 - 1

ポリヌクレオチド SEQ ID NO:	Incyte ポリヌクレオチド ID	配列長	選択断片	配列断片	5'位置	3'位置
10	1577952CB1	3424	1-39 2099-2292 3272-3424	70687032V1	2938	3424
				7438458H1 (ADRETUE02)	2439	3019
				71111635V1	574	1240
				70647387V1	2364	2984
				8009276H1 (NOSEDIC02)	1529	2189
				71112240V1	665	1360
				71114546V1	1294	1881
				7993001H1 (UTRSDIC01)	1906	2435
				70767606V1	1	627
				1438701F1 (PANCNOT08)	1361	1899
11	4983705CB1	1033	732-1033	GBI:G6524214_00026.raw.1	674	1033
				95810426	1	445
				4983705H1 (HELATXT05)	266	535
				6901767R8 (MUSLTDR02)	291	867
				7610385H1 (KIDCTME01)	39	305
12	1310465CB1	3902	1-1452 1764-2846	7164475F8 (PLACNOR01)	416	1016
				71991186V1	2676	3281
				7714340H1 (SINTFEE02)	1136	1732
				7714340J1 (SINTFEE02)	1433	2056
				8117813H1 (TONSDIC01)	2011	2714
				6542726H1 (LNODNON02)	3386	3902
				71987767V1	2777	3431
				71991460V1	2136	2767
				7080102F6 (STOMTMR02)	466	1222
				7693543H2 (LNODTUE01)	1	438

10

20

30

40

表 4-2

ポリヌクレオチド SEQ ID NO:	Incyte ポリヌクレオチド ID	配列長	選択断片	配列断片	5'位置	3'位置
13	4291779CB1	2574	979-1592	70735272V1	1460	2082
				2526437F6 (BRAITUT21)	1720	2198
				6154207H1 (ENDMUNT04)	1341	1668
				6996766H1 (BRAKTR17)	635	1150
				7061302H1 (PENITMN02)	946	1663
				1869872F6 (SKINBIT01)	2185	2574
				70738661V1	1973	2557
				7278956H1 (EMARTXE01)	1	565
				4822574F9 (PROSTUT17)	479	937
				2079722F6 (UTRSNOT08)	1964	2462
14	4728247CB1	2878	539-2368	70867423V1	2297	2878
				71171848V1	1383	1966
				71399565V1	641	1331
				224668R6 (PANCNOT01)	369	868
				3354730F6 (PROSNOT28)	1	579
				224668T6 (PANCNOT01)	1831	2455
				7631669H1 (BRAFTUE03)	1309	1902
				7396961H1 (KIDEUNE02)	1374	2096
				7761273H1 (THYMNOE02)	1	498
				6842194H1 (BRSTNON02)	4134	4726
15	7472259CB1	5628	1-2080 5302-5628 2575-4411	2394503F6 (THPIAZT01)	5028	5628
				91963660	4411	4860
				92063588	4767	5235
				GNN_96006353_010	1	4922

10

20

30

40

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	選択断片	配列断片	5'位置	3'位置
16	7476740CB1	1482	1-566 750-831	GNN.G7547222_000015_002	175	1365
				92064117	946	1482
				6247775F8 (TESTNOT17)	1	755
17	74737740B1	2511	1-113 639-1330 2471-2511	70555497V1	1470	2036
				72049138V1	716	1538
				7585860H2 (BRAIFEC01)	1	630
				71045227V1	1640	2282
				72050239V1	587	1476
18	7946329CB1	1680		71539112V1	2184	2511
				7228832H1 (BRAXTDR15)	1	480
				71801792V1	460	1167
				71801805V1	607	1231
				6996340H1 (BRAXTDR17)	1091	1680

10

20

30

40

表 5

ポリヌクレオチド SEQ ID NO:	Incyte プロジェクト ID	代表的ライブラリ
10	1577952CB1	LIVRTUT13
11	4983705CB1	BRAFNOT02
12	1310465CB1	SINTFEE02
13	4291779CB1	BRAITUT21
14	4728247CB1	PANCN0T01
15	7472259CB1	THELAZT01
16	7476740CB1	TESTNOT17
17	7473774CB1	BRAIFEC01
18	7946329CB1	SCOMDICO1

10

20

30

40

【表 1 5】

表 6 - 1

ライブラリ	ベクター	ライブラリの説明
BRAIFEC01	pINCY	この大型分画のライブラリは、妊娠 23 週間で死産の左心低形成症胎児(白人男性)から除去された脳組織から単離した RNA を用いて作製した。
BRAITUT21	pINCY	ライブラリは、61 才の白人女性の脳腫瘍病変の切除時、正中線前頭葉から採取した脳腫瘍組織から単離した RNA を用いて作製した。病理は非定型を伴わない前頭下髄膜皮髄腫(subfrontal meningothelial meningioma)を示した。1 つの篩骨と粘膜炎組織サンプルは髄膜腫を示した。家族歴には、脳血管疾患、老人性痴呆症、高脂血症、良性の高血圧、アテローム硬化型冠状動脈疾患、鬱血性心不全および乳がんが含まれる。
BRAFNOT02	pINCY	ライブラリは、摘出大動脈瘤および虚血の腸疾病で死亡の 45 才白人女性から得た線条体、淡蒼球および被殻組織から単離した RNA を用いて作製した。病理学的に、大脳皮質の白質および基底核を含む穏やかな動脈硬化症を示した。肉眼で見ても、脊椎の動脈ばかりでなく中央の大脳動脈に軽度の髄膜繊維症および軽度の病巣のアテローム性動脈硬化のラクが双方にあった。微視的には、大脳半球、脳幹および小脳は、構形的に血管を含んでいて白質中の病巣の部位を明らかにした。このとき、フィリヒョーロペンソベースにヘモジデリン含有マクロファージがあった。さらに、扁桃の側底核内に神経原線維濃縮体が散在した。患者の病歴は、動脈瘤により大動脈と冠状動脈の軽度のアテロマトシス、腸および肝臓梗塞部、肥満に関連した生理学的な脂肪肝、軽度の広汎性肺気腫、腸間膜・門脈の血栓症、左心室の異常発達による心臓肥大症、動脈性の高血圧症、急性肺水腫、脾腫、肥満(300 ポンド)、子宮平滑筋腫、睡眠無呼吸、鉄分欠乏貧血があった。LIVRTUT13
LIVRTUT13	pINCY	ライブラリは 62 才の白人女性の部分肝切除と診査開腹術時に採取した肝腫瘍組織から単離した RNA を用いて作製した。病理は、島細胞腫と一致した転移性で中グレードの神経内分泌悪性腫瘍を示し、腫瘍は、内側と外側の左肝葉でサイズの異なる結節を形成していた。脾臓には、偽性嚢胞と一致する、繊維症、慢性炎症及び脂肪壊死が見られた。胆嚢には軽度の慢性胆嚢炎が見られた。患者の病歴には、腓骨部の悪性腫瘍、肺塞栓症、高脂血症、血栓性脈管炎、複数の関節における関節痛、II 型糖尿病、良性高血圧症、脳血管疾患が含まれる。家族の病歴には、脾臓癌、二次性肝臓癌、良性高血圧症、高脂血症が見られた。
PANCN0T01	PBLUESCRIPT	ライブラリは、脳外傷で死亡した 29 才白人男子の脾臓組織から単離された RNA を用いて作製された。
SCOMDIC01	PSPORT1	この大型分画ライブラリは脳血管事故により死亡した 57 才の白人男性の延髄底部から採取した罹病脊髄組織から単離した RNA を用いて作製した。血清検査は陰性だった。患者の病歴は、ハンチントン病、気腫およびタバコ乱用(40 年間 1 日 3~4 箱)などであった。
SINTFEE02	PCDNA2.1	この 5'偏向無作為ライブラリは妊娠 20 週にてパーセントリソミーで死亡した白人男胎児から採取した小

10

20

30

40

【表 16】

表 6 - 2

ライブラリ	ベクター	ライブラリの説明
TESTNOT17	pINCY	腸組織から単離した RNA を用いて作製した。血清学は陰生であった。 ライブラリは、自動車事故による頭部外傷で死亡した 26 才の白人男性から取り除いた精巢組織から作成された。血清検査は陰性だった。患者の病歴は、誕生時のヘルニア、喫煙 (1.5 ppd)、マリファナ吸引および毎日の飲酒(ビールと蒸留酒)があった。THP1AZT01
THP1AZT01	pINCY	ライブラリは、0.8 ミクロモルで3日間処理されたTHP-1プロモノサイト細胞から単離したRNAを用いて作製した。THP-1(ATCC TIB 202)は、急性単球性白血病を持った 1 才の白人男性の末梢血液から由来したヒトの単球支持ラインである。( Int. J. Cancer (1980) 26:171).

10

20

30

40

【表 17】

【表 18】

プログラム	説明	引用文献	パラメータ閾値
ABI FACTURA	核酸配列においてベクター配列を除去して、不定の塩基をマスクするプログラム	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	Fast Data Finderは、アミノ酸または核酸配列の比較、注釈付けに有用である。	Applied Biosystems, Foster City, CA. Paracel Inc., Pasadena, CA.	不一致<50%
ABI オートアセンブラ	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) (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: 確率値=1.0E-8 以下 完全長配列: 確率値= 1.0E-10以下
FASTA	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 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; および Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E 値=1.06E-6 構築された ESTs: fasta 同一性=95%以上 一致した長さ=200塩基以上 fastx E 値=1.0E-8 以下 完全長配列: fastx スコア=100以上
BLIMPS	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アインゲンブリント領域を検索するBLOCKS IMPROVED サーチャャー。	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; および Atwood, T.K. 他. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	確率値= 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 (1988) Nucleic Acids Res. 26:320-322; Durbin, R. 他. (1998) Our World View, in a Nutshell. Cambridge Univ. Press, pp. 1-350.	PFAM ヒット: 確率値= 1.0E-3 以下 シグナルペプチド ヒット: スコア =0以上

プログラム	説明	引用文献	パラメータ閾値
ProfileScan	Prosit 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; Bairoch, 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 Prosit 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ト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. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的なインプリメンテーションに基づく SWAT や CrossMatch を含む Phils Revised Assembly プログラムで、配列相同性の検索や 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; and 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. and 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蛋白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) を使って蛋白配列の膜貫通セグメントを描写し、配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., Biol., Glasgow 他, eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	Prosit で定義された配列と一致したパターン氨基酸配列を検索するプログラム。	Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, 第9版, M51-59頁, Genetics Computer Group, Madison, WI	

10

20

30

40

## 【国際公開パンフレット】

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

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number  
WO 02/02610 A2

(51) International Patent Classification: C07K 14/435

(21) International Application Number: PCT/US01/20704

(22) International Filing Date: 28 June 2001 (28.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60215.465 29 June 2000 (29.06.2000) US  
60239.384 10 October 2000 (10.10.2000) US  
60253.639 28 November 2000 (28.11.2000) US(71) Applicant (for all designated States except US): INCYTE  
GENOMICS, INC. [US/US], 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

(72) Inventors and

(75) Inventors/Applicants (for US only): LEE, Ernestine,  
A. [US/US], 624 Kaine Street, Albany, CA 94706 (US).  
LU, Yan [CN/US], 3885 Corina Way, Palo Alto, CA  
94303 (US). LAL, Preeti [IN/US], P.O. Box 5142, Santa  
Clara, CA 95056 (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).  
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). DAS, Debopriya [IN/US], 1179 Bonita Avenue,  
#3, Mountain View, CA 94040 (US). RAMKUMAR,  
Jayalaxmi [IN/US], 34359 Maybird Circle, Fremont, CA  
94555 (US). TRIBOULEY, Catherine, M. [FR/US],  
1121 Tennessee Street, #5, San Francisco, CA 94107 (US).  
LU, Dyang, Aina, M. [US/US], 233 Coy Drive, San Jose,  
CA 95123 (US). HAFALIA, April [US/US], 2227 Calle  
de Primavera, Santa Clara, CA 95054 (US). GANDHI,  
Ameena, R. [US/US], 837 Roble Avenue, #1, Menlo Park,  
CA 94025 (US). XU, Yanning [US/US], 1739 Walnut  
Drive, Mountain View, CA 94040 (US). BANDMAN,  
Olga [US/US], 366 Anna Avenue, Mountain View, CA94043 (US). ELLIOTT, Vicki, S. [US/US], 3770 Polton  
Place Way, San Jose, CA 95121 (US). NGUYEN, Dan-  
niel, B. [US/US], 1403 Ridgewood Drive, San Jose, CA  
95118 (US). BURRILL, John, D. [GB/US], 650 Emerald  
Hill Road, Redwood City, CA 94061 (US). MARCUS,  
Gregory, A. [US/US], 73 Maple Way, San Carlos, CA  
94070 (US). ZINGLER, Kurt, A. [US/US], 723 Ashbury  
Street #C, San Francisco, CA 94117 (US). YAO, Montique,  
G. [US/US], 111 Frederick Court, Mountain View, CA  
94043 (US). GURURAJAN, Rajagopal [US/US], 5591  
Dent Avenue, San Jose, CA 95118 (US). DING, Li  
[CN/US], 3353 Alma Street, #146, Palo Alto, CA 94306  
(US). WARREN, Bridget, A. [US/US], 10130 Parwood  
Drive #2, Cupertino, CA 95014 (US). THANGAVELU,  
Kavitha [IN/US], 1950 Montecito Avenue, #23, Mountain  
View, CA 94043 (US). LEE, Sally [US/US], 825 East  
Evelyn, #425, Sunnyvale, CA 94086 (US).(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,  
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,  
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,  
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,  
TH, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished  
upon receipt of that reportFor two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: SECRETION AND TRAFFICKING MOLECULES

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



WO 02/02610 A2

WO 02/02610

PCT/US01/20704

## SECRETION AND TRAFFICKING MOLECULES

## TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of secretion and trafficking  
5 molecules and to the use of these sequences in the diagnosis, treatment, and prevention of vesicle  
trafficking, transport, neurological, autoimmune/inflammatory, and cell proliferative disorders, and in  
the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino  
acid sequences of secretion and trafficking molecules.

## 10 BACKGROUND OF THE INVENTION

Eukaryotic cells are bound by a lipid bilayer membrane and subdivided into functionally  
distinct, membrane-bound compartments. The membranes maintain the essential differences between  
the cytosol, the extracellular environment, and the luminal space of each intracellular organelle.  
Eukaryotic proteins including integral membrane proteins, secreted proteins, and proteins destined for  
15 the lumen of organelles are synthesized within the endoplasmic reticulum (ER), delivered to the Golgi  
complex for post-translational processing and sorting, and then transported to specific intracellular and  
extracellular destinations. Material is internalized from the extracellular environment by endocytosis, a  
process essential for transmission of neuronal, metabolic, and proliferative signals; uptake of many  
essential nutrients; and defense against invading organisms. This intracellular and extracellular  
20 movement of protein molecules is termed vesicle trafficking. Trafficking is accomplished by the  
packaging of protein molecules into specialized vesicles which bud from the donor organelle  
membrane and fuse to the target membrane (Rothman, J.E and Wieland, F.T. (1996) *Science*  
*272:227-234*).

The transport of proteins across the ER membrane involves a process that is similar in  
25 bacteria, yeast, and mammals (Gorlich, D. et al. (1992) *Cell* 71: 489-503). In mammalian systems,  
transport is initiated by the action of a cytoplasmic signal recognition particle (SRP) which recognizes  
a signal sequence on a growing, nascent polypeptide and binds the polypeptide and its ribosome  
complex to the ER membrane through an SRP receptor located on the ER membrane. The signal  
peptide is cleaved and the ribosome complex, together with the attached polypeptide, becomes  
30 membrane bound. The polypeptide is subsequently translocated across the ER membrane and into a  
vesicle (Blobel, G. and B. Dobberstein (1975) *J. Cell Biol.* 67:852-862).

Proteins implicated in the translocation of polypeptides across the ER membrane in yeast  
include SEC61p, SEC62p, and SEC63p. Mutations in the genes encoding these proteins lead to  
defects in the translocation process. SEC61 may be of particular importance since certain mutations

WO 02/02610

PCT/US01/20704

in the gene for this protein inhibit the translocation of many proteins (Gorlich, *supra*).

Mammalian homologs of yeast SEC61 (mSEC61) have been identified in dog and rat (Gorlich, *supra*). Mammalian SEC61 is also structurally similar to SECYp, the bacterial cytoplasmic membrane translocation protein. mSEC61 is found in tight association with membrane-bound ribosomes. This association is induced by membrane-targeting of nascent polypeptide chains and is weakened by dissociation of the ribosomes into their constituent subunits. mSEC61 is postulated to be a component of a putative protein-conducting channel, located in the ER membrane, to which nascent polypeptides are transferred following the completion of translation by ribosomes (Gorlich, *supra*).

Several steps in the transit of material along the secretory and endocytic pathways require the formation of transport vesicles. Specifically, vesicles form at the transitional endoplasmic reticulum (tER), the rim of Golgi cisternae, the face of the Trans-Golgi Network (TGN), the plasma membrane (PM), and tubular extensions of the endosomes. Vesicle formation occurs when a region of membrane buds off from the donor organelle. The membrane-bound vesicle contains proteins to be transported and is surrounded by a proteinaceous coat, the components of which are recruited from the cytosol. The initial budding and coating processes are controlled by a cytosolic ras-like GTP-binding protein, ADP-ribosylating factor (Arf), and adapter proteins (AP). Cytosolic GTP-bound Arf is also incorporated into the vesicle as it forms. Different isoforms of both Arf and AP are involved at different sites of budding. For example, Arfs 1, 3, and 5 are required for Golgi budding, Arf4 for endosomal budding, and Arf6 for plasma membrane budding. Two different classes of coat protein have also been identified. Clathrin coats form on vesicles derived from the TGN and PM, whereas coatamer (COP) coats form on vesicles derived from the ER and Golgi (Mellman, I. (1996) *Annu. Rev. Cell Dev. Biol.* 12:575-625).

In clathrin-based vesicle formation, APs bring vesicle cargo and coat proteins together at the surface of the budding membrane. APs are heterotetrameric complexes composed of two large chains: one chain comprised of an  $\alpha$ ,  $\gamma$ ,  $\delta$ , or  $\epsilon$  chain with a  $\beta$  chain, a medium chain ( $\mu$ ), and a small chain ( $\sigma$ ). Clathrin binds to APs via the carboxy-terminal appendage domain of the  $\beta$ -adaptin subunit (Le Bourgne, R. and Hoflack, B. (1998) *Curr. Opin. Cell. Biol.* 10:499-503). AP-1 functions in protein sorting from the TGN and endosomes to compartments of the endosomal/lysosomal system. AP-2 functions in clathrin-mediated endocytosis at the plasma membrane, while AP-3 is associated with endosomes and/or the TGN and recruits integral membrane proteins for transport to lysosomes and lysosome-related organelles. The recently isolated AP-4 complex localizes to the TGN or a neighboring compartment and may play a role in sorting events thought to take place in post-Golgi compartments (Dell'Angelica, E. C. et al. (1999) *J. Biol. Chem.* 274:7278-7285). Cytosolic GTP-bound Arf is also incorporated into the vesicle as it forms. Another GTP-binding protein, dynamin,

WO 02/02610

PCT/US01/20704

forms a ring complex around the neck of the forming vesicle and provides the mechanochemical force required to release the vesicle from the donor membrane. The coated vesicle complex is then transported through the cytosol. During the transport process, Arf-bound GTP is hydrolyzed to GDP and the coat dissociates from the transport vesicle (West, M.A. et al. (1997) *J. Cell Biol.* 138:1239-1254).

Coatamer (COP) coats, a second class of coat proteins, form on vesicles derived from the ER and Golgi. COP coats can further be classified as COPI, involved in retrograde traffic through the Golgi and from the Golgi to the ER, and COPII, involved in anterograde traffic from the ER to the Golgi (Mellman, *supra*). The COP coat consists of two major components, a GTP-binding protein (Arf or Sar) and coat protomer (coatamer). Coatamer is an equimolar complex of seven proteins, termed  $\alpha$ -,  $\beta$ -,  $\beta'$ -,  $\gamma$ -,  $\Delta$ -,  $\epsilon$ - and Z-COP. The coatamer complex binds to dilysine motifs contained on the cytoplasmic tails of integral membrane proteins. These include the dilysine-containing retrieval motif of membrane proteins of the ER and dibasic/diphenylamine motifs of members of the p24 family. The p24 family of type I membrane proteins represents the major membrane proteins of COPI vesicles. (Harter, C. and Wieland, F.T. (1998) *Proc. Natl. Acad. Sci. USA* 95:11649-11654.)

