

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2005-501526

(P2005-501526A)

(43) 公表日 平成17年1月20日(2005.1.20)

(51) Int.Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	2 G O 4 5
A 6 1 K 38/00	A 6 1 K 45/00	4 B O 2 4
A 6 1 K 45/00	A 6 1 K 49/00	A 4 B O 2 9
A 6 1 K 49/00	A 6 1 P 1/04	4 B O 6 3
A 6 1 P 1/04	A 6 1 P 1/12	4 B O 6 4
審査請求 未請求 予備審査請求 未請求 (全 280 頁) 最終頁に続く		

(21) 出願番号	特願2003-500229 (P2003-500229)	(71) 出願人	301005050
(86) (22) 出願日	平成14年5月22日 (2002.5.22)		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15年11月25日 (2003.11.25)		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/US2002/018354		
(87) 国際公開番号	W02002/097060	(74) 代理人	100089266
(87) 国際公開日	平成14年12月5日 (2002.12.5)		弁理士 大島 陽一
(31) 優先権主張番号	60/293,768	(72) 発明者	スウォーナカール、アニータ
(32) 優先日	平成13年5月25日 (2001.5.25)		アメリカ合衆国カリフォルニア州94122・サンフランシスコ・#5ディー・ロックスリーアベニュー 8
(33) 優先権主張国	米国 (US)	(72) 発明者	ゴード、アン・イー
(31) 優先権主張番号	60/309,548		アメリカ合衆国カリフォルニア州94550・リバモア・マリーコモン 369
(32) 優先日	平成13年8月1日 (2001.8.1)		最終頁に続く
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/314,400		
(32) 優先日	平成13年8月23日 (2001.8.23)		
(33) 優先権主張国	米国 (US)		

(54) 【発明の名称】 糖質結合タンパク質

(57) 【要約】

本発明はヒトの糖質結合タンパク質(CHOP)、およびCHOP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、CHOP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 (f) 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:1-10 (配列番号 1 乃至 10) 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:1-3、SEQ ID NO:5、SEQ ID NO:7-10 からなる群から選択したアミノ酸配列と少なくとも 90% が同一であるような天然のアミノ酸配列を含むポリペプチド。

(c) SEQ ID NO:4 のアミノ酸配列に対して少なくとも 93% が同一であるような天然アミノ酸配列を含むポリペプチド

(d) SEQ ID NO:6 のアミノ酸配列に対して少なくとも 99% が同一であるような天然アミノ酸配列を含むポリペプチド 10

(e) SEQ ID NO:1-10 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

(f) SEQ ID NO:1-10 からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片

【請求項 2】

SEQ ID NO:1-10 からなる群から選択したアミノ酸配列を含む、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。 20

【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:11-20 からなる群から選択したポリヌクレオチド配列を含む、請求項 4 に記載の単離されたポリヌクレオチド。

【請求項 6】

請求項 3 に記載のポリヌクレオチドに機能的に連結したプロモーター配列を含む組換えポリヌクレオチド。

【請求項 7】

請求項 6 に記載の組換えポリヌクレオチドを用いて形質転換した細胞。 30

【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、

(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連結したプロモーター配列を含む組換えポリヌクレオチドで形質転換される細胞を培養する過程と、

(b) そのように発現した前記ポリペプチドを回収する過程とからなり、前記組換えポリヌクレオチドが、請求項 1 に記載の前記ポリペプチドをコードするポリヌクレオチドに機能的に連結したプロモーター配列を含むことを特徴とする方法。 40

【請求項 10】

前記ポリペプチドが、SEQ ID NO:1-10 からなる群から選択した或るアミノ酸配列を含むことを特徴とする、請求項 9 に記載の方法。

【請求項 11】

請求項 1 に記載のポリペプチドと特異結合するような単離された抗体。

【請求項 12】

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

(a) SEQ ID NO:11-20 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:11-20 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 9 50

0 % が同一であるような天然のポリヌクレオチド配列を含むポリヌクレオチド

(c) (a) のポリヌクレオチドに相補的なポリヌクレオチド

(d) (b) のポリヌクレオチドに相補的なポリヌクレオチド、および

(e) (a) ~ (d) の RNA 等価物

【請求項 13】

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

【請求項 14】

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

10

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

(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、とを含み、前記プローブと前記標的ポリヌクレオチドあるいはその断片との間でハイブリダイゼーション複合体が形成される条件下で、プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズすることを特徴とする方法。

【請求項 15】

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

20

【請求項 16】

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

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

(b) 前記の増幅した標的ポリヌクレオチドまたはその断片の有無を検出し、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。

【請求項 17】

30

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ことを特徴とする組成物。

【請求項 18】

前記ポリペプチドが、SEQ ID NO:1-10からな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17に記載の組成物。

【請求項 19】

機能的CHOPの発現低下に関連する疾患又は病状を治療する方法であって、そのような治療を必要とする患者に対して請求項17に記載の組成物を投与する過程を含むことを特徴とする方法。

【請求項 20】

40

請求項1に記載のポリペプチドのアゴニストとして有効性を確認するために化合物をスクリーニングする方法であって、

(a) 請求項1のポリペプチドを含むサンプルを化合物に曝すステップと、

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

【請求項 21】

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

【請求項 22】

機能的なCHOP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

50

必要な患者に請求項21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23】

請求項1に記載のポリペプチドのアンタゴニストとして有効性を確認するために化合物をスクリーニングする方法であって、

- (a) 請求項1のポリペプチドを含むサンプルを化合物に曝すステップと、
- (b) 前記サンプルにおいてアンタゴニスト活性を検出するステップとを含むことを特徴とするスクリーニング方法。

【請求項24】

請求項23に記載の方法によって同定したアンタゴニスト化合物と、薬剤として許容できる賦形剤とを含むことを特徴とする組成物。

10

【請求項25】

機能的CHOP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、そのような治療を要する患者への請求項24の組成物の投与を含む治療方法。

【請求項26】

請求項1に記載のポリペプチドに特異結合する化合物をスクリーニングする方法であって、

- (a) 請求項1のポリペプチドを適切な条件下で少なくとも1つの試験化合物と混合する過程と、
- (b) 請求項1のポリペプチドの試験化合物との結合を検出し、それによって請求項1のポリペプチドに特異結合する化合物を同定する過程を含む方法。

20

【請求項27】

請求項1に記載のポリペプチドの活性を調節する化合物をスクリーニングする方法であって、

- (a) 請求項1のポリペプチドの活性が許容される条件下で、請求項1のポリペプチドを少なくとも1つの試験化合物と混合する過程と、
- (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のポリペプチドの活性と比較する過程を含み、試験化合物の存在下での請求項1のポリペプチドの活性の変化が、請求項1のポリペプチドの活性を調節する化合物を標示するような方法。

30

【請求項28】

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容させるのに効果的な化合物をスクリーニングする方法であって、

- (a) 前記標的ポリヌクレオチドの発現に好適な条件下で、該標的ポリヌクレオチドを含むサンプルを化合物に曝露する過程と、
- (b) 前記標的ポリヌクレオチドの発現改変を検出する過程と、
- (c) 可変量の前記化合物の存在下と前記化合物の不存在下で、前記標的ポリヌクレオチドの発現を比較する過程とを含むことを特徴とする方法。

【請求項29】

試験化合物の毒性を算定する方法であって、

40

- (a) 核酸を含む生物学的サンプルを前記試験化合物で処理する過程と、
- (b) 処理した前記生物学的サンプルの核酸と、請求項12のポリヌクレオチドの少なくとも20の連続するヌクレオチドを持つプローブをハイブリダイズさせる過程であって、このハイブリダイゼーションが、前記プローブと前記生物学的サンプル中の標的ポリヌクレオチドとの間で特異的なハイブリダイゼーション複合体が形成される条件下で行われ、前記標的ポリヌクレオチドが、請求項12のポリヌクレオチドまたはその断片のポリヌクレオチド配列を含むポリヌクレオチドである、前記過程と、
- (c) ハイブリダイゼーション複合体の収量を定量する過程と、
- (d) 前記処理された生物学的サンプル中のハイブリタイゼーション複合体の量を、処理されていない生物学的サンプル中のハイブリタイゼーション複合体の量と比較する過程