Vesicles can undergo homotypic, fusing with a same type vesicle, or heterotypic, fusing with a different type vesicle, fusion. Molecules required for appropriate targeting and fusion of vesicles include proteins in the vesicle membrane, the target membrane, and proteins recruited from the cytosol. During budding of the vesicle from the donor compartment, an integral membrane protein, VAMP (vesicle-associated membrane protein) is incorporated into the vesicle. Soon after the vesicle uncoats, a cytosolic prenylated GTP-binding protein, Rab, is inserted into the vesicle membrane. The amino acid sequence of Rab proteins reveals conserved GTP-binding domains characteristic of Ras superfamily members. In the vesicle membrane, GTP-bound Rab interacts with VAMP. Once the vesicle reaches the target membrane, a GTPase activating protein (GAP) in the target membrane converts the Rab protein to the GDP-bound form. A cytosolic protein, guanine-nucleotide dissociation inhibitor (GDI) then removes GDP-bound Rab from the vesicle membrane. Several Rab isoforms have been identified and appear to associate with specific compartments within the cell. For example, Rabs 4, 5, and 11 are associated with the early endosome, whereas Rabs 7 and 9 associate with the late endosome. These differences may provide selectivity in the association between vesicles and their target membranes. (Novick, P., and Zerial, M. (1997) *Cur. Opin. Cell Biol.* 9:496-504.)

Docking of the transport vesicle with the target membrane involves the formation of a complex between the vesicle SNAP receptor (v-SNARE), target membrane (t-) SNAREs, and certain other membrane and cytosolic proteins. Many of these other proteins have been identified although their exact functions in the docking complex remain uncertain (Tellam, J.T. et al. (1995) *J.*

WO 02/02610

PCT/US01/20704

Biol. Chem. 270:5857-5863; Hata, Y. and Sudhof, T.C. (1995) J. Biol. Chem. 270:13022-13028). N-ethylmaleimide sensitive factor (NSF) and soluble NSF-attachment protein ( $\alpha$ -SNAP and  $\beta$ -SNAP) are two such proteins that are conserved from yeast to man and function in most intracellular membrane fusion reactions. Sec1 represents a family of yeast proteins that function at many different stages in the secretory pathway including membrane fusion. Recently, mammalian homologs of Sec1, called Munc-18 proteins, have been identified (Katagiri, H. et al. (1995) J. Biol. Chem. 270:4963-4966; Hata et al. *supra*).

The SNARE complex involves three SNARE molecules, one in the vesicular membrane and two in the target membrane. Together they form a rod-shaped complex of four  $\alpha$ -helical coiled-coils. The membrane anchoring domains of all three SNAREs project from one end of the rod. This complex is similar to the rod-like structures formed by fusion proteins characteristic of the enveloped viruses, such as myxovirus, influenza, filovirus (Ebola), and the HIV and SIV retroviruses (Skehel, J.J., and Wiley, D.C. (1998) Cell 95:871-874). It has been proposed that the SNARE complex is sufficient for membrane fusion, suggesting that the proteins which associate with the complex provide regulation over the fusion event (Weber, T. et al. (1998) Cell 92:759-772). For example, in neurons, which exhibit regulated exocytosis, docked vesicles do not fuse with the presynaptic membrane until depolarization, which leads to an influx of calcium (Bennett, M.K., and Scheller, R.H. (1994) Annu. Rev. Biochem. 63:63-100). Synaptotagmin, an integral membrane protein in the synaptic vesicle, associates with the t-SNARE syntaxin in the docking complex. Synaptotagmin binds calcium in a complex with negatively charged phospholipids, which allows the cytosolic SNAP protein to displace synaptotagmin from syntaxin and fusion to occur. Thus, synaptotagmin is a negative regulator of fusion in the neuron. (Littleton, J.T. et al. (1993) Cell 74:1125-1134.)

The most abundant membrane protein of synaptic vesicles appears to be the glycoprotein synaptophysin, a 38 kDa protein with four transmembrane domains and two intravesicular loops. Synaptophysin monomers associate into homopolymers which form channels in the synaptic vesicle membrane. Synaptophysin's calcium-binding ability, tyrosine phosphorylation, and widespread distribution in neural tissues suggest a potential role in neurosecretion (Bennett, *supra*).

The transport of proteins into and out of vesicles relies on interactions between cell membranes and a supporting membrane cytoskeleton consisting of spectrin and other proteins. A large family of related proteins called ankyrins participate in the transport process by binding to the membrane skeleton protein spectrin and to a protein in the cell membrane called band 3, a component of an anion channel in the cell membrane. Ankyrins therefore function as a critical link between the cytoskeleton and the cell membrane.

Originally found in association with erythroid cells, ankyrins are also found in other tissues as

WO 02/02610

PCT/US01/20704

well (Birkenmeier, C.S. et al. (1993) *J. Biol. Chem.* 268:9533-9540). Ankyrins are large proteins (~1800 amino acids) containing an N-terminal, 89 kDa domain that binds the cell membrane proteins band 3 and tubulin, a central 62 kDa domain that binds the cytoskeletal proteins spectrin and vimentin, and a C-terminal, 55 kDa regulatory domain that functions as a modifier of the binding activities of the other two domains. Individual genes for ankyrin are able to produce multiple ankyrin isoforms by various insertions and deletions. These isoforms are of nearly identical size but may have different functions. In addition, smaller transcripts are produced which are missing large regions of the coding sequences from the N-terminal (band 3 binding), and central (spectrin binding) domains. The existence of such a large family of ankyrin proteins and the observation that more than one type of ankyrin may be expressed in the same cell type suggests that ankyrins may have more specialized functions than simply binding the membrane skeleton to the plasma membrane (Birkenmeier, *supra*).

In humans, two isoforms of ankyrin are expressed, alternatively, in developing erythroids and mature erythroids, respectively (Lambert, S. et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1730-1734). A deficiency in erythroid spectrin and ankyrin has been associated with the hemolytic anemia, hereditary spherocytosis (Coetzer, T.L. et al. (1988) *New Engl. J. Med.* 318:230-234).

Correct trafficking of proteins is of particular importance for the proper function of epithelial cells, which are polarized into distinct apical and basolateral domains containing different cell membrane components such as lipids and membrane-associated proteins. Certain proteins are flexible and may be sorted to the basolateral or apical side depending upon cell type or growth conditions. For example, the kidney anion exchanger (kAE1) can be retargeted from the apical to the basolateral domain if cells are cultured at higher density. The protein kanadaplin was isolated as a protein which binds to the cytoplasmic domain of kAE1. It also colocalizes with kAE1 in vesicles, but not in the membrane, suggesting that kanadaplin's function is to guide kAE1-containing vesicles to the basolateral target membrane (Chen, J. et al. (1998) *J. Biol. Chem.* 273:1038-1043).

Vesicle trafficking is crucial in the process of neurotransmission. Synaptic vesicles carry neurotransmitter molecules from the cytoplasm of a neuron to the synapse. Rab3's are a family of GTP-binding proteins located on synaptic vesicles. The RIM family of proteins are thought to be effectors for Rab3's (Wang, Y. et al. (2000) *J. Biol. Chem.* 275:20033-20044). Rabphilin-3 is a synaptic vesicle protein. Granuphilins are proteins with homology to rabphilins, and may have a unique role in exocytosis (Wang, J. et al. (1999) *J. Biol. Chem.* 274:28542-28548).

The etiology of numerous human diseases and disorders can be attributed to defects in the trafficking of proteins to organelles or the cell surface. Defects in the trafficking of membrane-bound receptors and ion channels are associated with cystic fibrosis (cystic fibrosis transmembrane conductance regulator; CFTR), glucose-galactose malabsorption syndrome (Na<sup>+</sup>/glucose

WO 02/02610

PCT/US01/20704

cotransporter), hypercholesterolemia (low-density lipoprotein (LDL) receptor), and forms of diabetes mellitus (insulin receptor). Abnormal hormonal secretion is linked to disorders including diabetes insipidus (vasopressin), hyper- and hypoglycemia (insulin, glucagon), Grave's disease and goiter (thyroid hormone), and Cushing's and Addison's diseases (adrenocorticotropic hormone; ACTH).

5 Cancer cells secrete excessive amounts of hormones or other biologically active peptides. Disorders related to excessive secretion of biologically active peptides by tumor cells include: fasting hypoglycemia due to increased insulin secretion from insulinoma-islet cell tumors; hypertension due to increased epinephrine and norepinephrine secreted from pheochromocytomas of the adrenal medulla and sympathetic paraganglia; and carcinoid syndrome, which includes abdominal cramps, diarrhea, and  
10 valvular heart disease, caused by excessive amounts of vasoactive substances (serotonin, bradykinin, histamine, prostaglandins, and polypeptide hormones) secreted from intestinal tumors. Ectopic synthesis and secretion of biologically active peptides (peptides not expected from a tumor) includes ACTH and vasopressin in lung and pancreatic cancers; parathyroid hormone in lung and bladder cancers; calcitonin in lung and breast cancers; and thyroid-stimulating hormone in medullary thyroid  
15 carcinoma.

Various human pathogens alter host cell protein trafficking pathways to their own advantage. For example, the HIV protein Nef down-regulates cell surface expression of CD4 molecules by accelerating their endocytosis through clathrin coated pits. This function of Nef is important for the spread of HIV from the infected cell (Harris, M. (1999) *Curr. Biol.* 9:R449-R461). A recently  
20 identified human protein, Nef-associated factor 1 (Naf1), a protein with four extended coiled-coil domains, has been found to associate with Nef. Overexpression of Naf1 increased cell surface expression of CD4, an effect which could be suppressed by Nef (Pukushii, M. et al. (1999) *FEBS Lett.* 442:83-88).

The discovery of new secretion and trafficking molecules and the polynucleotides encoding  
25 them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of vesicle trafficking, transport, neurological, autoimmune/inflammatory, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of secretion and trafficking molecules.

#### 30 SUMMARY OF THE INVENTION

The invention features purified polypeptides, secretion and trafficking molecules, referred to collectively as "SAT" and individually as "SAT-1," "SAT-2," "SAT-3," "SAT-4," "SAT-5," "SAT-6," "SAT-7," "SAT-8," and "SAT-9." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected

WO 02/02610

PCT/US01/20704

from the group consisting of SEQ ID NO:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-9.

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:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-9. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:10-18.

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:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. 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:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with

WO 02/02610

PCT/US01/20704

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:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

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:10-18, 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:10-18, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, 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:10-18, 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:10-18, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90%

WO 02/02610

PCT/US01/20704

identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, 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)

5 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:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional SAT, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. 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 SAT, comprising administering to a patient in need of such treatment the composition.

30 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:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid

WO 02/02610

PCT/US01/20704

sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. 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 SAT, comprising administering to a patient in need of such treatment the composition.

10 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:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

20 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:1-9, 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:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. 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.

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

WO 02/02610

PCT/US01/20704

selected from the group consisting of SEQ ID NO:10-18, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, 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:10-18, 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:10-18, 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:10-18, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

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

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

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.

WO 02/02610

PCT/US01/20704

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.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

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

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

#### 10 DESCRIPTION OF THE INVENTION

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

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

20 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in 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

30 "SAT" refers to the amino acid sequences of substantially purified SAT 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 SAT. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other

WO 02/02610

PCT/US01/20704

compound or composition which modulates the activity of SAT either by directly interacting with SAT or by acting on components of the biological pathway in which SAT participates.

An "allelic variant" is an alternative form of the gene encoding SAT. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or 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 SAT include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as SAT or a polypeptide with at least one functional characteristic of SAT. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding SAT, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding SAT. 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 SAT. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of SAT 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

WO 02/02610

PCT/US01/20704

of SAT. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of SAT either by directly interacting with SAT or by acting on components of the biological pathway in which SAT participates.

5 The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind SAT 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,  
10 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  
15 immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense"  
20 (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 SAT, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid

WO 02/02610

PCT/US01/20704

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 SAT or fragments of SAT 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

WO 02/02610

PCT/US01/20704

Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

---

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

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

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

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

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

          A "fragment" is a unique portion of SAT or the polynucleotide encoding SAT which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up  
25           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, 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  
30           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 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

35           A fragment of SEQ ID NO:10-18 comprises a region of unique polynucleotide sequence that

WO 02/02610

PCT/US01/20704

specifically identifies SEQ ID NO:10-18, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:10-18 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:10-18 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:10-18 and the region of SEQ ID NO:10-18 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:1-9 is encoded by a fragment of SEQ ID NO:10-18. A fragment of SEQ ID NO:1-9 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-9. For example, a fragment of SEQ ID NO:1-9 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-9. The precise length of a fragment of SEQ ID NO:1-9 and the region of SEQ ID NO:1-9 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

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

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

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

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

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment

WO 02/02610

PCT/US01/20704

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*

15 *Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

20 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.

25 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.

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

WO 02/02610

PCT/US01/20704

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

WO 02/02610

PCT/US01/20704

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

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

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

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

WO 02/02610

PCT/US01/20704

to which cells or their nucleic acids have been fixed).

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

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

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of SAT 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 SAT which is useful in any of the antibody production methods disclosed herein or known in the art.

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

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

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

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

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

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

"Post-translational modification" of an SAT may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the

WO 02/02610

PCT/US01/20704

art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of SAT.

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

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers  
15 may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

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

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

WO 02/02610

PCT/US01/20704

avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

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

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

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

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

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

WO 02/02610

PCT/US01/20704

instead of deoxyribose.

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

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

WO 02/02610

PCT/US01/20704

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

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may 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

WO 02/02610

PCT/US01/20704

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.

#### THE INVENTION

5 The invention is based on the discovery of new human secretion and trafficking molecules (SAT), the polynucleotides encoding SAT, and the use of these compositions for the diagnosis, treatment, or prevention of vesicle trafficking, transport, neurological, autoimmune/inflammatory, and cell proliferative disorders.

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

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

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and  
25 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI).  
30 Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are secretion and trafficking molecules. For

WO 02/02610

PCT/US01/20704

example, SEQ ID NO:2 is 93% identical to mitsugutin29 (GenBank ID g3077703), a synaptophysin family member, as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $2.9e-136$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a synaptophysin/synaptoporin 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 PROFILESCAN analyses, and BLAST comparisons to protein signature sequences in the DOMO and PRODOM databases provide further corroborative evidence that SEQ ID NO:2 is a synaptophysin family member. SEQ ID NO:3 is 72% identical to rat apical endosomal glycoprotein (GenBank ID g777776) with a BLAST probability score of 0.0. Data from BLAST analyses against the PRODOM database provide further corroborative evidence that SEQ ID NO:3 is an apical endosomal glycoprotein. SEQ ID NO:8 is 95% identical to *Rattus norvegicus* synaptotagmin III (GenBank ID g484296) with a BLAST probability score of 0.0. SEQ ID NO:8 also contains a C2 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, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:8 is a C2 domain-containing protein, most likely a member of the synaptotagmin family. SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, and SEQ ID NO:9 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-9 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO.) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:10-18 or that distinguish between SEQ ID NO:10-18 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

WO 02/02610

PCT/US01/20704

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 1438701F1 is the identification number of an Incyte cDNA sequence, and PANCNOT02 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70767606V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5810426) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (i.e., those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 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 identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX 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<sub>1,2,3,4</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL\_XXXXXX\_gAAAAA\_gBBBBB\_1\_N is the identification number of a "stretched" sequence, with XXXXXX 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
--------	--

WO 02/02610

PCT/US01/20704

GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).

5

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

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

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

20 The invention also encompasses polynucleotides which encode SAT. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:10-18, which encodes SAT. The polynucleotide sequences of SEQ ID NO:10-18, 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.

25 The invention also encompasses a variant of a polynucleotide sequence encoding SAT. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding SAT. 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:10-18 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:10-18. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of SAT.

30 It will be appreciated by those skilled in the art that as a result of the degeneracy of the

WO 02/02610

PCT/US01/20704

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

Although nucleotide sequences which encode SAT and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring SAT under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding SAT or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding SAT 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 SAT and SAT derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding SAT or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:10-18 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE 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

WO 02/02610

PCT/US01/20704

(Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

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

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

WO 02/02610

PCT/US01/20704

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  
5 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 SAT may be cloned in recombinant DNA molecules that direct expression of SAT, or  
10 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 SAT.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter SAT-encoding sequences for a variety of purposes including, but not  
15 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 Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve  
20 the biological properties of SAT, 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  
25 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.

WO 02/02610

PCT/US01/20704

In another embodiment, sequences encoding SAT 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, SAT 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 SAT, 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 SAT, the nucleotide sequences encoding SAT 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 SAT. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding SAT. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding SAT 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 SAT and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in*

WO 02/02610

PCT/US01/20704

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 SAT. 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 SAT. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding SAT can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding SAT 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 SAT are needed, e.g. for the production of antibodies, vectors which direct high level expression of SAT 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 SAT. A number of vectors

WO 02/02610

PCT/US01/20704

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;

5 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 SAT. Transcription of sequences encoding SAT 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.*

10 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,

15 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 SAT 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

20 infective virus which expresses SAT 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

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

For long term production of recombinant proteins in mammalian systems, stable expression of

30 SAT in cell lines is preferred. For example, sequences encoding SAT can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a

WO 02/02610

PCT/US01/20704

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),  $\beta$  glucuronidase and its substrate  $\beta$ -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 SAT is inserted within a marker gene sequence, transformed cells containing sequences encoding SAT can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding SAT 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 SAT and that express SAT may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of SAT 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

WO 02/02610

PCT/US01/20704

monoclonal antibodies reactive to two non-interfering epitopes on SAT is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

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

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

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid

WO 02/02610

PCT/US01/20704

sequences encoding SAT 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 SAT protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of SAT 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 SAT encoding sequence and the heterologous protein sequence, so that SAT 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 SAT 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, <sup>35</sup>S-methionine.

SAT of the present invention or fragments thereof may be used to screen for compounds that specifically bind to SAT. At least one and up to a plurality of test compounds may be screened for specific binding to SAT. 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 SAT, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which SAT 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 SAT, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing SAT or cell membrane fractions which contain SAT are then contacted with a test

WO 02/02610

PCT/US01/20704

compound and binding, stimulation, or inhibition of activity of either SAT 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 SAT, either in solution or affixed to a solid support, and detecting the binding of SAT 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.

SAT of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of SAT. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for SAT activity, wherein SAT is combined with at least one test compound, and the activity of SAT in the presence of a test compound is compared with the activity of SAT in the absence of the test compound. A change in the activity of SAT in the presence of the test compound is indicative of a compound that modulates the activity of SAT. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising SAT under conditions suitable for SAT activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of SAT 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 SAT or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential

WO 02/02610

PCT/US01/20704

therapeutic or toxic agents.

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

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

#### 15 THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of SAT and secretion and trafficking molecules. In addition, the expression of SAT is closely associated with brain, spinal cord, lymphatic, and reproductive tissues. Therefore, SAT appears to play a role in vesicle trafficking, transport, neurological, autoimmune/inflammatory, and cell  
20 proliferative disorders. In the treatment of disorders associated with increased SAT expression or activity, it is desirable to decrease the expression or activity of SAT. In the treatment of disorders associated with decreased SAT expression or activity, it is desirable to increase the expression or activity of SAT.

Therefore, in one embodiment, SAT or a fragment or derivative thereof may be administered  
25 to a subject to treat or prevent a disorder associated with decreased expression or activity of SAT. Examples of such disorders include, but are not limited to, a vesicle trafficking disorder such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions  
30 associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS); allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminthic, and protozoal

WO 02/02610

PCT/US01/20704

infections; a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer; cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis; neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia; and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; 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, 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

WO 02/02610

PCT/US01/20704

frontotemporal dementia; an autoimmune/inflammatory disorder such as 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, 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 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; 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 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, and uterus.

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

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

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

In a further embodiment, an antagonist of SAT may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SAT. Examples of such disorders include, but are not limited to, those vesicle trafficking, transport, neurological, autoimmune/inflammatory, and cell proliferative disorders described above. In one aspect, an antibody

WO 02/02610

PCT/US01/20704

which specifically binds SAT 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 SAT.