50

とを含み、前記処理された生物学的サンプル中のハイブリタイゼーション複合体の量の差が、前記試験化合物の毒性を標示するような方法。

【請求項 30】

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

- (a) 前記生物学的サンプルと請求項11の抗体との混合を、前記抗体が前記ポリペプチドに結合し、抗体とポリペプチドとの複合体を形成するのに適した条件下で行う過程と、
- (b) 前記複合体を検出する過程とを含み、前記複合体の存在が、前記生物学的サンプル中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 31】

請求項11の抗体であって、

- (a) キメラ抗体
- (b) 単鎖抗体
- (c) Fab断片
- (d) $F(ab')_2$ 断片
- (e) またはヒト化抗体である抗体。

【請求項 32】

請求項 11 に記載の抗体と、許容できる賦形剤とを含む組成物。

【請求項 33】

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

【請求項 34】

前記抗体が標識されることを特徴とする請求項 32 に記載の組成物。

【請求項 35】

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

【請求項 36】

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

- (a) 抗体反応を誘発する条件下で、SEQ ID NO:1-10からなる群から選択したアミノ酸配列またはその免疫原性断片を含むポリペプチドを用いて動物を免疫化する過程と、
- (b) 前記動物から抗体を単離する過程と、
- (c) 前記単離された抗体を前記ポリペプチドでスクリーニングし、それによって、SEQ ID NO:1-10からなる群から選択したアミノ酸配列を含むポリペプチドに特異結合するポリクローナル抗体を同定する過程とを含むような方法。

【請求項 37】

請求項36に記載の方法で産出したポリクローナル抗体。

【請求項 38】

請求項37に記載のポリクローナル抗体及び適切なキャリアーを含む組成物。

【請求項 39】

請求項 11 に記載の抗体の特異性を有するモノクローナル抗体を作製する方法であって、

- (a) 抗体反応を誘発する条件下で、SEQ ID NO:1-10からなる群から選択したアミノ酸配列またはその免疫原性断片を含むポリペプチドを用いて動物を免疫化する過程と、
- (b) 前記動物から抗体産出細胞を単離する過程と、
- (c) 前記抗体産出細胞と不死化した細胞とを融合して、モノクローナル抗体を産出するハイブリドーマ細胞を形成する過程と、
- (d) 前記ハイブリドーマ細胞を培養する過程と、
- (e) SEQ ID NO:1-10からなる群から選択したアミノ酸配列を含むポリペプチドに特異結合するようなモノクローナル抗体を前記培養物から単離する過程とを含むことを特徴とする方法。

【請求項 40】

10

20

30

40

50

請求項39に記載の方法で産出したモノクローナル抗体。

【請求項41】

請求項40に記載のモノクローナル抗体及び適切なキャリアーを含む組成物。

【請求項42】

Fab発現ライブラリをスクリーニングすることにより産出されることを特徴とする請求項11に記載の抗体。

【請求項43】

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11に記載の抗体。

【請求項44】

SEQ ID NO:1-10 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をサンプル中に検出する方法であって、

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

(b) 特異結合を検出する過程とを含み、該特異結合が、SEQ ID NO:1-10からなる群から選択したアミノ酸配列を有するポリペプチドがサンプル中に存在することを標示することを特徴とする方法。

【請求項45】

SEQ ID NO:1-10 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を精製する方法であって、

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

(b) 前記サンプルから前記抗体を分離し、SEQ ID NO:1-10からなる群から選択したアミノ酸配列を含む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。

【請求項46】

マイクロアレイの少なくとも1つが請求項13に記載のポリヌクレオチドであることを特徴とするマイクロアレイ。

【請求項47】

ポリヌクレオチドを含むサンプルの発現プロファイルを作製する方法であって、

(a) サンプル中のポリヌクレオチドを標識化する過程

(b) ハイブリダイゼーション複合体が形成されるのに適した条件下で請求項46のマイクロアレイの要素とサンプル中の標識化ポリヌクレオチドとを接触させる過程と

、

(c) サンプル中のポリヌクレオチドの発現を定量する過程を含む方法

【請求項48】

或る固体基板上の固有の物理的位置に付着された種々のヌクレオチド分子を有するアレイであって、少なくとも1つの前記ヌクレオチド分子が、或る標的ポリヌクレオチドの少なくとも30の連続したヌクレオチド群と特異的にハイブリダイズ可能な最初のオリゴヌクレオチドまたはポリヌクレオチド配列を含み、前記の標的ポリヌクレオチドが請求項12に記載のポリヌクレオチドであることを特徴とするアレイ。

【請求項49】

請求項48に記載のアレイで、前記の最初のオリゴヌクレオチドまたはポリヌクレオチドの配列が前記の標的ポリヌクレオチドの少なくとも30の連続したヌクレオチドに完全に相補的であることを特徴とするアレイ。

【請求項50】

請求項48に記載のアレイで、前記の最初のオリゴヌクレオチドまたはポリヌクレオチドの配列が前記の標的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に完全に相補的であることを特徴とするアレイ。

【請求項51】

請求項48に記載のアレイで、前記の最初のオリゴヌクレオチドまたはポリヌクレオチドの

10

20

30

40

50

配列が前記の標的ポリヌクレオチドに完全に相補的であることを特徴とするアレイ。

【請求項 5 2】

請求項48に記載のアレイで、マイクロアレイであることを特徴とするアレイ。

【請求項 5 3】

請求項48に記載のアレイで、前記のオリゴヌクレオチドまたはポリヌクレオチドの最初の配列を含むヌクレオチド分子にハイブリダイズした前記の標的ポリヌクレオチドを有することを特徴とするアレイ。

【請求項 5 4】

請求項48に記載のアレイで、リンカーが少なくとも1つの前記のヌクレオチド分子と前記の固体基板を連結していることを特徴とするアレイ。

10

【請求項 5 5】

請求項48に記載のアレイで、基板上の固有の物理的位置の各々が複数のヌクレオチド分子を含み、任意の単一の固有の物理的位置でのその複数のヌクレオチド分子は同一の配列を有し、基板上の固有の物理的位置の各々は、基板上の別の固有の物理的位置でのヌクレオチド分子の配列とは異なる配列を有するヌクレオチド分子を含むことを特徴とするアレイ。

【請求項 5 6】

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 5 7】

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。

20

【請求項 5 8】

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 5 9】

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 6 0】

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。

【請求項 6 1】

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 6 2】

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。

30

【請求項 6 3】

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 6 4】

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 6 5】

SEQ ID NO:10のアミノ酸配列を含む請求項1に記載のポリペプチド。

【請求項 6 6】

SEQ ID NO:11のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 6 7】

SEQ ID NO:12のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

40

【請求項 6 8】

SEQ ID NO:13のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 6 9】

SEQ ID NO:14のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 0】

SEQ ID NO:15のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 1】

SEQ ID NO:16のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 2】

SEQ ID NO:17のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

50

【請求項 7 3】

SEQ ID NO:18のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 4】

SEQ ID NO:19のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【請求項 7 5】

SEQ ID NO:20のポリヌクレオチド配列を含む請求項12に記載のポリヌクレオチド。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、糖質結合タンパク質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糖質代謝、細胞増殖異常、自己免疫/炎症性の疾患、生殖系疾患、遺伝疾患、輸送疾患、神経系疾患および癌の診断・治療・予防に関する。本発明はさらに、糖質結合タンパク質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。 10