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

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

An antagonist of SAT may be produced using methods which are generally known in the art. In particular, purified SAT may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind SAT. Antibodies to SAT 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 SAT 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 SAT 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 SAT 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 SAT may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited

WO 02/02610

PCT/US01/20704

to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

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

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

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

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

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for SAT. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of SAT-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$

WO 02/02610

PCT/US01/20704

determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple SAT epitopes, represents the average affinity, or avidity, of the antibodies for SAT. The  $K_a$ , determined for a preparation of monoclonal antibodies, which are monospecific for a particular SAT 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 SAT-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 SAT, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

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

In another embodiment of the invention, the polynucleotides encoding SAT, 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 SAT. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding SAT. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

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

WO 02/02610

PCT/US01/20704

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 SAT may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in SAT expression or regulation causes disease, the expression of SAT 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 SAT are treated by constructing mammalian expression vectors encoding SAT and introducing these vectors by mechanical means into SAT-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 SAT include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). SAT may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus

WO 02/02610

PCT/US01/20704

(RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the 5 ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding SAT from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID 10 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these 15 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to SAT expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding SAT under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive 20 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for 25 receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses 30 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<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene 1 therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998)

WO 02/02610

PCT/US01/20704

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

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

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding SAT 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

WO 02/02610

PCT/US01/20704

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 SAT into the alphavirus genome in place of the capsid-coding region results in the production of a large number of SAT-coding RNAs and the synthesis of high levels of SAT 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 SAT into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.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 SAT.

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

WO 02/02610

PCT/US01/20704

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.

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

10 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, 15 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.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding SAT. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not 20 limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased SAT 25 expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding SAT may be therapeutically useful, and in the treatment of disorders associated with decreased SAT expression or activity, a compound which specifically promotes expression of the polynucleotide encoding SAT may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in 30 altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a

WO 02/02610

PCT/US01/20704

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

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

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

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient.

Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of SAT, antibodies to SAT, and mimetics, agonists, antagonists, or inhibitors of SAT.

The compositions utilized in this invention may be administered by any number of routes

WO 02/02610

PCT/US01/20704

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.

5 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  
10 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.

15 Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising SAT or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, SAT 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  
20 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

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

A therapeutically effective dose refers to that amount of active ingredient, for example SAT or fragments thereof, antibodies of SAT, and agonists, antagonists or inhibitors of SAT, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by  
30 standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD<sub>50</sub>/ED<sub>50</sub> ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are

WO 02/02610

PCT/US01/20704

used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

5 The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response  
10 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  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

15 Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

#### DIAGNOSTICS

In another embodiment, antibodies which specifically bind SAT may be used for the diagnosis  
20 of disorders characterized by expression of SAT, or in assays to monitor patients being treated with SAT or agonists, antagonists, or inhibitors of SAT. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for SAT include methods which utilize the antibody and a label to detect SAT 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  
25 covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

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

WO 02/02610

PCT/US01/20704

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

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding SAT or closely related molecules may be used to identify nucleic acid sequences which encode SAT. 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 SAT, 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 SAT 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:10-18 or from genomic sequences including promoters, enhancers, and introns of the SAT gene.

Means for producing specific hybridization probes for DNAs encoding SAT include the cloning of polynucleotide sequences encoding SAT or SAT 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 <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding SAT may be used for the diagnosis of disorders associated with expression of SAT. Examples of such disorders include, but are not limited to, a vesicle trafficking disorder such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS); allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage;

WO 02/02610

PCT/US01/20704

and viral, bacterial, fungal, helminthic, and protozoal infections; a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis,

5 Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer; cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy,

10 dermatomyositis, inclusion body myositis, infectious myositis, polymyositis; neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia; and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal

15 deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease,

20 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

25 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

30 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

WO 02/02610

PCT/US01/20704

supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as 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 lymphocytotoxicity, 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 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; 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, and uterus. The polynucleotide sequences encoding SAT 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 SAT expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding SAT may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding SAT 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 SAT 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

WO 02/02610

PCT/US01/20704

the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of SAT, 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 SAT, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

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

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

Additional diagnostic uses for oligonucleotides designed from the sequences encoding SAT 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 SAT, or a fragment of a polynucleotide complementary to the polynucleotide encoding SAT, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding SAT 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 SAT are used to amplify DNA using the

WO 02/02610

PCT/US01/20704

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 fSCCP, the oligonucleotide primers are 5 fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual 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 10 errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of SAT include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from 15 standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplax, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

20 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene 25 function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and 30 display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, SAT, fragments of SAT, or antibodies specific for SAT 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.

WO 02/02610

PCT/US01/20704

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

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

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

WO 02/02610

PCT/US01/20704

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

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global  
10 pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is  
15 achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot  
20 is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass  
25 spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for SAT to quantify the  
30 levels of SAT 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

WO 02/02610

PCT/US01/20704

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 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 SAT 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

WO 02/02610

PCT/US01/20704

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

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding SAT 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, SAT, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between SAT 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

WO 02/02610

PCT/US01/20704

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

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

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

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/215,465, U.S. Ser. No. 60/239,384, and U.S. Ser. No. 60/253,639, are hereby expressly incorporated by reference.

#### EXAMPLES

##### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively,

WO 02/02610

PCT/US01/20704

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<sup>T</sup> 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., PBLUESCRIP<sup>T</sup> plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

### II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such

WO 02/02610

PCT/US01/20704

as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

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

WO 02/02610

PCT/US01/20704

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:10-18. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative secretion and trafficking molecules 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 secretion and trafficking molecules, the encoded polypeptides were analyzed by querying against PFAM models for secretion and trafficking molecules. Potential secretion and trafficking molecules were also identified by homology to Incyte cDNA sequences that had been annotated as secretion and trafficking molecules. These selected Genscan-predicted sequences were then compared by 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

WO 02/02610

PCT/US01/20704

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.

5 **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  
10 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  
15 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.  
20 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  
25 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  
30 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

WO 02/02610

PCT/US01/20704

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.

5 VI. Chromosomal Mapping of SAT Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:10-18 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:10-18 were assembled into clusters of contiguous and overlapping sequences using  
10 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.

15 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  
20 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.

25 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.)

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

WO 02/02610

PCT/US01/20704

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

5 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  
 10 (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  
 15 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 SAT are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at  
 20 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;  
 25 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  
 30 libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding SAT. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of SAT Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate

WO 02/02610

PCT/US01/20704

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 Mg<sup>2+</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 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 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

WO 02/02610

PCT/US01/20704

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/Zx 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. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:10-18 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 µCi of [<sup>32</sup>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<sup>7</sup> 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 & Schnell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### X. Microarrays

WO 02/02610

PCT/US01/20704

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).

5 Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schna, M. et al. (1995) Science  
10 270:467-470; 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  
15 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  
20 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  
25 poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ 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)<sup>+</sup> RNA with  
30 GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> 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.

WO 02/02610

PCT/US01/20704

(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  $\mu$ l 5X SSC/0.2% SDS.

#### 5 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  $\mu$ g.

10 Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and  
15 coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic  
20 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%  
25 SDS and distilled water as before.

#### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ 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  
30 an 1.8 cm<sup>2</sup> 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  $\mu$ 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.

WO 02/02610

PCT/US01/20704

**Detection**

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

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

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).

WO 02/02610

PCT/US01/20704

**XI. Complementary Polynucleotides**

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

**XII. Expression of SAT**

Expression and purification of SAT is achieved using bacterial or virus-based expression systems. For expression of SAT 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 SAT upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of SAT 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 SAT 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, SAT 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 SAT at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-

WO 02/02610

PCT/US01/20704

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 SAT obtained by these methods can be used directly in the assays shown in Examples XVI and XVII, where applicable.

### 5 XIII. Functional Assays

SAT function is assessed by expressing the sequences encoding SAT 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  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of SAT on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding SAT 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 SAT and other genes of interest can be analyzed by northern analysis or microarray techniques.

WO 02/02610

PCT/US01/20704

**XIV. Production of SAT Specific Antibodies**

SAT 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.

5 Alternatively, the SAT 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.)

10 Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for  
15 antipeptide and anti-SAT activity by, for example, binding the peptide or SAT to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

**XV. Purification of Naturally Occurring SAT Using Specific Antibodies**

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

**XVI. Identification of Molecules Which Interact with SAT**

30 SAT, or biologically active fragments thereof, are labeled with <sup>125</sup>I 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 SAT, washed, and any wells with labeled SAT complex are assayed. Data obtained using different concentrations of SAT are used to calculate values for the number, affinity, and association of SAT with the candidate

WO 02/02610

PCT/US01/20704

molecules.

Alternatively, molecules interacting with SAT 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).

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

#### XVII. Demonstration of SAT Activity

- 10 SAT activity is measured by its inclusion in coated vesicles. SAT can be expressed by transforming a mammalian cell line such as COS7, HeLa, or CHO with an eukaryotic expression vector encoding SAT. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A small amount of a second plasmid, which expresses any one of a number of marker genes, such as  $\beta$ -galactosidase, is co-
- 15 transformed into the cells in order to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of SAT and  $\beta$ -galactosidase.

- Transformed cells are collected and cell lysates are assayed for vesicle formation. A non-
- 20 hydrolyzable form of GTP, GTP $\gamma$ S, and an ATP regenerating system are added to the lysate and the mixture is incubated at 37 °C for 10 minutes. Under these conditions, over 90% of the vesicles remain coated (Orci, L. et al. (1989) *Cell* 56:357-368). Transport vesicles are salt-released from the Golgi membranes, loaded under a sucrose gradient, centrifuged, and fractions are collected and analyzed by SDS-PAGE. Co-localization of SAT with clathrin or COP coatamer is indicative of SAT activity in
- 25 vesicle formation. The contribution of SAT in vesicle formation can be confirmed by incubating lysates with antibodies specific for SAT prior to GTP $\gamma$ S addition. The antibody will bind to SAT and interfere with its activity, thus preventing vesicle formation.

In the alternative, SAT activity is measured by its ability to alter vesicle trafficking pathways. Vesicle trafficking in cells transformed with SAT is examined using fluorescence microscopy.

- 30 Antibodies specific for vesicle coat proteins or typical vesicle trafficking substrates such as transferrin or the mannose-6-phosphate receptor are commercially available. Various cellular components such as ER, Golgi bodies, peroxisomes, endosomes, lysosomes, and the plasmalemma are examined. Alterations in the numbers and locations of vesicles in cells transformed with SAT as compared to control cells are characteristic of SAT activity.

WO 02/02610

PCT/US01/20704

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.

- 5 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.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1577952	1	1577952CD1	10	1577952CBI
4983705	2	4983705CD1	11	4983705CBI
1310465	3	1310465CD1	12	1310465CBI
4291779	4	4291779CD1	13	4291779CBI
4728247	5	4728247CD1	14	4728247CBI
7472259	6	7472259CD1	15	7472259CBI
7476740	7	7476740CD1	16	7476740CBI
7946329	8	7946329CD1	17	7946329CBI
7946329	9	7946329CD1	18	7946329CBI

WO 02/02610

PCT/US01/20704

Table 2

Polypeptide SEQ ID NO:	Inverte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
2	4983705CDL	g5077703	2.90E-136	[Oryctolagus cuniculus] mitsurumin29 Takeshima, H. et al. (1998) Biochem. J. 331:317-322.
3	1310465CDL	g777776	0	[Rattus norvegicus] apical endosomal glycoprotein Speelman, B.A. et al. (1995) J. Biol. Chem. 270:1583-1588
4	4291779CDL	g3378150	2.30E-45	[Trypanosoma brucei rhodesiense] Kemp, D. et al. (1999) J. Biol. Chem. 274:28542-28548
5	4728247CDL	g5926736	0	[Rattus norvegicus] protein p67 Kelley, R.J. et al. (1999) Mol. Biochem. Parasitol. 98:17-28
6	7472259CDL	g6925888	3.70E-101	[Mus musculus] gramphilin-a Wang, J. et al. (1999) J. Biol. Chem. 274:28542-28548
7	7476740CDL	g6977950	5.60E-43	[Rattus norvegicus] REM binding protein 1A Wang, Y. et al. (2000) J. Biol. Chem. 275:20093-20098
8	7473774CDL	g484296	0	[Rattus norvegicus] putative v- src binding protein, v-src domain Mizuta, M. et al. (1994) J. Biol. Chem. 269:11675-11678
9	7946329CDL	g6136794	1.30E-221	[Mus musculus] synaptotagmin XI Fukuda, M. et al. (1999) J. Biol. Chem. 274:31427-31437

WO 02/02610

PCT/US01/20704

Table 3

SEQ Inocyte ID Polypeptide No. ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1 1577952CD1	315			signal peptide:K1-A33 Transmembrane domain:S101-E121 Transmembrane domain:K214-E234 Transmembrane domain:K214-E234 SYNAPTOPHYSIN / SYNAPTOPORIN: DMG2395 E20488 21-148:638-V166 SYNAPTOPHYSIN: PD05837:R27-A268 SYNAPTOPHYSIN / synaptophysin: BL00604A:L41-L95 BL00604B:T104-E133 BL00604C:K214-E162 BL00604D:C203-E162 SYNAPTOPHYSIN/SYNAPTOPORIN: PR00220A:I38-T60 PR00220B:A62-H87 PR00220C:F117-N141 PR00220D:F149-G172 PR00220E:V216-E234 Synaptophysin / synaptoporin signature motif:Q42-I89 SYNAPTOPHYSIN / SYNAPTOPORIN Transmembrane domains: P114-V136, I214-E234 Synaptophysin / synaptoporin: SYNAPTOPHYSIN:R27-Q272	HMMER HMMER HMMER_Pfam BLAST_DOMO BLAST_PRODOM BLIMPS_BLOCKS BLIMPS_PRINTS PROFILESSCAN HMMER HMMER_Pfam
2 4983705CD1	272	S101 S14 S2 S262	N213		



WO 02/02610

PCT/US01/20704

Table 3 (cont.)

SFO ID	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	4231773CD1	589	S157 S172 S308 S456 S506 T208 T3161 T3299 Y196	N116 N231 N416 N465 N515 N88	Sigmal Sequences: MI-A41 F09812 F09813 PD145700-C0403-D589 F09812.3 protein, IAMA lysosomal/endosomal p67 PD043621-C244-D532 Protein splicing site: LZ58-T263 C2 domains: L373-E462, L52E-R617 C2 domain signatures/profiles: F408-H574, G466-L636 Synaptosomal signature: PR00399: I360-V375, P413-H448 C2 domain signatures PR00360: A543-L555, K572-N585 C7-domain DM00150   P40748   294-423: G358-E460, E527-I618 C2-domain DM00150   JC2473   249-373: G358-E459, L528-D634 C2-domain DM00150   P46097   140-266: G358-E460, E527-I618 C2-domain DM00150   P24507   233-362: G358-E460, E527-I618	SPSCAN BLAST-PRODOM BLAST-PRODOM MOTIFS HMMER-PFAM PROFILESKAN BLIMPS-PRINFS BLIMPS-PRINFS BLAST-DOMO BLAST-DOMO BLAST-DOMO
5	4728247CD1	671	S11 S19 S147 S210 S231 S232 S274 S289 S363 S393 S414 S467 S494 S502 S520 S55 S563 S59 S625 S635 S652 S74 T115 T131 T137 T142 T143 T144 T427 T567 T600 T636 T79 T99 Y120	N537 N579 N97		



WO 02/02610

PCT/US01/20704

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7473774CD1	590	S160 S166 S240 S242 S243 S427 S488 S584 S75 T184 T223 T487 T493 T497	R365 N436	PROTEIN KINASE C 2 REGION BL03932 F40748 1-292:MI-G290 SYNAPSE III SYNTIII C P40748 292-397-G496, K430-L454 P40748 425-552-S427-S585, K398-V423 P24507 364-491-S427-N556, A398-V423 SYNAPTOTAGMIN TRANSMEMBRANE REPEAT SYNAPSE III SYNTIII VI C SYNAPTIC VESICLE PD022173:F142-Q315 SYNAPTOTAGMIN TRANSMEMBRANE REPEAT SYNAPSE III SYNTIII C SYNAPTIC VESICLE PROTEIN P102 PD012608:MI-P102 PD007398:RS36-S580 PROTEIN C REPEAT SYNAPTOTAGMIN PHOSPHOLIPASE TRANSMEMBRANE SYNAPSE BINDING PHOSPHOLIPASE KINASE PD00136:L316-Q400, L448-CE35 C2 domain proteins L475-E500 SYNAPTOTAGMIN REPEAT PRO0399:P373-D388, S393-L403, L303-V318, V318-S331 C2 domain signature PRO0360:S331-L343, K358-S371, L382-D390 Transmembrane Domain:F51-V74 C2 domain C2:L316-V402, L448-R536 C2 domain signature gcs_motif: L432-439 C2 domain signature gcs_motif: c2.domain.prf:L435-K490, E303-K358	BLAST_DOMO BLAST_DOMO BLAST_PRODUM BLIMPS_PPFAM BLIMPS_PRINTMS BLIMPS_PRINTMS PFAM PFAM MOTIFS PROFILES PROFILES



Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
10	1577952CB1	3424	1-39 2095-2092 3272-3424	70687032V1	2938	3424
				7438458HL (ADREUR02)	2439	3019
				1441635V1	574	1240
				7041635V1	266	2584
				8009276HL (NOSEBIC02)	662	2189
				21112240V1	180	180
				71114548V1	1904	189
				79931001HL (UTRSDIC01)	1906	2435
				70767606V1	1	627
				1438701F1 (PANCHOT08)	1361	1899
11	4983705CB1	1033	732-1033	GBI:96524214_000026_raw.1	674	1033
				45810426	1	445
				4983705HL (HELA/MT05)	266	535
				6901767R8 (MUS/ADRO2)	291	867
				7610385HL (KIDC/MRE01)	39	305
12	1310465CB1	3902	1-1452 1764-2846	7164475F8 (FLAC/NOR01)	416	1016
				71991166V1	2676	3281
				7144340HL (SINTFEB02)	136	1732
				714434001 (SINTFEB02)	1433	2056
				8144340HL (FONSDIC01)	2011	2714
				8542248HL (LOND/NOR02)	386	3902
				71987767V1	177	2423
				71991166V1	2136	3423
				7080102P6 (STOMME02)	466	4723
				7693543H2 (LOND/TUR01)	1	439

WO 02/02610

PCT/US01/20704

Table 4 (cont.)

Polynucleotide SEQ ID NO.	Inocyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
13	4291779CB1	2574	979-1592	70735773V1	1460	2082
				252443766 (BRATYU021)	1460	2082
				6154207H1 (BRMDM004)	1470	2128
				6996766H1 (BRAXYDR17)	635	1508
				7061302H1 (PENYMM002)	846	1663
				1869872F6 (SKINBI101)	2185	2574
				70738661V1	1973	2557
				7278956H1 (BMAPXXE01)	1	565
				4822574F9 (PROSTUT17)	479	937
				2079722F6 (UTKSN008)	1964	2462
14	4728247CB1	2878	539-2368	70867423V1	2397	2878
				71171848V1	1383	1966
				71399565V1	641	1331
				2246668E (PANCNOT01)	369	868
				3394730F6 (PROSN028)	1	579
				4631668H1 (PANCNOT01)	1831	2455
				4631668H1 (PANCNOT01)	1309	1902
				73369631H1 (KUPEN003)	1374	2096
				7761273H1 (HYPMNO03)	4134	479
				6842194H1 (BRSTWNO02)	5028	5628
15	7472359CB1	5628	1-2080 5302-5628 2575-4411	2394503F6 (THELIZ001)	4411	4860
				619633660	4767	5233
				62063588	1	4922
				GNN_46006353_010	1	4922

WO 02/02610

PCT/US01/20704

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Intyre Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
16	7476740CB1	1482	1-566 750-831	GNN_07547222_000015_002 4206417 6247775F8 (TESTM0717)	175 946 1	1355 1482 755
17	7473774CB1	2511	1-113 639-1330 2471-2511	70555497V1 72049138V1 7585860H2 (BRAIFEC01) 71045227V1 72050239V1 71539112V1 7228832H1 (BRAXTDR15) 71801792V1 71801805V1 6996340H1 (BRAXTDR17)	1470 716 1 1640 587 2184 1 460 607 1091	2035 1538 630 2282 1475 2511 480 1167 1231 1680
18	7946329CB1	1680				

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
10	1577952CBI	LIVRUT13
11	4983705CBI	BRANOT01
12	1310465CBI	SINTFER02
13	4291779CBI	BRAITUT21
14	4728247CBI	PANCR0T01
15	7472359CBI	THEP1A2T01
16	4877464CBI	TESTROT17
17	7437747CBI	BRALFEC01
18	7946329CBI	SCOMPIC01

WO 02/02610

PCT/US01/20704

Table 6

Library	Vector	Library Description
BRA1F001	P1NCY	This large size-fractionated library was constructed using RNA isolated from the placenta removed from a Caucasian male fetus who was stillborn with a hyoplastic left ventricle.
BRA1W021	P1NCY	This library was constructed using RNA isolated from a brain tumor removed from the midline frontal lobe of a 61-year-old Caucasian female during resection of a cerebral meningial lesion. Pathology indicated subfrontal meningothelial meningioma with no atypia. One atypical and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
BRA2M001	P1NCY	This library was constructed using RNA isolated from striatum, globus pallidus and cerebral putamen of a 70-year-old Caucasian female who died from a dissecting aortic aneurysm and ischemic bowel disease. Pathology indicated mild arteriosclerosis involving the cerebral cortical white matter and basal ganglia. Grossly, there was mild meningeal fibrosis and mild focal atherosclerotic plaque in the middle cerebral artery, as well as vertebral arteries bilaterally. Microscopically, the cerebral hemispheres, brain stem and cerebellum revealed focal areas in the white matter that contained blood vessels with cholesterol-shaped plaques. The cerebellum contained scattered neurofibrillary tangles within the basolateral nuclei of the amygdala. The brain stem had mild atheromatosis of aorta and coronary arteries, bowel and liver infarct due to aneurysm, physiologic fatty liver associated with obesity, mild diffuse emphysema, thrombosis of mesenteric and portal veins, cardiomegaly due to hypertrophy of left ventricle, arterial hypertension, acute pulmonary edema, splenomegaly, obesity (300 lb.), leiomyoma of uterus, sleep apnea, and iron deficiency anemia.