【背景技術】

【0002】

糖または糖類、澱粉およびセルロースを含む糖質は複数の水酸基を持つアルデヒドまたはケトン化合物である。糖質は哺乳動物の細胞において3つの重要な役割を持っている。糖質は、エネルギー貯蔵分子、燃料および代謝中間体として機能する。糖質は解糖によって分解されてエネルギーを放出するか、あるいはグリコーゲンとして後の使用のために貯蔵することができる。糖質代謝の重要性は、血液のグルコースレベルを維持するために敏感な調節システムが用意されていることからわかる。インスリンとグルカゴンという2つの膵臓ホルモンがそれぞれ細胞によるグルコースの取り込みと貯蔵および細胞によるグルコースの放出を促進する。デオキシリボースとリボースという糖はそれぞれDNAとRNAの構造的サポートの一部を形成していて、糖質機能の2番目の例となっている。3番目として、糖質は分泌性タンパク質および膜タンパク質および脂質の翻訳後修飾の手段を提供する。実際、真核細胞膜の内容の2~10%は、膜の糖タンパク質および糖脂質のオリゴ糖である。糖タンパク質および糖脂質の糖質修飾により、広範な構造多様性が生み出され、これらが主に原形質膜の細胞外の側にあるため、細胞間認識に重要な役割を果たす(Stryer, L. (1988) Biochemistry, W.H. Freeman および Company, New York NY, 298-299、 331-347ページ)。 20 30

【0003】

タンパク質はいくつかの方法で糖質に会合している。糖タンパク、糖脂質、グリコサミノグリカン、およびプロテオグリカンなどの糖質含有高分子は細胞表面や細胞外マトリックスに見つけられる。細胞外マトリックスは、細胞から分泌された多様な糖タンパクおよび糖質結合タンパク質から構成され、細胞表面と密接した組織的な網状にアセンブリされている。周囲のマトリックスと細胞との相互作用は細胞の形、強度、柔軟性、運動性、および接着に影響する。これらの動的な性質は、細胞増殖、細胞分化、組織の構成、および胚発生を制御するシグナル伝達経路と密接に関連している。

【0004】

糖タンパクは共有結合で付着した糖質を有しており、これらの炭水化物はこれらのタンパク質が分泌経路を通過する時に付加されたものである。糖質を結合、修飾、または分解する目的で、いくつかのタンパク質は糖質を含む高分子と非共有結合的に会合する。糖タンパク質には、細胞接着分子、受容体、血液型抗原、成長因子および抗体が含まれる。これらのタンパク質は細胞間認識およびシグナル伝達、神経伝達物質の認識および/または破壊、神経インパルスの伝達および免疫機能などの細胞プロセスに関与している。 40

【0005】

オリゴ糖修飾により、広範な構造多様性が得られる。N-およびO-結合型のオリゴ糖がタンパク質に受け渡され、修飾されるが、これは小胞体(ER)およびゴルジで起こる一連の酵素反応によって達成される。オリゴ糖はフォールディングの途中およびその後でタンパク 50

質を安定化し、膜内でタンパク質を配向させ、タンパク質の可溶性を改善し、リソソーム標的のシグナルの役割を果たす。

【 0 0 0 6 】

重度にグリコシル化された糖タンパク質は、プロテオグリカンとも呼ばれる。軟骨などの結合組織の細胞外マトリックスのプロテオグリカンは重量を支える関節において負荷を分散するために必須である。細胞表面に付着したプロテオグリカンは細胞を細胞外マトリックスに固着させる。細胞外プロテオグリカンおよび細胞表面プロテオグリカンはいずれも成長因子に結合することにより細胞表面受容体への結合とその後のシグナル伝達経路の誘発を促進する(Lodish, H. 他 (1995) Molecular Cell Biology, Scientific American Books, New York NY, 1139-1142ページ)。

10

【 0 0 0 7 】

糖質はグリコサミノグリカン (GAG) も形成するが、これは二糖ユニットの反復から構成される直鎖状の分岐のない多糖である。GAGは遊離状態で存在するか、あるいはコアタンパク質が1つ以上のGAGに結合してできた大分子であるプロテオグリカンの一部として存在する。GAGは細胞表面、細胞内部および細胞外マトリックスに見られる。GAGヒアルロン酸 (HA) は、特に柔らかい結合組織で多数の細胞の細胞外マトリックスに見られ、滑液に豊富である (Pitsillides, A.A. 他 (1993) *Int. J. Exp. Pathol.* 74:27-34)。水および原形質タンパク質のホメオスタシスに関する機能を持つHAは細胞の制御、発生、および分化で重要な役割を果たしていると思われる。

【 0 0 0 8 】

糖脂質は、リン脂質およびコレステロールとともに細胞の膜を形成する。糖脂質の例としては、赤血球上の血液型抗原およびニューロンの髄鞘中のガングリオシドが挙げられる。原形質膜の細胞外側の糖タンパク質と糖脂質の修飾は細胞間認識にとって重要である (Stryer, 前出, pp. 298-299, 331-347; Lodish, 他, 前出, 612-615ページ)。

20

【 0 0 0 9 】

レクチンは、細胞表面の糖質に特異的かつ可逆的に結合し、その結果、細胞の凝集をもたらす細胞外糖タンパク質である (Drickamer, K. および Taylor, M. E. (1993) *Annu. Rev. Cell Biol.* 9:237-264)。この機能は、免疫反応の活性化に特に重要である。レクチンは炎症部位でリンパ球の凝集と細胞分裂促進の刺激を仲介する (Lasky, L.A. (1991) *J. Cell. Biochem.* 45:139-146; Palletta, E. 他 (1989) *J. Immunol.* 143:2850-2857)。

30

【 0 0 1 0 】

レクチンは、糖結合特異性に基づいてサブファミリーに分類される。ガレクチンサブファミリーは、特に、チオール依存性の方法でβガラクトシド糖質部分に結合するレクチンを含む (Hadari, Y.R. 他 (1995) *J. Biol. Chem.* 270:3447-3453)。ガレクチンは広範囲に発現され、発生的に制御されている。すべてのガレクチンはN末端のシグナルペプチドを欠いているため、ガレクチンは通常でない分泌機構によって細胞外に出されていると示唆されている。2つのクラスのガレクチンが、分子量とオリゴマー形成の特徴に基づいて定義されている。ガレクチンは、ガラプチン (galaptin) ドメインとしても知られる、特徴的な糖認識ドメイン (CRD) を含む。このドメインは長さが約140アミノ酸残基で、いくつかの保存された残基を含んでいる (Prosit PD0C00279 Vertebrate galactoside-binding lectin signatureを参照)。

40

【 0 0 1 1 】

もう一つの例はインテレクチンで、これはガラクトフラノシル (galactofuranosyl) 残基および細菌性アラビノガラクトンに結合するCa²⁺依存性のレクチンである。インテレクチンは細菌の糖質認識および微生物への免疫反応の誘発において役割を持っている可能性がある。

【 0 0 1 2 】

糖質修飾酵素

グルタミン：フルクトース-6-リン酸アミドトランスフェラーゼ (GFAT) はアミノトランスフェラーゼとしても知られる酵素であり、L-グルタミンとD-フルクトース-6-リン酸がL

50

-グルタミン酸とD-グルコサミン-6-リン酸を形成する可逆反応を触媒する。この反応はヘキソサミンの生合成経路の律速過程である(ExpASY ENZYME: EC 2.6.1.16)。D-グルコサミン-6-リン酸は糖タンパク質とプロテオグリカンに組み込まれるUDP-N-アセチルグルコサミン(UDP-GlcNAc) およびその他のヘキソサミンの生合成において役割を持っている。GFATはN-およびO-結合型のグリコシル化の前駆体の利用可能性を調節する。グルコサミンはトランスフォーミング成長因子(TGF)- β 1産生を促進する(Kolm-Litty, V. 他 (1998) J. Clin. Invest. 101:160-169)。GFAT活性はタイプIIの糖尿病のインスリン耐性において役割を持っており、GFATの過剰発現はインスリン耐性につながる。ヘキソサミン代謝は、グリコーゲン合成の律速酵素であるグリコーゲンシンターゼをはじめ、PP1G(グリコーゲン結合タンパク質ホスファターゼ)、ピルビン酸キナーゼおよびグルコース輸送体GLUT1を調節しているように見える(McClain, D.A. および Crook, E.D. (1996) Diabetes 45:1003-1009)。