WO 02/02610

PCT/US01/20704

Table 6 (cont.)

Library	Vector	Library Description
LIVERWU13	P1NCY	Library was constructed using RNA isolated from liver tumor tissue removed from a 62-year-old Caucasian female who underwent a laparotomy and exploratory laparotomy. Pathology indicated metastatic intermediate grade carcinoma, consistent with islet cell tumor, forming nodules ranging in size, in the lateral and medial left liver lobe. The pancreas showed fibrosis, chronic inflammation and fat necrosis consistent with pseudocyst. The gall bladder showed mild chronic cholecystitis. Patient history included malignant neoplasm of the pancreas tail, pulmonary embolism, hyperlipidemia, hypertension, Type II diabetes, benign prostatic hyperplasia, hypercholesterolemia, hypertriglyceridemia, hypertension, and cerebrovascular disease. Patient history also includes cancer, secondary liver cancer, benign hypertension, and hyperlipidemia.
PANCR0701	PBLUESCRIPT	Library was constructed using RNA isolated from the pancreatic tissue of a 29-year-old Caucasian male who died from head trauma.
SCOMD101	PSPORT1	This large size-fractionated library was constructed using RNA isolated from diseased spinal cord tissue removed from the base of the medulla of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day).
SMITFE02	PCDNA2.1	This 5' biased random library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male who died from a small intestine tumor.
TESTNOT17	P1NCY	Patau's syndrome (trisomy 13) at 20-weeks' gestation. Serology was negative. Library was constructed from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor).
TRPLA2T01	P1NCY	Library was constructed using RNA isolated from WHP-1 promonocyte cells treated for 48 hours with interferon- $\gamma$ (1000 IU/ml), ara-2'-deoxycytidine (100 ng/ml), and phytohemagglutinin (PHA) (100 ng/ml). Patient history included a heart attack and is a human promonocyte leukemia (ATCC TIB 202).
		Library was constructed using RNA isolated from peripheral blood of a 62-year-old Caucasian male with acute monocytic leukemia (Am. J. Cancer (1989), 20:111).

WO 02/02610

PCT/US01/20704

Table 7

Program	Description	Reference	Parameter/Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI PARACEL.PDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ZSYs: 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, fasta, fastx, frstax, and search.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ZSYs: fasta E value=1.06E-6 Assembled ZSYs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value= 1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLOMPS	A BLOKS IMProved Searcher that matches a sequence against those in BLOKS, 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. 19:656S-657Z; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Altwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1998) Nucleic Acids Res. 26:320-322; 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 hit: Score= 0 or greater

WO 02/02610

PCT/US01/20704

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Baroch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores, GCC-specified "HIGH" value for that particular Prosite motif. Generally, score=14-2.1.
Pired	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-182; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Pirap	A Phils 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.	Scores=120 or greater; Match length>= 56 or greater
Consed	A graphical tool for viewing and editing Pileup assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Scores=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.	Sommadossi, 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.	Baroch, 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 02/02610

PCT/US01/20704

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of  
5 SEQ ID NO:1-9,
  - 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:1-9,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected  
from the group consisting of SEQ ID NO:1-9, and
  - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from  
the group consisting of SEQ ID NO:1-9.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-9.
- 15 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 selected from the group consisting of SEQ ID  
20 NO:10-18.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a  
polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method for producing a polypeptide of claim 1, the method comprising:
  - 30 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
  - b) recovering the polypeptide so expressed.

WO 02/02610

PCT/US01/20704

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting
  - 5 of SEQ ID NO:10-18,
  - 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:10-18,
  - c) a polynucleotide complementary to a polynucleotide of a),
  - d) a polynucleotide complementary to a polynucleotide of b), and
  - 10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex
  - 20 is formed between said probe and said target polynucleotide or fragments thereof, and
  - b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
  - 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

WO 02/02610

PCT/US01/20704

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

18. A method for treating a disease or condition associated with decreased expression of functional SAT, comprising administering to a patient in need of such treatment the composition of claim 16.

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

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

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional SAT, comprising administering to a patient in need of such treatment a composition of claim 20.

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

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

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional SAT, comprising administering to a patient in need of such treatment a composition of claim 23.

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

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

WO 02/02610

PCT/US01/20704

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

WO 02/02610

PCT/US01/20704

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

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

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

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

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

32. A method of diagnosing a condition or disease associated with the expression of SAT in a subject, comprising administering to said subject an effective amount of the composition of claim 31.  
20

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

34. A method of diagnosing a condition or disease associated with the expression of SAT in a subject, comprising administering to said subject an effective amount of the composition of claim 33.  
25

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

- 30 a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, 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 having an amino acid sequence selected from the

WO 02/02610

PCT/US01/20704

group consisting of SEQ ID NO:1-9.

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

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

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

10 a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, 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;

15 d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

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

20

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

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

25

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

43. A method for detecting a polypeptide having an amino acid sequence selected from the 30 group consisting of SEQ ID NO:1-9 in a sample, comprising the steps of:

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

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide

WO 02/02610

PCT/US01/20704

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9 from a sample, the method comprising:

- 5 a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

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

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

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

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

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

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

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

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

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

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

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

56. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:12.

WO 02/02610

PCT/US01/20704

- 57. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:13.
- 58. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:14.
- 5 59. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:15.
- 60. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:16.
- 61. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:17.
- 10 62. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:18.

WO 02/02610

PCT/US01/20704

&lt;110&gt; INCYTE GENOMICS, INC.

LAL, Preeti  
 TANG, Y. Tom  
 YUE, Henry  
 WALIA, Narinder K.  
 BAUGHN, Mariah R.  
 DAS Debopriya  
 RAMKUMAR Jayalaxmi  
 TRIBOULEY, Catherine M.  
 LU, Dyung Aina M.  
 HAPALIA, April  
 GANDHI, Ameena R.  
 LEE, Ernestine A.  
 XU, Yuming  
 BANDMAN, Olga  
 ELLIOT, Vicki S.  
 NGUYEN, Damiel B.  
 BURRILL John D.  
 MARCUS, Gregory A.  
 ZINGLER, Kurt A.  
 LU, Van  
 YAO Monique G.  
 GURURAJAN, Rajagopal  
 DING, Li  
 WARREN, Bridget A.  
 THANGAVELU, Kavitha  
 LEE, sally

&lt;120&gt; SECRETION AND TRAFFICKING MOLECULES

&lt;130&gt; PF-0801 PCT

&lt;140&gt; To Be Assigned

&lt;141&gt; Herewith

&lt;150&gt; 60/215,465; 60/239,384; 60/253,639

&lt;151&gt; 2000-06-29; 2000-10-10; 2000-11-28

&lt;160&gt; 18

&lt;170&gt; PERL Program

&lt;210&gt; 1

&lt;211&gt; 315

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1577952CD1

&lt;400&gt; 1

Met	Gln	Arg	Arg	Ser	Arg	Gly	Ile	Asn	Thr	Gly	Leu	Ile	Leu	Leu
1				5					10					15
Leu	Ser	Gln	Ile	Phe	His	Val	Gly	Ile	Asn	Asn	Ile	Pro	Pro	Val
				20					25					30
Thr	Leu	Ala	Thr	Leu	Ala	Leu	Asn	Ile	Trp	Phe	Phe	Leu	Asn	Pro

WO 02/02610

PCT/US01/20704

```

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

```

&lt;210&gt; 2

&lt;211&gt; 272

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4983705CD1

&lt;400&gt; 2

```

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

```



WO 02/02610

PCT/US01/20704

Leu Gly Thr Asp Leu Gly Trp Tyr Met Ala Val Gly Thr His Arg  
 110 115 120  
 Gly Lys Glu Ala Ser Thr Ala Ala Leu Arg Ser Pro Thr Leu Arg  
 125 130 135  
 Glu Ala Ala Ser Ser Cys Lys Leu Arg Leu Trp Tyr His Ala Ala  
 140 145 150  
 Ser Gly Asp Val Ala Glu Leu Arg Val Glu Leu Thr His Gly Ala  
 155 160 165  
 Glu Thr Leu Thr Leu Trp Gln Ser Thr Gly Pro Trp Gly Pro Gly  
 170 175 180  
 Trp Gln Glu Leu Ala Val Thr Thr Gly Arg Ile Arg Gly Asp Phe  
 185 190 195  
 Arg Val Thr Phe Ser Ala Thr Arg Asn Ala Thr His Arg Gly Ala  
 200 205 210  
 Val Ala Leu Asp Asp Leu Glu Phe Trp Asp Cys Gly Leu Pro Thr  
 215 220 225  
 Pro Gln Ala Asn Cys Pro Pro Gly His His His Cys Gln Asn Lys  
 230 235 240  
 Val Cys Val Glu Pro Gln Gln Leu Cys Asp Gly Glu Asp Asn Cys  
 245 250 255  
 Gly Asp Leu Ser Asp Glu Asn Pro Leu Thr Cys Gly Arg His Ile  
 260 265 270  
 Ala Thr Asp Phe Glu Thr Gly Leu Gly Pro Trp Asn Arg Ser Glu  
 275 280 285  
 Gly Trp Ser Arg Asn His Arg Ala Gly Gly Pro Glu Arg Pro Ser  
 290 295 300  
 Trp Pro Arg Arg Asp His Ser Arg Asn Ser Ala Gln Gly Ser Phe  
 305 310 315  
 Leu Val Ser Val Ala Glu Pro Gly Thr Pro Ala Ile Leu Ser Ser  
 320 325 330  
 Pro Glu Phe Gln Ala Ser Gly Thr Ser Asn Cys Ser Leu Val Phe  
 335 340 345  
 Tyr Gln Tyr Leu Ser Gly Ser Glu Ala Gly Cys Leu Gln Leu Phe  
 350 355 360  
 Leu Gln Thr Leu Gly Pro Gly Ala Pro Arg Ala Pro Val Leu Leu  
 365 370 375  
 Arg Arg Arg Arg Gly Glu Leu Gly Thr Ala Trp Val Arg Asp Arg  
 380 385 390  
 Val Asp Ile Gln Ser Ala Tyr Pro Phe Gln Ile Leu Leu Ala Gly  
 395 400 405  
 Gln Thr Gly Pro Gly Gly Val Val Gly Leu Asp Asp Leu Ile Leu  
 410 415 420  
 Ser Asp His Cys Arg Pro Val Ser Glu Val Ser Thr Leu Gln Pro  
 425 430 435  
 Leu Pro Pro Gly Pro Arg Ala Pro Ala Pro Gln Pro Leu Pro Pro  
 440 445 450  
 Ser Ser Arg Leu Gln Asp Ser Cys Lys Gln Gly His Leu Ala Cys  
 455 460 465  
 Gly Asp Leu Cys Val Pro Pro Glu Gln Leu Cys Asp Phe Glu Glu  
 470 475 480  
 Gln Cys Ala Gly Gly Glu Asp Glu Gln Ala Cys Gly Thr Thr Asp  
 485 490 495  
 Phe Glu Ser Pro Glu Ala Gly Gly Trp Glu Asp Ala Ser Val Gly  
 500 505 510  
 Arg Leu Gln Trp Arg Arg Val Ser Ala Gln Glu Ser Gln Gly Ser  
 515 520 525

WO 02/02610

PCT/US01/20704

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

WO 02/02610

PCT/US01/20704

Ala Val Ala Ala Gly Val Ala His Ser Tyr Val Ala Leu Asp Asp  
 950 955 960  
 Leu Leu Leu Gln Asp Gly Pro Cys Pro Gln Pro Gly Ser Cys Asp  
 965 970 975  
 Phe Glu Ser Gly Leu Cys Gly Trp Ser His Leu Ala Gly Pro Gly  
 980 985 990  
 Leu Gly Gly Tyr Ser Trp Asp Trp Gly Gly Ala Thr Pro Ser  
 995 1000 1005  
 Arg Tyr Pro Gln Pro Pro Val Asp His Thr Leu Gly Thr Glu Ala  
 1010 1015 1020  
 Gly His Phe Ala Phe Phe Glu Thr Gly Val Leu Gly Pro Gly Gly  
 1025 1030 1035  
 Arg Ala Ala Trp Leu Arg Ser Glu Pro Leu Pro Ala Thr Pro Ala  
 1040 1045 1050  
 Ser Cys Leu Arg Phe Trp Tyr His Met Gly Phe Pro Glu His Phe  
 1055 1060 1065  
 Tyr Lys Gly Glu Leu Lys Val Leu Leu His Ser Ala Gln Gly Gln  
 1070 1075 1080  
 Leu Ala Val Trp Gly Ala Gly Gly His Arg Arg His Gln Trp Leu  
 1085 1090 1095  
 Glu Ala Gln Val Glu Val Ala Ser Ala Lys Glu Phe Gln Ile Val  
 1100 1105 1110  
 Phe Glu Ala Thr Leu Gly Gly Gln Pro Ala Leu Gly Pro Ile Ala  
 1115 1120 1125  
 Leu Asp Asp Val Glu Tyr Leu Ala Gly Gln His Cys Gln Gln Pro  
 1130 1135 1140  
 Ala Pro Ser Pro Gly Asn Thr Ala Ala Pro Gly Ser Val Pro Ala  
 1145 1150 1155  
 Val Val Gly Ser Ala Leu Leu Leu Leu Met Leu Leu Val Leu Leu  
 1160 1165 1170  
 Gly Leu Gly Gly Arg Arg Trp Leu Gln Lys Lys Gly Ser Cys Pro  
 1175 1180 1185  
 Phe Gln Ser Asn Thr Glu Ala Thr Ala Pro Gly Phe Asp Asn Ile  
 1190 1195 1200  
 Leu Phe Asn Ala Asp Gly Val Thr Leu Pro Ala Ser Val Thr Ser  
 1205 1210 1215  
 Asp Pro

<210> 4  
 <211> 589  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4291779CD1

<400> 4  
 Met Val Gly Gln Met Tyr Cys Tyr Pro Gly Ser His Leu Ala Arg  
 1 5 10 15  
 Ala Leu Thr Arg Ala Leu Ala Leu Ala Leu Val Leu Ala Leu Leu  
 20 25 30  
 Val Gly Pro Phe Leu Ser Gly Leu Ala Gly Ala Ile Pro Ala Pro  
 35 40 45  
 Gly Gly Arg Trp Ala Arg Asp Gly Pro Val Pro Pro Ala Ser Arg

WO 02/02610

PCT/US01/20704

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

WO 02/02610

PCT/US01/20704

```

470          475          480
Tyr Asn Asp Phe Leu His Asp Pro Leu Ser Leu Cys Lys Ala Cys
485
Asn Pro Gln Pro Asn Gly Glu Asn Ala Ile Ser Ala Arg Ser Asp
500
Leu Asn Pro Ala Asn Gly Ser Tyr Pro Phe Gln Ala Leu Arg Gln
515
Arg Ser His Gly Gly Ile Asp Val Lys Val Thr Ser Met Ser Leu
530
Ala Arg Ile Leu Ser Leu Leu Ala Ala Ser Gly Pro Thr Trp Asp
545
Gln Val Pro Pro Phe Gln Trp Ser Thr Ser Pro Phe Ser Gly Leu
560
Leu His Met Gly Gln Pro Asp Leu Trp Lys Phe Ala Pro Val Lys
575
Val Ser Trp Asp
580

```

&lt;210&gt; 5

&lt;211&gt; 671

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4728247CD1

&lt;400&gt; 5

```

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

```

WO 02/02610

PCT/US01/20704

```

.Ser Thr Ser Arg Arg Asp Ser Leu Asp Lys Ser Gly Leu Phe Pro
  215 220 225
Glu Trp Lys Lys Met Ser Ala Pro Lys Ser Gln Val Glu Lys Glu
  230 235 240
Thr Gln Pro Gly Gly Gln Asn Val Val Phe Val Asp Glu Gly Glu
  245 250 255
Met Ile Phe Lys Lys Asn Thr Arg Lys Ile Leu Arg Pro Ser Glu
  260 265 270
Tyr Thr Lys Ser Val Ile Asp Leu Arg Pro Glu Asp Val Val His
  275 280 285
Glu Ser Gly Ser Leu Gly Asp Arg Ser Lys Ser Val Pro Gly Leu
  290 295 300
Asn Val Asp Met Glu Glu Glu Glu Glu Glu Glu Asp Ile Asp His
  305 310 315
Leu Val Lys Leu His Arg Gln Lys Leu Ala Arg Ser Ser Met Gln
  320 325 330
Ser Gly Ser Ser Met Ser Thr Ile Gly Ser Met Met Ser Ile Tyr
  335 340 345
Ser Glu Ala Gly Asp Phe Gly Asn Ile Phe Val Thr Gly Arg Ile
  350 355 360
Ala Phe Ser Leu Lys Tyr Glu Gln Gln Thr Gln Ser Leu Val Val
  365 370 375
His Val Lys Glu Cys His Gln Leu Ala Tyr Ala Asp Glu Ala Lys
  380 385 390
Lys Arg Ser Asn Pro Tyr Val Lys Thr Tyr Leu Leu Pro Asp Lys
  395 400 405
Ser Arg Gln Gly Lys Arg Lys Thr Ser Ile Lys Arg Asp Thr Val
  410 415 420
Asn Pro Leu Tyr Asp Glu Thr Leu Arg Tyr Glu Ile Pro Glu Ser
  425 430 435
Leu Leu Ala Gln Arg Thr Leu Gln Phe Ser Val Trp His His Gly
  440 445 450
Arg Phe Gly Arg Asn Thr Phe Leu Gly Glu Ala Glu Ile Gln Met
  455 460 465
Asp Ser Trp Lys Leu Asp Lys Lys Leu Asp His Cys Leu Pro Leu
  470 475 480
His Gly Lys Ile Ser Ala Glu Ser Pro Thr Gly Leu Pro Ser His
  485 490 495
Lys Gly Glu Leu Val Val Ser Leu Lys Tyr Ile Pro Ala Ser Lys
  500 505 510
Thr Pro Val Gly Gly Asp Arg Lys Lys Ser Lys Gly Gly Glu Gly
  515 520 525
Gly Glu Leu Gln Val Trp Ile Lys Glu Ala Lys Asn Leu Thr Ala
  530 535 540
Ala Lys Ala Gly Gly Thr Ser Asp Ser Phe Val Lys Gly Tyr Leu
  545 550 555
Leu Pro Met Arg Asn Lys Ala Ser Lys Arg Lys Thr Pro Val Met
  560 565 570
Lys Lys Thr Leu Asn Pro His Tyr Asn His Thr Phe Val Tyr Asn
  575 580 585
Gly Val Arg Leu Glu Asp Leu Gln His Met Cys Leu Glu Leu Thr
  590 595 600
Val Trp Asp Arg Glu Pro Leu Ala Ser Asn Asp Phe Leu Gly Gly
  605 610 615
Val Arg Leu Gly Val Gly Thr Gly Ile Ser Asn Gly Glu Val Val
  620 625 630

```

WO 02/02610

PCT/US01/20704

Asp Trp Met Asp Ser Thr Gly Glu Glu Val Ser Leu Trp Gln Lys  
 635 640  
 Met Arg Gln Tyr Pro Gly Ser Trp Ala Glu Gly Thr Leu Gln Leu  
 650 655 660  
 Arg Ser Ser Met Ala Lys Gln Lys Leu Gly Leu  
 665 670

<210> 6  
 <211> 1519  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7472259CD1

<400> 6  
 Met His Arg Glu Arg Asp Gly Val Val Arg Gln Ala Arg Glu Leu  
 1 5 10 15  
 Gln Arg Gln Leu Ala Glu Glu Leu Val Asn Arg Gly His Cys Ser  
 20 25 30  
 Arg Pro Gly Ala Ser Glu Val Ser Ala Ala Gln Cys Arg Cys Arg  
 35 40 45  
 Leu Gln Glu Val Leu Ala Gln Leu Arg Trp Gln Thr Asp Gly Glu  
 50 55 60  
 Gln Ala Ala Arg Ile Arg Tyr Leu Gln Ala Ala Leu Glu Val Glu  
 65 70 75  
 Arg Gln Leu Phe Leu Lys Tyr Ile Leu Ala His Phe Arg Gly His  
 80 85 90  
 Pro Ala Leu Ser Gly Ser Pro Asp Pro Gln Ala Val His Ser Leu  
 95 100 105  
 Glu Glu Pro Leu Pro Gln Thr Ser Ser Gly Ser Cys His Ala Pro  
 110 115 120  
 Lys Pro Ala Cys Gln Leu Gly Ser Leu Asp Ser Leu Ser Ala Glu  
 125 130 135  
 Val Gly Val Arg Ser Arg Ser Leu Gly Leu Val Ser Ser Ala Cys  
 140 145 150  
 Ser Ser Ser Pro Asp Gly Leu Leu Ser Thr His Ala Ser Ser Leu  
 155 160 165  
 Asp Cys Phe Ala Pro Ala Cys Ser Arg Ser Leu Asp Ser Thr Arg  
 170 175 180  
 Ser Leu Pro Lys Ala Ser Lys Ser Glu Glu Arg Pro Ser Ser Pro  
 185 190 195  
 Asp Thr Ser Thr Pro Gly Ser Arg Arg Leu Ser Pro Pro Pro Ser  
 200 205 210  
 Pro Leu Pro Pro Pro Pro Pro Ser Ala His Arg Lys Leu Ser  
 215 220 225  
 Asn Pro Arg Gly Gly Glu Gly Ser Glu Ser Gln Pro Cys Glu Val  
 230 235 240  
 Leu Thr Pro Ser Pro Pro Gly Leu Gly His His Glu Leu Ile Lys  
 245 250 255  
 Leu Asn Trp Leu Leu Ala Lys Ala Leu Trp Val Leu Ala Arg Arg  
 260 265 270  
 Cys Tyr Thr Leu Gln Glu Glu Asn Lys Gln Leu Arg Arg Ala Gly  
 275 280 285  
 Cys Pro Tyr Gln Ala Asp Glu Lys Val Lys Arg Leu Lys Val Lys

WO 02/02610

PCT/US01/20704

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

WO 02/02610

PCT/US01/20704

710 715 720  
 Tyr Asn Tyr Asn Pro Phe Glu Gly Pro Asn Asp His Pro Glu Gly  
 725 730 735  
 Glu Leu Pro Leu Thr Ala Gly Asp Tyr Ile Tyr Ile Phe Gly Asp  
 740 745 750  
 Met Asp Glu Asp Gly Phe Tyr Glu Gly Glu Leu Asp Asp Gly Arg  
 755 760 765  
 Arg Gly Leu Val Pro Ser Asn Phe Val Glu Gln Ile Pro Asp Ser  
 770 775 780  
 Tyr Ile Pro Gly Cys Leu Pro Ala Lys Ser Pro Asp Leu Gly Pro  
 785 790 795  
 Ser Gln Leu Pro Ala Gly Gln Asp Glu Ala Leu Glu Glu Asp Ser  
 800 805 810  
 Leu Leu Ser Gly Lys Ala Gln Gly Met Val Asp Arg Gly Leu Cys  
 815 820 825  
 Gln Met Val Arg Val Gly Ser Lys Thr Glu Val Ala Thr Glu Ile  
 830 835 840  
 Leu Asp Thr Lys Thr Glu Ala Cys Gln Leu Gly Leu Leu Gln Ser  
 845 850 855  
 Met Gly Lys Gln Gly Leu Ser Arg Pro Leu Leu Gly Thr Lys Gly  
 860 865 870  
 Val Leu Arg Met Ala Pro Met Gln Leu His Leu Gln Asn Val Thr  
 875 880 885  
 Ala Thr Ser Ala Asn Ile Thr Trp Val Tyr Ser Ser His Arg His  
 890 895 900  
 Pro His Val Val Tyr Leu Asp Asp Arg Glu His Ala Leu Thr Pro  
 905 910 915  
 Ala Gly Val Ser Cys Tyr Thr Phe Gln Gly Leu Cys Pro Gly Thr  
 920 925 930  
 His Tyr Arg Val Arg Val Glu Val Arg Leu Pro Trp Asp Leu Leu  
 935 940 945  
 Gln Val Tyr Trp Gly Thr Met Ser Ser Thr Val Thr Phe Asp Thr  
 950 955 960  
 Leu Leu Ala Gly Pro Pro Tyr Pro Pro Leu Asp Val Leu Val Glu  
 965 970 975  
 Arg His Ala Ser Pro Gly Val Leu Val Val Ser Trp Leu Pro Val  
 980 985 990  
 Thr Ile Asp Ser Ala Gly Ser Ser Asn Gly Val Gln Val Thr Gly  
 995 1000 1005  
 Tyr Ala Val Tyr Ala Asp Gly Leu Lys Val Cys Glu Val Ala Asp  
 1010 1015 1020  
 Ala Thr Ala Gly Ser Thr Val Leu Glu Phe Ser Gln Leu Gln Val  
 1025 1030 1035  
 Pro Leu Thr Trp Gln Lys Val Ser Val Arg Thr Met Ser Leu Cys  
 1040 1045 1050  
 Gly Glu Ser Leu Asp Ser Val Pro Ala Gln Ile Pro Glu Asp Phe  
 1055 1060 1065  
 Phe Met Cys His Arg Trp Pro Glu Thr Pro Pro Phe Ser Tyr Thr  
 1070 1075 1080  
 Cys Gly Asp Pro Ser Thr Tyr Arg Val Thr Phe Pro Val Cys Pro  
 1085 1090 1095  
 Gln Lys Leu Ser Leu Ala Pro Pro Ser Ala Lys Ala Ser Pro His  
 1100 1105 1110  
 Asn Pro Gly Ser Cys Gly Glu Pro Gln Ala Lys Phe Leu Glu Ala  
 1115 1120 1125  
 Phe Phe Glu Glu Pro Pro Arg Arg Gln Ser Pro Val Ser Asn Leu

WO 02/02610

PCT/US01/20704

1130 1135 1140  
 Gly Ser Glu Gly Glu Cys Pro Ser Ser Gly Ala Gly Ser Gln Ala  
 1145 1150 1155  
 Gln Glu Leu Ala Glu Ala Trp Glu Gly Cys Arg Lys Asp Leu Leu  
 1160 1165 1170  
 Phe Gln Lys Ser Pro Gln Asn His Arg Pro Pro Ser Val Ser Asp  
 1175 1180 1185  
 Gln Pro Gly Glu Lys Glu Asn Cys Tyr Gln His Met Gly Thr Ser  
 1190 1195 1200  
 Lys Ser Pro Ala Pro Gly Phe Ile His Leu Arg Thr Glu Cys Gly  
 1205 1210 1215  
 Pro Arg Lys Glu Pro Cys Gln Glu Lys Ala Ala Leu Glu Arg Val  
 1220 1225 1230  
 Leu Arg Gln Lys Gln Asp Ala Gln Gly Phe Thr Pro Pro Gln Leu  
 1235 1240 1245  
 Gly Ala Ser Gln Gln Tyr Ala Ser Asp Phe His Asn Val Leu Lys  
 1250 1255 1260  
 Glu Glu Gln Glu Ala Leu Cys Leu Asp Leu Trp Gly Thr Glu Arg  
 1265 1270 1275  
 Arg Glu Glu Arg Arg Glu Pro Glu Pro His Ser Arg Gln Gly Gln  
 1280 1285 1290  
 Ala Leu Gly Val Lys Arg Gly Cys Gln Leu His Glu Pro Ser Ser  
 1295 1300 1305  
 Ala Leu Cys Pro Ala Pro Ser Ala Lys Val Ile Lys Met Pro Arg  
 1310 1315 1320  
 Gly Gly Pro Gln Gln Leu Gly Thr Gly Ala Asn Thr Pro Ala Arg  
 1325 1330 1335  
 Val Phe Val Ala Leu Ser Asp Tyr Asn Pro Leu Val Met Ser Ala  
 1340 1345 1350  
 Asn Leu Lys Ala Ala Glu Glu Glu Leu Val Phe Gln Lys Arg Gln  
 1355 1360 1365  
 Leu Leu Arg Val Trp Gly Ser Gln Asp Thr His Asp Phe Tyr Leu  
 1370 1375 1380  
 Ser Glu Cys Asn Arg Gln Val Gly Asn Ile Pro Gly Arg Leu Val  
 1385 1390 1395  
 Ala Glu Met Glu Val Gly Thr Glu Gln Thr Asp Arg Arg Trp Arg  
 1400 1405 1410  
 Ser Pro Ala Gln Gly His Leu Pro Ser Val Ala His Leu Glu Asp  
 1415 1420 1425  
 Phe Gln Gly Leu Thr Ile Pro Gln Gly Ser Ser Leu Val Leu Gln  
 1430 1435 1440  
 Gly Asn Ser Lys Arg Leu Pro Leu Trp Thr Pro Lys Ile Met Ile  
 1445 1450 1455  
 Ala Ala Leu Asp Tyr Asp Pro Gly Asp Gly Gln Met Gly Gly Gln  
 1460 1465 1470  
 Gly Lys Gly Arg Leu Ala Leu Arg Ala Gly Asp Val Val Met Val  
 1475 1480 1485  
 Tyr Gly Pro Met Asp Asp Gln Gly Phe Tyr Tyr Gly Glu Leu Gly  
 1490 1495 1500  
 Gly His Arg Gly Leu Val Pro Ala His Leu Leu Asp His Met Ser  
 1505 1510 1515  
 Leu His Gly His

<210> 7  
 <211> 396

WO 02/02610

PCT/US01/20704

<212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7476740CD1

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

WO 02/02610

PCT/US01/20704

Leu Pro Glu Gly Glu Leu Pro Leu Cys Ser Arg Met Val Ser Gly  
 365 370 375  
 Gln Asp Glu Ser Ser Asp Lys Glu Ile Thr His Ser Val Met Asp  
 380 385 390  
 Ser Gly Arg Lys Glu His  
 395

<210> 8  
 <211> 590  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7473774CD1

<400> 8  
 Met Ser Gly Asp Tyr Glu Asp Asp Leu Cys Arg Arg Ala Leu Ile  
 1 5 10 15  
 Leu Val Ser Asp Leu Cys Ala Arg Val Arg Asp Ala Asp Thr Asn  
 20 25 30  
 Asp Arg Cys Gln Glu Phe Asn Asp Arg Ile Arg Gly Tyr Pro Arg  
 35 40 45  
 Gly Pro Asp Ala Asp Ile Ser Val Ser Leu Leu Ser Val Ile Val  
 50 55 60  
 Thr Phe Cys Gly Ile Val Leu Leu Gly Val Ser Leu Phe Val Ser  
 65 70 75  
 Trp Lys Leu Cys Trp Val Pro Trp Arg Asp Lys Gly Gly Ser Ala  
 80 85 90  
 Val Gly Gly Gly Pro Leu Arg Lys Asp Leu Gly Pro Gly Val Gly  
 95 100 105  
 Leu Ala Gly Leu Val Gly Gly Gly Gly His His Leu Ala Ala Gly  
 110 115 120  
 Leu Gly Gly His Pro Leu Leu Gly Gly Pro His His His Ala His  
 125 130 135  
 Ala Ala His His Pro Pro Phe Ala Glu Leu Leu Glu Pro Gly Ser  
 140 145 150  
 Leu Gly Gly Ser Asp Thr Pro Glu Pro Ser Tyr Leu Asp Met Asp  
 155 160 165  
 Ser Tyr Pro Glu Ala Ala Ala Ala Val Ala Ala Gly Val Lys  
 170 175 180  
 Pro Ser Gln Thr Ser Pro Glu Leu Pro Ser Glu Gly Gly Ala Gly  
 185 190 195  
 Ser Gly Leu Leu Leu Leu Pro Pro Ser Gly Gly Gly Leu Pro Ser  
 200 205 210  
 Ala Gln Ser His Gln Gln Val Thr Ser Leu Ala Pro Thr Thr Arg  
 215 220 225  
 Tyr Pro Ala Leu Pro Arg Pro Leu Thr Gln Gln Thr Leu Thr Ser  
 230 235 240  
 Gln Pro Asp Pro Ser Ser Glu Glu Arg Pro Pro Ala Leu Pro Leu  
 245 250 255  
 Pro Leu Pro Gly Gly Glu Glu Lys Ala Lys Leu Ile Gly Gln Ile  
 260 265 270  
 Lys Pro Glu Leu Tyr Gln Gly Thr Gly Pro Gly Gly Arg Arg Ser  
 275 280 285  
 Gly Gly Gly Pro Gly Ser Gly Glu Ala Gly Thr Gly Ala Pro Cys

WO 02/02610

PCT/US01/20704

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

&lt;210&gt; 9

&lt;211&gt; 431

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7946329CD1

&lt;400&gt; 9

Met Ala Glu Ile Thr Asn Ile Arg Pro Ser Phe Asp Val Ser Pro  
 1 5 10 15  
 Val Val Ala Gly Leu Ile Gly Ala Ser Val Leu Val Val Cys Val  
 20 25 30

WO 02/02610

PCT/US01/20704

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

&lt;210&gt; 10

17/27

WO 02/02610

PCT/US01/20704

<211> 3424  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1577952CB1

```

<400> 10
cgacgtgcy cggaagacg tggggacgca ggcgggtcgt agagagcgtt cagccgtctg 60
tatactccc cagatacctg aaactgacca cctgagtacg ttttccatt gctgagctgt 120
ttccctgata tctggccatg caacggagat caagagggat aaatactgga cttattctac 180
tcctttctca aatctccatc gttgggatca acaatattcc acctgtcacc ctagcaactt 240
tggccctcaa catctggttc ttcttgaacc ctcaagaacc actgtatagc tctctcctta 300
gtgtggagaa gtgttaccag caaaaagact ggcagcgttt actgctctct cccttcacc 360
atgctgatga ttggcatttg tatttcaata tggcatccat gctctggaaa ggaataaac 420
tagaaagagc actgggaagt agatggtttg cctatgttat caccgattt tctgtactta 480
ctggagtgtt atacctgctc ttgcaatttg ctgttggcga atttatggat gaacctgact 540
tcaaaaggag ctgtctgcta ggtttctcag gagttttggt tgctttgaaa gttcttaaca 600
accattattg cctggggagg ttgtcaaca ttttggcctt tctctaccg aacagatttg 660
ctgttgggt cgaactgtgt gctattcatt tattctcacc agggacttcc ttcgggtggc 720
atcgtgctgg gattctgttt ggactaatgt acactcaagg gctctggaag aaaatcattg 780
aagcatgtgc agggcgtttt tctccatgct tgggttacc aggacggcaa tactacttta 840
atagtccagc cagctctgga taccagattt attatccgca tggcagcca gatcactatg 900
aagaagacc caggaactat gacacgtaca cagcaggact gactgaaga gaacagctcg 960
agagagcatt acaagccagc ctctgggacc gagaaatac cagaatagc ccaccacct 1020
acgggtttca tctctacca gaagaatga ggaacagcgg gcttcacaga ttctgtagcc 1080
agtgggtgtg catctgggga agacatggcc tattctgcta attattgccc atttggctca 1140
ttcccraagc cctcaattca ttttaattca ttttaacaa aagcagagta caccggtatt 1200
gctccagatc gctccatcca cctgggacag tcccattggc cctatgagtc aactcacagc 1260
ttgcygggag tggcctctct cctggccttg ttctgtgcca taaacaggtc acttctctca 1320
tgaagagacc agtttccacg ctcccacttc tcactgtgga ctccagcagc cctctgcttc 1380
ggtctgcttt tgaagactgt gaccttcacc agggagtttt acttacacca gtcgggaaga 1440
ttagtccctc attctgctgt gagtgcctcg tgtttgactt ggcagcgggt tgggagccat 1500
ccccgctcc tcttggaaca ttgccaactg gctgtctcag gaacaggatg tgggtgctt 1560
ggcgaatgt tgtctactc tcccacacc cggcgctcca gctcctcagc tctctgggcc 1620
cctgctctgt gctgtgtttt gcaggccttt cgtctgctc tggattgct ctgcttttat 1680
agaaagtctt attgaagaag tgaagaagag acctaggttg gggaaagctc ctacacacac 1740
cattagtatc agtgacacca gcaatgttag ttcccagccc ctcccagtg gcagcttgtg 1800
tgtccaggag ataggacatc atttaacgca tcagcaaatg agcagcagat gccacataca 1860
gagttagagc aaggcatttg gtggatcgtt cactagagat ctatctgca gaaagtatgt 1920
tttctcat aaaagtgcct cttaattggc cattgtacca gccacttctc ctagccaaat 1980
gtccaaaaca cgccttggg ccccgccacg ttacaatcca cagattgtct gctgagctcg 2040
tttaaggcat ttctgggtgc ttgtgttcca tgaataaaag gacaaaagca gaagatcact 2100
gatgtcttac tgtcaacaga gatattttaa aagagaaag cagaaaaga tcttctttt 2160
ttgatctaca acttatatag ttttctgatt atgcacataa tagatatgcc ttccagatgc 2220
ataagcmeta cactctgaaa gaaatatacc caaatcttag caggggttat cttgggagat 2280
ggagtacatg ggaatttctt ttcttcattt ttataatttt atattactgt cttggaagat 2340
gtgtttatgt gtgtgtgta cttttcaact caggaaaaca tatttaataa catatagca 2400
agaaaacaga cttaaaaata aatactatgt gtccattgag aaaattcaca atataaacag 2460
aaatacaaat aaatacatac acaattttaa agtcacctgt agccctacc tttagagttac 2520
ccagggttaa cattttgtgt gtaattgttt atcaattttt ccgttgatac atccagcaa 2580
ttggagcacc attgacctg gagttttgt tccaaatcca atctgaatt acctggaaga 2640
ggccttgaca cctgcatgga aatgagctaa gaaaccact gggccttgg gagctcttg 2700
gctcctggc tggccagta atactgagc tctttggtt aattataac tgatataaa 2760
ctacatctc tttataaat aaattgtacc tgtgagtcta gaagcttaa atgtgttaa 2820

```

WO 02/02610

PCT/US01/20704

```

atataaat tcaagctaaa tgttactgct ctctccaaa ttctgtaagt ttgactcccg 2880
ttacccaat tagaagtaac ttctttggtt catgccaatt ttatagcatt tggtaattct 2940
gctataaac atcttgcccc tattattaac tgtgcacagc taacaaaagg tgtgctctct 3000
acgtgggaac atggattgtg aatgactctg taatgagggc tgagtcttag ttatctttcc 3060
actcactccc cyctcctccc ttccaacccc aaaggctcac gataggggct cactaaatgt 3120
cagtggttca ccaaagtatt ttttcoattg tattaagagt ccagtcactg tatatggaag 3180
tatittattt tttatttttt tatatcactt gagtccacta gtagtacttc cttgctctgt 3240
ttgactgttc agatacaaa acacgggatt agattttggg tggtaaaatt gtagatacga 3300
tggctgttga tggagtggaa catcttagtg atgtgagaaa ggtcatttta gttataaatg 3360
taaccaaat actttagcac aacaataaag atgttctgga aattaaaaaa aaaaaaaaaa 3420
aagg                                     3424

```

<210> 11  
 <211> 1033  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4983705CB1

```

<400> 11
agcggccyca gctctgaga gcacgaacag cagcgccccc gctgccage cagccagcca 60
gccagactgg actccggccc accgacggcc gctcgogctc cggcccctgt cgcctgctct 120
gccccggacc tgcagctccc cgtcccccc cgtgtcccgc cgcctccctg ccagagagcc 180
aagccccac gccgcgccca ggcctgccc cgcagcagc tcctcgaccg agagccctg 240
ccgcacggcg gacaagtgc cgcgccagca ggtggaccgc ctactcgtgg ggtcgccctg 300
gcgcggctg gaggagccgc tgggttcat caaagtctc cagtggctct ttgctatttt 360
cgccttcggg tctgtggct cctacagcgg ggagacagga gcaatggttc gctgcaacaa 420
cgaagccaag gacgtgagct ccatcatcgt tgcatttggc tatcccttca ggttgccacc 480
gatccaatat gagatgcccc tctgcgatga agagtccage tccaagacca tgcacctcat 540
gggggacttc tctgcacccg ccgagttctt cgtgacctt ggcatcttt ccttcttcta 600
tacctggct gccctagtta tctactcgtc cttccacac ctctacacag agaacaaacg 660
cttcccctg gtggacttct gtgtgactgt ctccctcacc ttcttctggc tggtagctgc 720
agtcgctgg gccaaggccc tgaccgatgt caagggggcc acacagccat ccagctgac 780
agcagccatg tcagtgctc atggagagga agcagtgctc agtgccgggg ccacgcctc 840
tatggcctg gccaacatct ccgtgctctt tggctttatc aactcttccc tgtggccgg 900
gaactgttg tttgtttca aggagacccc gtggcatgga cagggccagg gccaggacca 960
ggaccaggac caggaccagg gccagggctc cagccagggag agtcagctg agcagggagc 1020
agtggaagaag cag                                     1033

```

<210> 12  
 <211> 3902  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1310465CB1

```

<400> 12
cttccatca cagcggcgc tgcctccctt ggcaccacca tgcctctgct cagccaectg 60
ctgcccgcct tggctctgtt cctggcagca ggttctcag gctgggctg ggtccccaac 120
cactgcagga gccctggcca ggcctgtgct aactctgtgt gtagctcag gactgctca 180
gatgaggccc agtggtgta ccacggggcc tgcgccacce tggggccccc cttcgctgt 240
gacttcgagc aggaccctg cggctggcgg gacattagta cctcaggcta cagctggctc 300

```

WO 02/02610

PCT/US01/20704

```

cgagacaggg caggggccc actggaggg cctgggccc actcagacca cacactgggc 360
accgacttgg gctggtacat ggcggttgg acccaccgag ggaagaggg atccaccgca 420
gccctggcct cccaaccct gcgagagga gccctcctt gcaagctgag gctctggtag 480
caagcgccct ctggagatgt gctgaaact cgggtggagc tgaccatgg ccgagagacc 540
ctgaccctgt ggcagagcac agggccctgg ggccttggct ggcagaggtt ggcagtgacc 600
acaggccgca tccggggtga cttccagtg accttctctg ccaccgaaa tgccaccacc 660
aggggcccgt tggctctaga tgacctagag ttctgggact gtggtctgcc cacccccag 720
gccaactgtc ccccgggaca ccaccactgc cagaacaagg tctgctgga gccccagcag 780
ctgtgctgag ggaagacaa ctgcccggac ctgtctgat agaaccact cactgttggc 840
cgccacatag ccaccgactt tgagacagg ctgggcccat ggaaccgctc ggaagctggc 900
tcccggaaac accgtgctgg tggctctgag cgcctcctc ggccaccgcy tgaccacagc 960
cggaacagtg cacagggtcc cttctggtc tccgtggcc agcctggcac cctgtctata 1020
ctctccagcc ccgaattcca agcctcaggc acctccaact gctcgtggt cttctatcag 1080
taactcagtg ggtctgagtc tggctgctc cagctgttcc tgcaactct ggggcccggc 1140
gcccggcctc ccccgtctc gctcggagg cgcaggggg agctgggga cgcctgggtc 1200
cgagaccgtg tgacatcca gagccctac ccttccaga tctctctgc cgggagcaca 1260
ggcccggggg gctcgttggg tctggagcag cctcctctgt ctgaccactg cagaccagtc 1320
tccggaggtc caaccctgca gccctgctc cctggccccc gggcccacc ccccagccc 1380
ctgctcccca gctcgggct ccaggattcc tgcaggcagg ggcatttgc ctgcccggac 1440
ctgtgtgtgc ccccggaaca actgtgtgac ttccaggagc agtgccaggg ggcagagacc 1500
gagcagctct gtgccaccac agactttgag tcccgcagg ctggggcctg ggaagagccc 1560
agctgggggc gctcagctg cgggctgtc tccagccagg agaccaggg gtcagatgca 1620
gctgctctg gccacttctc gctctctgag cgggctggg ggcagctagg cctgagggcc 1680
egggtctcoa caccctcctc tggccttctc gcccaccagc gtgaactcca cctggttat 1740
tatttacaga gccagcccc aggttctctg gcactagtbg tggtagaca cggctcccgg 1800
gagctggcat ggcagccctc gacgagcagt gcaggcatct ggaagtgga caaggtcctt 1860
ctagggcccc gcccgccccc cttccgctg gagtttctg gtttggtaga cttggatggc 1920
cctgaccagc agggagctgg ggtggaaca gtgacctga gggactgtag ccccacagt 1980
accaccgaga gagacagaga ggtctcctgt aactttgagc gggacacaty cagctggtac 2040
ccagggccac tctcagacac aactggcgc tgggtggaga gcccgggccc tgaccacgac 2100
cacaccacag gccaaaggcca ctttgyctc ctggacccca cagaccctct ggcctggggc 2160
cacagtgccc acctgctctc cagggcccag gtcagcagc caccaccgga gtgtctcagc 2220
ttctgttacc acctccatgg gccccagatt gggactctgc gccatagcca gagagggaa 2280
gggggggaga cacacctgtg gtcgctgca ggcaccaggg gcaaccgctg gcaagagccc 2340
tgggcccacc tttcccacca gccctggctc catgccagc accagctgct gttcagggc 2400
ctccgggagc gataccaccg caccatggcg ctggacgatg tggcctgctg gccgggccc 2460
tgcctggccc ctaattactg ctctcttgag gactcagact gggcttctc cctggaggc 2520
caaggtctct ggaggcgca gcccaatgcc tgggacctg ctgctgggg cccccaaca 2580
gaccatacca ctgagacagc ccaaggcacc tacatggttg tggacacaag cccagacgca 2640
ctacccccgg gccagacggc ctccctgacc tccaaggagc acaggccctc ggcaccgct 2700
gcttctctga ccttctgcta ccacgggagc ctccgagccc caggcaccct cggggtctac 2760
ctggaggagc cggggaggca ccagggtctc agcctcagtg cccacggcgg gcttgcctgg 2820
cgctgggcca gcatggagct gcaggccgag cagacctgga ggggtgtgt tgaggcagtg 2880
gcccgagggc tggcacaact ctacgtggct ctggatgatc tggctctcca ggaagggccc 2940
tgcctcagc caggttcctg tgattttgag tctggctgtg tggctggag ccactggcc 3000
gggcccggcc tgggaggata cagctgggac tggggcgggg gagccacccc ctctgttac 3060
ccccagcccc ctgtggacca caccctgggc acagaggcag gccactttgc ctcttttaa 3120
actgggctgc tgggcccggg gggcccggcc gectggctgc gccagcagcc tctgcccggc 3180
accaccgctc cctgctccg cttctgttac cacatgggtt ttctgagca cttctacaag 3240
ggggagctga aggtactgct gcacagtgtc cagggccagc ttgctgtgtg gggcgaggg 3300
gggcatcggc ggcaccagtg gctggagccc caggtggagg tagccagtc caaggattc 3360
cagatcgtat ttgaagccac tctggggccc cagccagccc tggggcccat tgcctggat 3420
gagctggagt atctgctgg gcagcattgc cagcagcctg ccccagccc ggggaacaca 3480
gccgaccgg ggtctctgac agctgtggtt ggcagtgccc tctctatgct catgctctg 3540
gtgctgctgg gacttggggg acggcctgg ctgcagaaga aggggagctg cccctccag 3600
agcaacacag agggccacag ccttggcttt gacaacatcc tttcaatgc ggttggtytc 3660

```

WO 02/02610

PCT/US01/20704

```

aacctcccgg catctgtcac cagtgatccg tagaccacco cagacaaggc cccgcttcc 3720
cacgtgacat ccagcacttg gtcagacctt agccaggagc cggacaactg ccccgcccag 3780
gctgggacag gctgcaggtc tcaggatatt ctgagccctg ggcgttccct gccctgtgct 3840
gacctctgtt ctctgtgaat aaacacccctg gcccatgagg gcagcccaaa aaaaaaaaaa 3900
aa 3902

```

```

<210> 13
<211> 2574
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4291779CB1

```

```

<400> 13
cggctcaggg tgcggtcagt gtgggcccaga tgtactgcta ccccggcagc cacctggccc 60
gggagctgac ggcggcgctg gcgctggccc tgggtctggc cctgctggtc ggcccgcttc 120
tgagcggcct ggcggggcgc atcccagcgc cggggggcgc ctggcgccgc gatgggcccg 180
tcctccagcgc ctcccagcgc cgtctggtgc tctggagctg ctccggcggc cagctgctta 240
tgggtgacgg acgcacacct gaagcctggc cctgggcccac cctcaacaac gccatcccg 300
agactgggtg ggccttccctg gactgggcca caagtggcca atcaatgac agcttgaggg 360
cctatgcagc cgggtgtggtg gaggctgctg tgtcggagga gctcatctac atgcaactgga 420
tgaacacggt ggtgaattac tgggcctcct tcagatgaga agtcggctac tgcagagggc 480
tgaagagctt cctggagggc aacctagagt ggatgcagga agagatggag tcaaacccag 540
actcacctta ctggcaccag gtgcgctgca cctcctgca gctgaaggc ctggagggaca 600
gctacagagg cgtgtgagc ttcccagctg ggaagttcac catcaaaccc ttggggcttc 660
tcctgtgca gctctctggg gacttggaaag aactggagct ggcctggaac aagaccaga 720
tcaaaccttc tctgggctct ggtcctctgt ctgcccctat caagctgctc cctggcaga 780
gtgacctctt ggttgcacc acacactgga acaactacca gcacatgctg cgtgtcatca 840
agaagctact gctccagttc cgggaaggcc cctgggggga ctaccctgct gttcccggca 900
acaagctggt ctctctctcc taacccggca ccatcttctc ctgagacgac ttctacatcc 960
tgggcagtggt gctgggtgaca ctggagacca ccatgggcaa caagaacca gccctgtgga 1020
agtatgtcgg gcccaagggc tgtgtgctgg agtgggtacg caacatcgtg gcccaaccgc 1080
tggcctcggg tggggccacc tgggcagaca tcttcaagag gttcaacagc ggcacgtata 1140
acaaccagtg gatgatcgtg gactacaagg cgttcatccc ggtggggccc agccccggga 1200
gcccgggtgct taccatcctg gacagatccc cggcatggtt ggtgggtgct gacaagacct 1260
cggagctcta ccagaagacc tactgggcca gctacaacat accgtccttc gagactgtgt 1320
tcaatgccag tgggctgcag gcctagtgg cccagtatgg gactgtgtt tcttatgagc 1380
ggagcccccg ggcaccagatc ttccggcggg accagtcact ggtacaagac atggactcca 1440
tggtcaggct gatgaggtac aatgacttcc tccatgccc tctgtcactg tgaagacct 1500
gcaaccacca gcccaatggg gagaatgcta tctccgccc ctccgacctc aaccgggcca 1560
atgctccta ccccttccag gccctcctga agcctcctca tgggggtatc gatgtgaagg 1620
tgaccagcat gtcactggcc aggatcctga gccctctggc ggcagcgggt cccagtggg 1680
accaggtgcc cccgttccag tggagcactc cgcctctcag cggcctgctg cacatggccc 1740
agccagacct ctggaagtcc gcgcctgtca aggtttcatg ggaactgaagt tctgtccctg 1800
ctctgtgctt ttgcccctg ctgaccctgc tcaaggtcac ccccttccca aggccaccg 1860
acttetaact ccagcccctc ctgggggctt cgttctctga ttgggggtct gactcatctc 1920
ctctagagat ggttcacgaa cctgatgggg ctcaagaactg acccctctct tcccggagg 1980
tgggtgggca cctggcgttc tcttctgccc tgcctaact ctcccactct ctgtttctgt 2040
ctgttctcta ctgtctctct ctcaacctca ttcccactc tggggccccct tctctgtct 2100
tctcctctct gaggttttgg gaagttcctg ggcagactc tggggctccc atggggtgga 2160
aggagctgtt tccagcacc ttctcccag tgcattccca cgggtggccc tggagctggt 2220
gagctttgtc tggcgctgt ctccggctgg cattgtctct cccagctctg gccctctgc 2280
tcctcagga agcagtcctc tctgtctctc ttctggcag ctctcttgg gacagaacct 2340
tgaaaccaaa cacaaacca agtttttggc catctgtggc tggaggttct tgaatgtct 2400

```

WO 02/02610

PCT/US01/20704

```

ctctccatgt caggcagagg gtcagccccc atgcttctgc ctcaggcccc accccaacccc 2460
accocaggcc tgcccctcac ctcaggccca taccacagc gcocctgatgg aggaaccaga 2520
ccgcagagctg tgcccaccatt aaacaagagc ggcctgtgaaa aaaaaaaaaa aagg 2574

```

```

<210> 14
<211> 2878
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4728247CB1

```

```

<400> 14
gtgaaagagg cgtgttgtct agtttcaaag gagaggagag aaggcaactc tggtagctct 60
ccttctctgg ttgttttgaa gaaagaagag tagaagaaaa agttgagtaa atcatgtcgg 120
agtacttggg cttttttttt ctgtctgagg aggaaaaagg tttgattctc agtgttctac 180
agcagatga agagtcocgg aaagcagatg agaaaaagat tagggacta aagaatgagt 240
tactggagat aaaaaggaaa ggggccaaga ggggcagcca acactacagt gatcggacct 300
gtgcccggtg ccaggagagc ctggccgctt tgagtcctca aaccaatctc tgcgggggtt 360
gtaatcacct ggtgtgtcgg gactgcccga tacaggaaag caatggtacc tggaggtgca 420
aggtgtcggc caagaaata gagtgaaga aagcaactgg ggactggttt tatgaccaga 480
aagtgaatcg ctttgcttac cgcacaggta gtgagataat caggatgtcc ctgcgccaca 540
aacctgcagt gagtaaaaga gagacagtgg gacagtccct ccttcatcag acacagatgg 600
gtgacatctg gccaggaaga aagatcattc aggagcggca gaaggagccc agtgtgtctat 660
ttgagtgccc aaagctgaaa agtggaaaga gtgcattgga agctgagagt gagagtctgg 720
atagcttcac agctgactcg gatagcactc ccaggagaga ctctctggat aaatctggcc 780
tccttccaga atggaagagc atgtctctgc ccaaatctca agtgaagaa gaactctcagc 840
ctgaggtgca aantgtggtg tttgtggatg aggtgtgagt gatatttaag aagaacacca 900
gaaanaatct caggccttca gactacacta aatctgtgat agatcttccg ccagaagatg 960
tggtaacatga aagtggctcc ttgggagaca gaagcaaatc cgtcccagcc ctcaatgtgg 1020
atatggaaga ggaagaagaa gaagaagaca ttgaccactc agtgaagtta catgcccaga 1080
agctagccag aagcagcatg caaagtggct cctccatgag tacgatcggc agcatgatga 1140
gcattctacg tgaagctggg gatttcggga acatctttgt gactggcagg attgaccttt 1200
ccctgaagta tgaagcagaa acccagatgc tggttgtcca tgtgaaggag tgccatcagc 1260
tggcctatgc tgatgaagcc aagaagcgtc ctaaccata tgtgaagact tacctctcgc 1320
ctgacaagtc ccgccaagga aaaaagaaaa ccagcatcaa gcgggacact gtaatccac 1380
tatatgatga gacgtgagg fatgagatcc cagaatctct cctggcccag aggaacctgc 1440
agttctcagt ttggtcatcat ggtcgttttg gcagaaacac tttccttggg gaggcagaga 1500
tccagatgga ttccctggaag cttgataaga aactggatca ttgcctccct ttacatggaa 1560
agatcagtcg tgagtcoccg actgctctgc catcacacia aggcagatgt gtggtttcat 1620
tgaatacatc ccagcctcc aaaaaccctg ttggaggtga ccggaaaaag agtaaaggtg 1680
gggaaggggg agagctccag gtgtggatca aagaagccaa gaacttgacg gctgccaaag 1740
caggaaggac ttcagacagc tttgtcaagg gatacctcct tcccatgagg aacaagccca 1800
gtaaacgtaa aactcctgtg atgaagaaga cctgtaatcc tcaactacaac catacatttg 1860
tctacaatgg tgtgagcctg gaagatctac agcatatgtg cctggaaactg actgtgtggg 1920
accyggacc cctgcccagc aatgactctc tggaggggtc caggctgggt gttggactgt 1980
ggatcagtaa tggggaagtg gtggactgga tggactcgac tggggaagaa gtgagcctgt 2040
gycagaagat gccagactac ccagggtctt gggcagaagg gaactctgag ctccgttctc 2100
caatggccaa gccagaagctg ggtttatgag tccctgtcct cttctgagg tccagccctg 2160
gccagggcag gtcagagaaa gtgaagaat caagagcaaa gatttataat ttaatgtgta 2220
tgtgtgtatg tgtgtatgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtacaaca 2280
tgtatctctc gcaaatctca ttatctgggc tagagtgatg cagacttggt cttcttttta 2340
aagcagctcc aagaataagc atttctttaa aatgtttctg tgtataatct agtttattt 2400
cagagctcat tttttcttat gttcttataa ggttcaacta acttaaaac agctttttaa 2460
acaaattttt atctttctgc ttgctatcat tgttctactc tccctaggaa gcocctggca 2520

```

WO 02/02610

PCT/US01/20704

```

cctttcgcat taggaccagt ctgggtttta aggcctctggg aagcagggtt ggttagtaaa 2580
gacaggaaatg ttggggagag gtgagtatt ccttctctct tctctctccc aatttatgct 2640
tttaacttat tttctacctg gataaaactt tggaaacttg cttttaaatt taacttttct 2700
agtttttaag cagtttccac cttgcttttg tctaagtctt ttotittgaa tgctaacaga 2760
attcccaagc tttttccagt tctagatata tttactagac ctttggggga cttttataat 2820
ggagctgctt ttgaaaagca ctttaattag ataagtatt ttgactaaat caccagga 2878

```

```

<210> 15
<211> 5628
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472259CE1

```

```

<400> 15
atggccaagg actgcccag ccccttgggc gcgtgcacca agaagccggg ctgctccagc 60
ccggccggcg cagtgtctga gaaccagagg cgggagctgg agaagctacy ggcggagctg 120
gagccggagc gggcaggctg gcgggocgaa cggcggcgtt tccgtgcccg ggagcccagc 180
ctgctgaggg aggcagagcg gtagcggcgg cagctggctt accatctgcy cccaagtg 240
gagccacagc gcagccggga gttgcggcag ctgcaagagg agatgcagcg ggaacgcgag 300
gcccagatcc ggcagctgct gcctggaaac ggagggccag cagcggcagc tgcagcagct 360
gcattgcacc ggagcgcgat ggcctggctc gccaaagccc ggagctccag cggcagctgg 420
ccgagggagct ggtgaaccgc gcccactgta gcccccgggg ggcgtccgag gtttccgagg 480
cgagtgcccg ctgtgcacct caggaaggtt tggcgcagct tccgtggcag actgacggcg 540
agcagggcgg gcgcacccgc tatctgcagg cggcgtcggg ggtggagcgc cagctcttcc 600
tcaagtaaat cctggcgcac ttccgcccgc acccggcttt ctggggatca ccggaccccc 660
aagctgtgca ttccctggaa gaaccctgct cccagacctc cagcggctct tgcacgcccc 720
ccaaaccgcy ctgccaaact ggaatctctag acagcctgag tgcctgaagt ggtgtgctgt 780
cccctctgct aggcctgggt tccctctgct gctccagctc cccagacggc ctgctctcca 840
cgcacgcccag ctcccttgat tgcctctgca ctgctgtgtt ccgctgctt gacagcacc 900
ggagcctccc caaggcctcc aaatccgagg agcggccctc ctcaccagac acctccacc 960
ctgctcccg gaggctctcy ccgccaccat cgcactccc gccgccacca ccaccgtcag 1020
cccacaggaa actcagcaac ccgcccggag gagaaggctc tgagagccag cctcggcag 1080
tccctgactc ctcacccccg ggcctgggccc accacggact gataaagctg aactggctgc 1140
tggccaaggc gttgtgggtg ctgcccggcc gctgttatac cctgcaagag gagaacaagc 1200
agctgcggcg tgcagctgct cectaccagg cagacgagaa ggtgaaccgg ctcaaggtaa 1260
agcgcggcgg gctgaccggg ctgcggcggc gcttagctga ccgcccgcgc gactcagcag 1320
agaccaactc cggggcctgt agcgcgccta taccggcga gaggttgcgc ggcctggagg 1380
tgttccaagt ctttgcggc cagcgcctc gggacctgct ggagcaggcg agcgcggcgc 1440
tggccaagga caagcagatc jaagagctgc ggcaggagt ccacctctg caggcgcgtg 1500
tcgctcggg tccctgcagc gacctgcata ctggaagggg cggcccctgc acccagtgcc 1560
tcaactcag agacttagac cgcctgcagc gcyagtcaca gccgggaagt ctgcccctgc 1620
agaggcagtt gatgctcag cagggcaacg gttggcgttg gcccgaggcg gggggccaga 1680
gcgcaacttg cgaggaggtg cgcggcaga tgcctggcct ggcggcggag ctggaaccagc 1740
ggcggccgga gtcaccagg ctgggcggc agcggccccc ggcggcga cgtggcagg 1800
aggccgagac acagctgcag gcggcgtgc tcaaaaaagc ctgctggcg gaggagaatg 1860
ggcggctgca ggcacaagacc gactgggtgc gaaaggtgga ggcctgagaat agcgaagtgc 1920
ggcggccact ggcggcggcg ggtcaagagc gcgagctccc cggcttgatc gccgaacagc 1980
tgcctgcaica ggcggcggcg gggcaggaca ggcagcagca gctgcaaccg gaccggcaga 2040
aggccctgtg tgacctccat ccttccctga agagatata ggcctccag tgcggcctg 2100
gtcaccctcc tgaacagccc tgggagacca gtcaaatgcc ggagtcocaa gtttaaggta 2160
gcagaagcc caagtccac gcacggcctg aagactagc agtgtcacag cccaacagag 2220
acatacagga gaaaaggga gccctccctg aggagagccc agttgcccct ggggagtcag 2280
ccagtgctcc ccaagtttca gagacagctc ctgcccacca acctctgtcc aagaaaaaca 2340

```

WO 02/02610

PCT/US01/20704

```

gctcccagtc aaactcctcc tctgaggggt cgatgtgggc caccgtgccy tectccceta 2400
ctctggacag ggacacagcc agtgaggtgg atgacctgga gcctgacagc gtgtccctgy 2460
ccctggaaat ggggggtctg gcggctctgy ctgcccccaa gctcaagatc ttcattggtc 2520
agtataacta caaccattt gaggggccca atgatcaccg tgagggtagc ctgcccctca 2580
cagctgggga ctacatata atcttcgggg acatggatga ggatggcttc tatgaggggy 2640
agcttgacga tggccggcgg gggctggtgc cctccaactt cgtggagcag atccggaca 2700
gtcacatccc aggtgctgy cctgccaat ccoctgatct tggcccagc caactccag 2760
ggggcagga tgaagctctg gaggaaaca gcttattatc tgggaaagcc cagggaatgg 2820
tggacagagg gctgtccag atggtcaggg tgggtcccaa gacagaagta gcaacagaga 2880
tcttgatata caagaccgaa gcctgcagc tgggtctgct gcagagcatg gggagcagg 2940
gcctccag accctctctg gggaccaag ggggtctcgg talggtccc atgcagctac 3000
acctgcagaa tgtcacagcc acatcagcca acatcactg ggtctacagc agccaccycc 3060
accoccatgt ggtatatttt gatgaccgag agcatgccct gaccccagcg ggcgtgagct 3120
gtcacacctt ccagggcctg tgccecgcca cgcactaccg ggtcggggtg gaggtycggc 3180
tgcctaggga cttgctgtag gtttatggg gaactatgct ctcccactc acctcgaca 3240
caectttgct aggacctccc taccaccgc tggatgtgct ggtggagcgc catgctctgc 3300
caggtgctct ggtggtcagc tggctccctg tgaccatga ctoagctgg tctccaatg 3360
gagtcacagt caccggttat gctgtgtat cagatgggct taagttttg gaggtycggc 3420
atgcacatct tgggagcacc gtatbtgaa tctcccagct acaggtgccc cteactgyc 3480
agaaggtctc agtgagaacc atgtcaact gtggtgagtc cctgattca gtgctgctc 3540
agatcccaga gacttcttcc atgtgtcacc gatggccaga gactccacce tttagctaca 3600
cttggtggca ccatccacc tacagagtea ccttcccctg ctgcccaccg aagctgtcac 3660
tggctctcct gagtgccaag gccagcccgc acaacctgg aagctgccc gaggcccagg 3720
ccaagttctt agaagcattc ttgaaagac ccccaaggag gcaatcccga gtgtccacc 3780
tgggtctaga aggaatgtt ccgagttcag ggggtggcag ccaagcccag gactgtcag 3840
aggtctggga gggctgtaga aaggacctgc tctttcagaa gactcccag aaccacaggc 3900
caocttcagt cagtgcacag cctggggaga aggaaattg ctaccagcc atgggaccca 3960
gcaaaaagcc tgcctcagga ttcatccatc tacgaccaga gtgtggccc aggaaagaac 4020
cgtgtcagga aaaggtgccc cttgagaggg tacttgcca aaagcaagat gcccaaggtt 4080
tcacacotcc ccagctgggc gccagccaac agtatgcacc tgacttccat aacgttttga 4140
aggagagaga ggggcaactg tgcttgatc tgtggggcac agagagcca gaggagagga 4200
gggagcctga gccccacagc aggcaaggac aagctctggg ggtcaagaga ggggtgccagc 4260
tccatgagcc cagctcggca ctgtgtccag ctccatccgc caaagtcac aagatgccca 4320
gggtgtgccc ccaacagctg gggacgggg ccaacactcc agccagggtc ttgtgtgccc 4380
tctctgatta caaccocctg gtgatgtctg ccaacctcaa ggtgcagag gaggagctgg 4440
tcttccagaa aaggcagttg ctaagagtgt ggggctctca ggaaccccat gattctaac 4500
tcagcagtg caacagcga tgggcaata tccccggcgc cctagtggct gaggatgagg 4560
tggggacaga gcagactgat aggaggtggc gttctccggc ccaagggcac ctgctctctg 4620
tggcccactc cgagactttt caggggctca ccatcccaca gggttcctcc ctggtgctcc 4680
agggaaactc caagagactc ccaactgtga ctecaagat catgatagca gctctggact 4740
atgatcctgg gtaggggcaa atggggggcc agggaaagg caggctggcg ctgaggcag 4800
gagactgtgt catggtttac gggcccatgg atgaccaagg attctattat ggagagttgg 4860
cggccacag gggcctggtt cctgccacc tctggatca catgtcctc catggactc 4920
gagcaagcat ccttcccag gtatggcct ctgctgctc acacctgcc agaggagaag 4980
caagcgttca gacctcaca ccagaccccc tctcaccacc cataagtagc atgtgtcca 5040
agtgcactcg tgttaactg atggtagtcc ttaagcgtcc cctaggctct gaagtagca 5100
ggacttaagc ctgagttatt tgcaaaagc aacacaocaa gccoacccc gagagctga 5160
gaagccattt caaagtgtct gataactatg gcagglatac ggagaagcgc cttttctgt 5220
ggcactgtgt gttttctct gggaggttaa gtttatctgt ccattgctt gtacaaagt 5280
ctcaagaaaa gctcacatct caaaaaagaa aaagcaatc gagggttatt tttggatgt 5340
gggggtgctc tggctcgac atgtgtcacc ccatgtgaca tcaaggttga ttcgctgat 5400
ctgctgact agggcggat cccctctct cctcacact ccatgtcgt cctccagaa 5460
cctgtgtgct caatgaaaga ggatgacct cccogataga ggacgatcgg tcttcatga 5520
agagtaaaag agtagctgcy ctcccctgct agaacctcca aacgagctct cagaatgta 5580
ttttctgctc ctatgtccaa cccctcatta aaatgttcat agaaaaa 5628

```

WO 02/02610

PCT/US01/20704

<210> 16  
<211> 1482  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7476740CB1

<400> 16  
tgggagacag cccccagaca gatgagtgc ggcctctct gagagtgaa tgagcccgga 60  
cggtccctac ctaccaagtc ctgaggagca gggcaccaa cgacgcagc ccgcccocag 120  
ccgccagtga gccgccatg cctctgcta gccggcccg ccggggccc cccatgctg 180  
atcacctgt actcgtgag gaggaccto tccaggtca cctctctct ccaggtcagc 240  
ccgacttg agtccgaaa ctccaagtc ctctgcaag cggagtccag agtcccctc 300  
gaagagatcc agatcatcca catggagca ctctcatcg aggaccact ttccctgggc 360  
tcctacggcc tcaaatagg cgatatcgt gtttactgc agaaggaca tctgggacct 420  
cggctccag gccgtgccc gaaccacct cgtgtagct tcaagtggca tgcggtgct 480  
gggagctcca gctccgtcc acagcacct ggacagcag acagcgcac acccctgccc 540  
cagcgtcac agggcttggc gtcagagag aagtgcccg gcctgcaag tctggcagc 600  
ccgccctga tccgagcat gctctctcc aaccccacg atctgtccct gctcaagaa 660  
cgcaacctc cctggcggga agcctctcc agcggagcc ttgagacct ttctcaagtg 720  
ctgatggagc agcaaggaa aaagccttg agagagcaag agaggttcg tctctacaca 780  
ggcagccac tggatcggga agctcagcc aaaatagaag aggaatccg gcagcaaac 840  
attgaagaaa acatgaatat agcagatgaa gaggcccoc agagtcttg acaagtgcg 900  
atgctcaca ttaactgcaa agtgaatggg catccttga aggtcttgg tgcctgggc 960  
ggccagatga ccattatgag ccagctgtg gccgagcag gtaacatcat gaggctggt 1020  
gaccagcgtg ggcctgggt tgcataagg gtggccacac agagaattat tggctgtgt 1080  
catctagctc agattcaaat tgaagggat tctctacag gctcttctc catacttgag 1140  
gatcaacca tggatagct tctaggccta gatctgtcc ggagacatca atgttccate 1200  
gattgaaga aaaatgtct ggtcctcgg accactggca cgcagactta tttctctct 1260  
gagggagag tgccttatg ctctaggat gtaagtggc aagatgagc ttcggacaag 1320  
gaattacac attcagctat ggattcagga cgaaaagagc attaaagcag gttataata 1380  
tgttaccac ttaggggagc ctcaggctcc cggcaattat aagttaagag cttactggca 1440  
atgtaatcat taaaaaacat cagtaacaac taaaaaaaaa aa 1482

<210> 17  
<211> 2511  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7473774CB1

<400> 17  
gatgtgactg ttaagotgag cttttctccc cggcctcagc cctagatca gacattctct 60  
ctcattacc gccgctggg aacgggtcca cagccccttg tccgacctag aaccccctg 120  
ggctgcgccc ccgcccggcc tcggtgcca cccagagcac gccagagata agccagagc 180  
cagcagggcca ccatgctagg agactacgag gatgacctct gccggcggc actcatctg 240  
gtctcggacc tctgtgctg ggtccagat gctgaccca acgacaggt ccaggagttc 300  
aatgaccgaa tccgaggcta tcccgggt ccagatgca acatctcgt gacctctg 360  
tcggtcatg tgacattctg tggcattgct cttctgggt tctctctct cgtctctg 420  
aagttgtgt ggggtccctg gccggacaag gtagctcgg cagtggcgg tggcccctg 480  
cgcaagacc taggcccctg tctcggctg ccaggcctg taggggag cggccaccac 540  
ctgcygctg gcttgggtg ccactctctg ctgggaggc cacaccaca tggccatg 600

WO 02/02610

PCT/US01/20704

```

gcccaccatc cacccttgc tgagctgctg gagccaggea geotgggggg ttotgacacc 660
cctgagccct cctacttggg catggactcg tatccagagg ctgcagcagc agcagtgccc 720
gctggggtca aaccgagcca aacatccctc gagctgcccc ctgagggggg agcaggctct 780
gggttgctec tctgcccccc cagtggtggg ggcttgcccc gtgccagctc acatcagcag 840
gtcaaaagcc tggcaccac taccaggtac ccagccctgc ccagaccctc caccagcag 900
actctgacct ccagcggga ccccaagctc gagagctgccc caactgcccc gcccttacc 960
ctgcctggag gcgaggaaaa agccaaactc attgggcaga ttaagccaga gctgtaccag 1020
gggactggcc ctggtggccg gcggagcgtt gggggcccag gctctggaga ggcaggcaca 1080
ggggcaccct gtggcgtat cagcttccgc ctgcccagcc tctatggctc ggaccagctg 1140
gtggtgagga tectgaggcc cctggacctc cctgccaagg actccaaagg ctctcagac 1200
ccctacgtca agatctacct gctgacctgac cgaagaaaa agtttcagac caaggtgcac 1260
aggaagacc tgaacccctt cttcaatgag acgtttcaat tctcggtgcc cctggccgag 1320
ctggcccacc gcaaacctgca cttcagctgc tatgactttg accgcttctc gcgcagcagc 1380
ctcatcgccc aggtggtgct ggacaacctc ctggagctgg ccgagcagcc cctgaccgc 1440
ccgctctgga gggacatcgt gggggcggc tcggaaaaag cagatcttgg gtagctcaac 1500
ttctcaactc gctacctccc caccgcccgg cgcctcaacc tgacctcat caaagcctc 1560
aacctcaaaag cgatggacct cactggcttc tcagaccctc acgtgaagcc ctccctgac 1620
agcgaggggc ggcctctgaa gaagcggaaa acctccatca agaagaacac gctgaacccc 1680
acctataatg aggcctggtt gttcagctg gcctccgaga gcgtggagaa cgtggggctc 1740
agcatcgccc tgggtgacta ctagctcacc gggcacaacg aggtgatcgg cgtgtccgt 1800
gtggcccccg acgctgcccga cccgacggcc cgcagcaact gggcagagat gctggccaat 1860
ccccgaagc ccgtggagca ctggcatcag ctagtggagg aaaagactgt gaccagctc 1920
acaaaagacc gcaaaagact atcagagaaa gagaactccg agtgaggggt ctggcctagg 1980
ccccggatcg gaccagctc cctcagacc coactcttcc ctgcccggac cgtgaatca 2040
tctcctgaaa gccataagct ccgagctgct ggtcgggggc agccctggcc ctaggcttc 2100
taacctggga agcagagaga tgagagggag ccggcccagc tctcttctc aggtggggg 2160
tccactgctt tccactgctt ctgtcttttc tccctgggg ctcccctcg agccgaggg 2220
ccatgatgt ctgggggacc cctgcccccc aaaacctctc gctgtctctc gtctctttg 2280
tgtttgtcca agactcgttg tcccaccctc tgttctgccc gtgaatgta atgggccaat 2340
cctctctgtc ctttcagaca caccaacc tgtytccacc ccttctgttc gccaccctc 2400
gcgtctggcc ggtcccccca ctgctgctgc tctcaagcc agaataaaca cactctgtg 2460
gtctcctcc aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 2511

```

```

<210> 18
<211> 1680
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7946329CE1

```

```

<400> 18
ctacctctc atcaggacca gtctgactgc acctgcatcc ttagctcaga gcatccccg 60
agcatcttaa gagctgagcg cagctgacaa ctagggggcc gaccgtcgca ggaggcgtc 120
gctggatacc ttccccctc cctgacctag agctctacag ctgctgctc ggtactgacc 180
gagggttccc agagctgtct taccattgca aaaacgttat agcaacagcc totgattacg 240
acatgctga gatcaccat atccgaacta gctttgatgt gtaaccygtg tggccggcc 300
tcatcgggccc ctctgtgctg gtggtgtgtg tctcggtgac cgtctttgtc tggctatgct 360
gccaccagca ggcagagaaag aagcacaaga acccaccata caagtttatt cacatgctca 420
aaggcatcag catataccca gagaccctca gcaacaagaa gaaatcctc aaagtccgga 480
gagacaaaga tggctcctggg aggaaaggtg gacgtaggaa cctgttgggt gacgcagag 540
aggtgcccct gctaaagcga gacaaagatc ccagggggccc tagctctgga tctgtatag 600
accaattacc catcaaatg gactatgggg aagaactaag gagccctatt acaagccta 660
ccccgggga gaccaaaacc acctctccat catctccaga gtaggatgtc atgctaggat 720
ccctcacctt ctcagtggac tataacttcc cgaaaaagc cctggtggtg acaatccagg 780

```

WO 02/02610

PCT/US01/20704

```
aggcccacgg gctgccagtg atggatgacc a|aacccaggg atctgacccc tacatcaaaa 840
tgaccatcct tcttgacaaa cggcabcggg tgaagaccag agtgctgceg aagaccctgg 900
accctgtgtt tgacgagacc ttcaccttct atggcatccc ctacagccag ctgacggacc 960
tgggtctgca ctctctgtc ctacgctttg accgctcttc togggatgat gtcattggcg 1020
aggtcatggt gccactggca ggggtggacc ccagcacagg caaggtacaa ctgaccaggg 1080
acatcatcaa aaggantatc cagaagtgca tcagcacagg ggaactccag gtgtctctgt 1140
catatcagcc tgtggcacag agaattgacg tgglygtcct caaagccaga cacttgccga 1200
agatggatat cacoggtctc tcaagtaate cttatgtcaa ggtgaacgtc tactacggca 1260
gaagcycat tyccaagaag aaaaccctg tgaagaagtg caotttgaac cccatcttca 1320
atgaatcttt catctacgac atcccactg acctcctgcc tgatatcagc atcgagttec 1380
tcgttatcga ctctgacgc accaccaaga atgagtggt ggggagcctg atctggggg 1440
cacacagtgt cacagccagt ggtgctgaac actggagaga ggtctgag agcccccgca 1500
agcctgtggc caagtggcac agtctgagc agtaactaac ctgttctct ctctctaat 1560
ccccggggc caagctgggg agggatgtgg aggggaaaaa gatgacagag aagtggactc 1620
caaacctcat tttagttgta gaagaaaatt tcttcaaaa caaattccac aaagaacacc 1680
```

## 【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  
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number  
WO 02/002610 A3(51) International Patent Classification: C12N 15/12,  
5/10, C07K 14/47, 16/18, C12Q 1/68, G01N 33/577,  
33/68, A61K 38/17, A01K 67/027

(21) International Application Number: PCT/US01/20704

(22) International Filing Date: 28 June 2001 (28.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/215,465 29 June 2000 (29.06.2000) US  
60/239,384 10 October 2000 (10.10.2000) US  
60/253,630 28 November 2000 (28.11.2000) US(71) Applicant (for all designated States except US): INCYTE  
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

(72) Inventors: and

(75) Inventors/Applicants (for US only): LEE, Ernestine,  
A. [US/US]; 624 Kains Street, Albany, CA 94706 (US).  
LI, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA  
94303 (US). LAL, Preeti [IN/US]; P.O. Box 5142, Santa  
Clara, CA 95056 (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).  
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). DAS, Debopriya [IN/US]; 1179 Bonita Avenue,  
#3, Mountain View, CA 94040 (US). RAMKUMAR,  
Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA  
94555 (US). TRIBOLEY, Catherine, M. [FR/US];  
1121 Tennessee Street, #5, San Francisco, CA 94107 (US).  
LI, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose,  
CA 95123 (US). HAFALIA, April [US/US]; 2227 Calle  
de Primavera, Santa Clara, CA 95054 (US). GANDHI,  
Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park,  
CA 94025 (US). XU, Yuming [US/US]; 1739 Walnut  
Drive, Mountain View, CA 94040 (US). BANDMAN,  
Olga [US/US]; 366 Anna Avenue, Mountain View, CA94043 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Pollon  
Place Way, San Jose, CA 95121 (US). NGUYEN, Dan-  
niel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA  
95118 (US). BURRILL, John, D. [GB/US]; 650 Emerald  
Hill Road, Redwood City, CA 94061 (US). MARCUS,  
Gregory, A. [US/US]; 73 Maple Way, San Carlos, CA  
94070 (US). ZINGLER, Kurt, A. [US/US]; 723 Ashbury  
Street #C, San Francisco, CA 94117 (US). YAO, Montique,  
G. [US/US]; 111 Frederick Court, Mountain View, CA  
94043 (US). GURURAJAN, Rajagopal [US/US]; 5591  
Dent Avenue, San Jose, CA 95118 (US). DING, Li  
[CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306  
(US). WARREN, Bridget, A. [US/US]; 10130 Parwood  
Drive #2, Cupertino, CA 95014 (US). THANGAVELU,  
Kavitha [IN/US]; 1950 Montecito Avenue, #23, Mountain  
View, CA 94043 (US). LEE, Sally [US/US]; 825 East  
Evelyn, #425, Sunnyvale, CA 94086 (US).(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,  
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,  
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,  
MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,  
TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, JI, MC, NL, PT, SI, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:  
19 September 2002For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: SECRETION AND TRAFFICKING MOLECULES

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



WO 02/002610 A3

## 【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 01/20704
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 C12N15/12 C12N5/10 C07K14/47 C07K16/18 C12Q1/68 G01N33/577 G01N33/68 A61K38/17 A01K67/027		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K C12Q G01N A61K A01K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) WPI Data, PAJ, CAB Data, SEQUENCE SEARCH, BIOSIS, EPO-Internal, GENSEQ		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 00 56767 A (HUMAN GENOME SCIENCES INC; ROSEN CRAIG A (US); RUBEN STEVEN M (US)) 28 September 2000 (2000-09-28) SEQ ID No. 39; claims 1-23	11-15
P,X	-& DATABASE GENSEQ SEQUENCE DATABASE [Online] Hinxtion, UK; 2 February 2001 (2001-02-02) HUMAN GENOME SCI INC: "Human secreted protein sequence #39" XP002193348 Genseq Sequence Accession no. AAC59662; abstract --- -/--	11-15
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "Z" document member of the same patent family		
Date of the actual completion of the international search 18 March 2002		Date of mailing of the international search report 21.06.2002
Name and mailing address of the ISA European Patent Office, P. B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx. 31 851 epo nl, Fax: (+31-70) 340-3016		Authorized officer HORNIG H.

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

## INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/US 01/20704

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EP 1 033 401 A (GENSET SA) 6 September 2000 (2000-09-06) SEQ ID No. 23886; claims 1-60	11-15
P,X	-& DATABASE GENSEQ DATABASE [Online] NCBI, US; 6 October 2000 (2000-10-06) GENSET: "Human secreted proteins 5' EST , SEQ ID No: 23886" XP002193416 GENSEQ Sequence Accession no. AAC19811; abstract ---	11-15
X	DATABASE EMBL SEQUENCE DATABASE [Online] Hinxton, UK; 18 June 2000 (2000-06-18) R.H. WATERSTON: "Homo sapiens chromosome 2 clone RP11-803F15" XP002193349 EMBL Sequence Accession no. AC073149; abstract ---	11-15
X	DATABASE EMBL SEQUENCE DATABASE [Online] Hinxton, UK; 23 September 1999 (1999-09-23) J.E. SULSTON AND R. WATERSON: "Homo sapiens BAC clone RP11-395N3 from 2, complete sequence" XP002193417 EMBL Sequence Accession no. AC010735; abstract ---	11-15
A	WO 00 12703 A (INCYTE PHARMA INC ;PATTERSON CHANDRA (US); CORLEY NEIL C (US); YUE) 9 March 2000 (2000-03-09) the whole document ---	
A	WO 99 16875 A (UNIV YALE ;MORROW JON S (US); DEVARAJAN PRASAD (US)) 8 April 1999 (1999-04-08) the whole document ---	
A	WO 98 37094 A (GENETICS INST.) 27 August 1998 (1998-08-27) the whole document ---	
A	WO 93 25699 A (UNIV CALIFORNIA) 23 December 1993 (1993-12-23) the whole document -----	

3

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

<b>INTERNATIONAL SEARCH REPORT</b>	International application No. PCT/US 01/20704
<b>Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)</b>	
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.: (20,21,23 and 24)-partially because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically: see FURTHER INFORMATION sheet PCT/ISA/210	
3. <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).	
<b>Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)</b>	
This International Searching Authority found multiple inventions in this international application, as follows:	
see additional sheet	
1. <input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.	
2. <input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.	
3. <input type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:	
4. <input checked="" type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: (45, 54) complete; (1-44) partially	
<b>Remark on Protest</b>	<input type="checkbox"/> The additional search fees were accompanied by the applicant's protest. <input type="checkbox"/> No protest accompanied the payment of additional search fees.

International Application No. PCT/US 01/20704

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. Claims: (45,54)-complete; (1-44)-partially

An isolated polypeptide sequence of SEQ ID Nos. 1; an isolated polynucleotide sequence of SEQ ID No. 10 encoding said polypeptide; a cell transformed with said polynucleotide sequence; a transgenic organism comprising said polynucleotide sequence; a method for producing said polypeptide encoded by said polynucleotide sequence; an isolated polyclonal and/or monoclonal antibody which binds to said polypeptide sequence; a method for detecting a target polynucleotide sequence comprising SEQ ID No. 10; a composition comprising said polypeptide and a pharmaceutical acceptable excipient; a method for screening a compound for effectiveness as an agonist and/or antagonist of said polypeptide; a method for assessing toxicity of a test compound using said polynucleotide sequence; a diagnostic test for a condition or disease using said antibody; a method for detecting said polypeptide in a sample;

2. Claims: (46,55)-complete, (1-44)-partially

Idem as invention 1 but limited to SEQ ID Nos. 2 and 11;

3. Claims: (47,56)-complete, (1-44)-partially

Idem as invention one but limited to SEQ ID Nos. 3 and 12;

4. Claims: (48,57)-complete, (1-44)-partially

Idem as invention 1 but limited to SEQ ID Nos. 4 and 13;

5. Claims: (49,58)-complete, (1-44)-partially

Idem as invention 1 but limited to SEQ ID Nos. 5 and 14;

6. Claims: (50,59)-complete, (1-44)-partially

Idem as invention 1 but limited to SEQ ID Nos. 6 and 15;

7. Claims: (51,60)-complete, (1-44)-partially

Idem as invention 1 but limited to SEQ ID Nos. 7 and 16;

8. Claims: (52,61)-complete, (1-44)-partially

International Application No. PCT/US 01/20704

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Idem as invention 1 but limited to SEQ ID Nos. 8 and 17;

9. Claims: (53,62)-complete, (1-44)-partially

Idem as invention 1 but limited to SEQ ID Nos. 9 and 18;

International Application No. PCT/US 01/20704

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

## Continuation of Box I.1

Although claims 18,35 and 38 are 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.

Although claim(s) 32 and 34 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.

## Continuation of Box I.2

Claims Nos.: (20,21,23 and 24)-partially

Present claim 20, 21, 23 and 24 relate to a composition comprising an agonist and/or antagonist and a method for treating a disease, comprising administering to a patient in need of such treatment said agonist and/or antagonist without giving a true technical characterization. Moreover no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to antibodies, antisense molecules, ribozymes, polypeptides and nucleic acids, the structure which can be directly derived from SEQ ID Nos. 1 and 10.

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

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 01/20704

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO 0056767	A	28-09-2000	AU 3628600 A	09-10-2000
			EP 1171457 A1	16-01-2002
			WO 0056767 A1	28-09-2000
EP 1033401	A	06-09-2000	EP 1033401 A2	06-09-2000
			JP 2001269182 A	02-10-2001
WO 0012703	A	09-03-2000	AU 5587799 A	21-03-2000
			EP 1108019 A2	20-06-2001
			WO 0012703 A2	09-03-2000
WO 9916875	A	08-04-1999	AU 1063599 A	23-04-1999
			CA 2304484 A1	08-04-1999
			EP 1021539 A1	26-07-2000
			JP 2001518295 T	16-10-2001
			WO 9916875 A1	08-04-1999
WO 9837094	A	27-08-1998	AU 6337398 A	09-09-1998
			EP 0971950 A2	19-01-2000
			JP 2002514073 T	14-05-2002
			WO 9837094 A2	27-08-1998
WO 9325699	A	23-12-1993	AU 4635593 A	04-01-1994
			US 5688936 A	18-11-1997
			WO 9325699 A1	23-12-1993

## フロントページの続き

(51) Int. Cl. <sup>7</sup>	F I	テーマコード(参考)
A 6 1 P 35/00	A 6 1 P 37/02	4 C 0 8 4
A 6 1 P 37/02	A 6 1 P 43/00	4 H 0 4 5
A 6 1 P 43/00	C 0 7 K 1/22	
C 0 7 K 1/22	C 0 7 K 14/47	
C 0 7 K 14/47	C 0 7 K 16/18	
C 0 7 K 16/18	C 1 2 N 1/15	
C 1 2 N 1/15	C 1 2 N 1/19	
C 1 2 N 1/19	C 1 2 N 1/21	
C 1 2 N 1/21	C 1 2 P 21/02	C
C 1 2 N 5/10	C 1 2 P 21/08	
C 1 2 P 21/02	C 1 2 Q 1/02	
C 1 2 P 21/08	C 1 2 Q 1/68	A
C 1 2 Q 1/02	G 0 1 N 33/15	Z
C 1 2 Q 1/68	G 0 1 N 33/50	Z
G 0 1 N 33/15	G 0 1 N 33/53	D
G 0 1 N 33/50	G 0 1 N 33/53	M
G 0 1 N 33/53	G 0 1 N 33/566	
G 0 1 N 33/566	C 1 2 N 5/00	A
	A 6 1 K 37/00	

(81) 指定国 AP(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW

(72) 発明者 リュ、ヤン

アメリカ合衆国カリフォルニア州 9 4 3 0 3 ・ パロアルト・ コリーナウェイ 3 8 8 5

(72) 発明者 ラル、フリーティ

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

(72) 発明者 タング、トム・ワイ

アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・ サンノゼ・ ランウィックコート 4 2 3 0

(72) 発明者 ユエ、ヘンリー

アメリカ合衆国カリフォルニア州 9 4 0 8 7 ・ サニーバイル・ ルイスアベニュー 8 2 6

(72) 発明者 チョーラ、ナリンダー・ケイ

アメリカ合衆国カリフォルニア州 9 4 5 8 7 ・ ユニオンシティ・ # 7 1 2 ・ ユニオンスクエア 3 3

(72) 発明者 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 9 4 5 7 7 ・ サンレアンドロ・ サンティアゴロード 1 4 2 4 4

(72) 発明者 ダス、デボプリバ

アメリカ合衆国カリフォルニア州 9 4 0 4 0 ・ マウンテンビュー・ # 3 ・ ボニータアベニュー 1 1 7 9

(72) 発明者 ランクマール、ジャヤラクシミ

アメリカ合衆国カリフォルニア州 9 4 5 5 5 ・ フレモント・ メイパードサークル 3 4 3 5 9

(72) 発明者 トリボレー、キャサリン・エム

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

- (72)発明者 リュ、デュング・アイナ・エム  
アメリカ合衆国カリフォルニア州9 5 1 2 3・サンノゼ・コイドライブ 2 3 3
- (72)発明者 ハファリア、エープリル  
アメリカ合衆国カリフォルニア州9 5 0 5 4・サンタクララ・コーレデプリマベータ 2 2 2 7
- (72)発明者 ガンディー、アミーナ・アール  
アメリカ合衆国カリフォルニア州9 4 0 2 5・メンロパーク・# 1・ローブルアベニュー 8 3 7
- (72)発明者 スー、ユーミング  
アメリカ合衆国カリフォルニア州9 4 0 4 0・マウンテンビュー・ウォルナットドライブ 1 7 3 9
- (72)発明者 バンドマン、オルガ  
アメリカ合衆国カリフォルニア州9 4 0 4 3・マウンテンビュー・アンナアベニュー 3 6 6
- (72)発明者 エリオット、ビッキー・エス  
アメリカ合衆国カリフォルニア州9 5 1 2 1・サンノゼ・ポルトンプレイスウェイ 3 7 7 0
- (72)発明者 ニュエン、ダニエル・ビー  
アメリカ合衆国カリフォルニア州9 5 1 1 8・サンノゼ・リッジウッドドライブ 1 4 0 3
- (72)発明者 バリル、ジョン・ディー  
アメリカ合衆国カリフォルニア州9 4 0 6 1・レッドウッドシティ・エメラルドヒルロード 6 5 0
- (72)発明者 マーカス、グレゴリー・エイ  
アメリカ合衆国カリフォルニア州9 4 0 7 0・サンカルロス・メープルウェイ 7 3
- (72)発明者 ジングラー、カート・エイ  
アメリカ合衆国カリフォルニア州9 4 1 1 7・サンフランシスコ・#シー・アッシュベリーストリート 7 2 3
- (72)発明者 ヤオ、モニック・ジー  
アメリカ合衆国カリフォルニア州9 4 0 4 3・マウンテンビュー・フレデリックコート 1 1 1
- (72)発明者 ガルラジャン、ラジャゴバル  
アメリカ合衆国カリフォルニア州9 5 1 1 8・サンノゼ・デントアベニュー 5 5 9 1
- (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9 4 3 0 6・パロアルト・# 1 4 6・アルマストリート 3 3 5 3
- (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9 5 0 1 4・クーペルティノー・# 2・パーウッドドライブ 1 0 1 3 0
- (72)発明者 サンガベル、カピサ  
アメリカ合衆国カリフォルニア州9 4 0 4 3・マウンテンビュー・# 2 3・モンテシトアベニュー 1 9 5 0
- (72)発明者 リー、サリー  
アメリカ合衆国カリフォルニア州9 4 0 8 6・サニーバイル・# 4 2 5・イーストイープリン 8 2 5

F ターム(参考) 2G045 AA25 AA40 BB03 BB20 CB01 CB21 DA12 DA13 DA14 DA36  
DA77 FB02 FB03  
4B024 AA01 AA11 BA80 CA04 CA09 CA11 DA02 DA06 EA02 FA02  
HA11 HA14 HA17  
4B063 QA13 QA18 QA19 QQ08 QQ42 QQ79 QR32 QR48 QR55 QR62  
QR77 QS25 QS33 QS34  
4B064 AG01 AG26 AG27 CA19 CA20 CC24 DA01 DA13  
4B065 AA26X AA90X AA93Y AB01 CA24 CA44 CA46  
4C084 AA02 AA06 AA07 BA01 BA22 CA53 DC50 ZA012 ZB072 ZB112  
ZB212 ZB262  
4H045 AA10 AA11 AA20 AA30 BA10 CA40 DA75 DA76 EA20 EA50

FA74 GA26

專利名稱(譯)	分泌和轉運分子		
公開(公告)號	<a href="#">JP2004528002A</a>	公開(公告)日	2004-09-16
申請號	JP2002507862	申請日	2001-06-28
[標]申請(專利權)人(譯)	洞察Genomics公司		
申請(專利權)人(譯)	洞察基因組公司		
[標]發明人	リーアーンステイーンエイ リュヤン ラルプリーティ タングトムワイ ユエヘンリー チョーラナリンダーケイ ボーグンマライアアール ダスデボプリバ ランクマールジャヤラクシミ トリボレーキャサリーンエム リュデユングアイナエム ハファリアエープリル ガンディーアミーナアール スーユーミング バンドマンオルガ エリオットビッキーエス ニュエンダニエルビー バリルジョンディー マーカスグレゴリーエイ ジングラーカートエイ ヤオモニークジー ガルラジャンラジャゴバル デイングリー ワレンブリジットエイ サンガベルカピサ リーサリー		
發明人	リー、アーンステイーン・エイ リュ、ヤン ラル、プリーティ タング、トム・ワイ ユエ、ヘンリー チョーラ、ナリンダー・ケイ ボーグン、マライア・アール ダス、デボプリバ ランクマール、ジャヤラクシミ トリボレー、キャサリーン・エム リュ、デユング・アイナ・エム ハファリア、エープリル ガンディー、アミーナ・アール スー、ユーミング バンドマン、オルガ エリオット、ビッキー・エス ニュエン、ダニエル・ビー バリル、ジョン・ディー		

マーカス、グレゴリー・エイ  
 ジングラー、カート・エイ  
 ヤオ、モニーク・ジー  
 ガルラジャン、ラジャゴバル  
 ディング、リー  
 ワレン、ブリジット、エイ  
 サンガベル、カピサ  
 リー、サリー

IPC分类号	G01N33/50 A61K38/00 A61K45/00 A61P25/00 A61P29/00 A61P35/00 A61P37/02 A61P43/00 C07K1/22 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12P21/02 C12P21/08 C12Q1/02 C12Q1/68 G01N33/15 G01N33/53 G01N33/566
CPC分类号	A01K2217/05 A61P25/00 A61P29/00 C07K14/47
FI分类号	C12N15/00.ZNA.A A61K45/00 A61P25/00 A61P29/00 A61P35/00 A61P37/02 A61P43/00.105 C07K1/22 C07K14/47 C07K16/18 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/02 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 C12N5/00.A A61K37/00
F-TERM分类号	2G045/AA25 2G045/AA40 2G045/BB03 2G045/BB20 2G045/CB01 2G045/CB21 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/DA77 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/BA80 4B024/CA04 4B024/CA09 4B024/CA11 4B024/DA02 4B024/DA06 4B024/EA02 4B024/FA02 4B024/HA11 4B024/HA14 4B024/HA17 4B063/QA13 4B063/QA18 4B063/QA19 4B063/QQ08 4B063/QQ42 4B063/QQ79 4B063/QR32 4B063/QR48 4B063/QR55 4B063/QR62 4B063/QR77 4B063/QS25 4B063/QS33 4B063/QS34 4B064/AG01 4B064/AG26 4B064/AG27 4B064/CA19 4B064/CA20 4B064/CC24 4B064/DA01 4B064/DA13 4B065/AA26X 4B065/AA90X 4B065/AA93Y 4B065/AB01 4B065/CA24 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA06 4C084/AA07 4C084/BA01 4C084/BA22 4C084/CA53 4C084/DC50 4C084/ZA012 4C084/ZB072 4C084/ZB112 4C084/ZB212 4C084/ZB262 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/CA40 4H045/DA75 4H045/DA76 4H045/EA20 4H045/EA50 4H045/FA74 4H045/GA26
優先権	60/215465 2000-06-29 US 60/239384 2000-10-10 US 60/253639 2000-11-28 US
外部リンク	<a href="http://Espacenet">Espacenet</a>

摘要(译)

本发明提供鉴定和编码SAT的人分泌和运输分子 ( SAT ) 和多核苷酸。  
 本发明还提供表达载体，宿主细胞，抗体，激动剂和拮抗剂。本发明还提供了用于诊断，治疗或预防与SAT异常表达有关的疾病的方法。

		(P2004-528002A) (43) 公表日 平成16年9月16日 (2004.9.16)	
(51) Int. Cl. <sup>7</sup>	FI	テーマコード (参考)	
C 1 2 N 15/00	C 1 2 N 15/00	Z N A A	2 G 0 4 5
A 6 1 K 38/00	A 6 1 K 45/00		4 B 0 2 4
A 6 1 K 45/00	A 6 1 P 25/00		4 B 0 6 3
A 6 1 P 25/00	A 6 1 P 29/00		4 B 0 6 4
A 6 1 P 29/00	A 6 1 P 35/00		4 B 0 6 5
		審査請求 未請求	予備審査請求 有 (全 233 頁) 最終頁に続く
(21) 出願番号	特願2002-507862 (P2002-507862)	(71) 出願人	301005050
(86) (22) 出願日	平成13年6月28日 (2001.6.28)		インサイト・ゲノミクス・インコーポレイテッド
(85) 翻訳文提出日	平成14年12月26日 (2002.12.26)		アメリカ合衆国カリフォルニア州94304・パロアルト・ボータードドライブ 3160
(86) 国際出願番号	PCT/US2001/020704	(74) 代理人	100089266
(87) 国際公開番号	W02002/002610		弁理士 大島 陽一
(87) 国際公開日	平成14年1月10日 (2002.1.10)	(72) 発明者	リー、アーンステイーン・エイ
(31) 優先権主張番号	60/215,465		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
(32) 優先日	平成12年6月29日 (2000.6.29)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/239,384		
(32) 優先日	平成12年10月10日 (2000.10.10)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/253,639		
(32) 優先日	平成12年11月28日 (2000.11.28)		
(33) 優先権主張国	米国 (US)		

最終頁に続く