10

【0013】

グルコサミン-6-リン酸デアミナーゼ(GNPDA)はイソメラーゼとしても知られているが、D-グルコサミン-6-リン酸と水がD-フルクトース-6-リン酸とアンモニアを形成する可逆反応を触媒する(ExpASY ENZYME EC 5.3.1.10)。この反応によりヘキソサミン系と解糖経路が連結され、糖タンパク質、糖脂質およびシアル酸含有高分子においてヘキソサミンの異化によるエネルギー源を供給している可能性がある。GNPDAはエネルギー要求量の大きな組織で発現している(Wolosker, H. 他 (1998) FASEB J. 12:91-99)。

【0014】

UDP-グルコースデヒドロゲナーゼ(UDPGD)という酵素はUDP-グルコース、2 NAD⁺および水がUDP-グルクロン酸と2 NADHを形成する可逆反応を触媒する(ExpASY ENZYME EC 1.1.1.22)。UDP-グルクロン酸はGAGの生合成に必要である。GAGはシグナル伝達経路において或る役割を果たしているように見える(Binari, R.C. 他 (1997) Development 124:2623-2632)。

20

【0015】

Man₉-マンノシダーゼはN結合型オリゴ糖の早期プロセッシングに関与する α 1,2-マンノシダーゼ(グリコシルヒドロラーゼ)である。この酵素はMan₉-(GlcNAc)₂ および Man₅-(GlcNAc)₂における α 1,2-マンノシド結合の特異的切断を触媒する。哺乳動物の細胞で複数の α 1,2-マンノシダーゼが同定されており、N-糖タンパク質の特定のクラスのプロセッシングに必要である可能性がある。Man₉-マンノシダーゼは短い細胞質内尾部、単一の膜貫通ドメインおよび大きな管腔触媒ドメインを持つタイプIIの膜タンパク質である。ヒトの腎臓酵素はゴルジに局在している(Bause, E. 他 (1993) Eur. J. Biochem. 217:535-540; Bieberich, E. および Bause, E. (1995) Eur. J. Biochem. 233:644-649)。

30

【0016】

DPM1は小胞体の中にある酵素であり、GDP-マンノースとドリコールリン酸からドリコールリン酸マンノース(DPM)を産生する反応を触媒する。DPM1の活性はDPM2によって調節されるが、DPM2はDPM1を小胞体(ER)に誘導し、ドリコールリン酸への親和性を強める。DPM2は(ER)膜内に存在し、2つの推定上の膜貫通ドメインと1つの推定上のER局在化シグナル(C末端の近く)を含んでいる。

40

【0017】

グリコシル化

グリコシル化は、任意の数の糖質鎖(オリゴ糖)がタンパク質の特定の部位(グリコシル化部位)に共有結合することを指す。グリコシル化はグリコシル化タンパク質(糖タンパク質)のコンフォメーション、安定性、輸送、分泌、抗原性、クリアランスおよび活性に必須な翻訳後修飾である。結合するオリゴ糖の組成はタンパク質に特異的であり、単純な場合(主にマンノース残基)もあり、複雑な場合(N-アセチルグルコサミン(GlcNAc)残基、シアル酸残基およびガラクトース残基が加わる)もある。糖タンパク質は少数の糖質基をもつ場合もあり、分子量的にタンパク質よりも糖質がより大きな割合を有することもある。後者のグリコシル化の程度が重度の糖タンパク質は分子の組成が主に糖質であること

50

を強調するためにプロテオグリカンとも呼ばれる。任意の2つの構成糖質残基の間に形成される糖類結合のタイプ(たとえば α , β , 1,2-, 1,4-) も糖タンパク質の構造と機能の決定に重要な分子的特徴である。

【0018】

グリコシル化により、タンパク質の親水性が増加する。キャリアタンパク質、抗体およびリソソームタンパク質など、多くの糖タンパク質は溶液(たとえば血漿)中に遊離した状態で見出される。その他の糖タンパク質は膜に結合している。膜結合糖タンパク質の場合は、糖質側鎖が膜脂質二重層内における糖タンパク質の配向を決めるのに役立つ。分子のグリコシル化領域は膜の内側または外側の水性環境と相互作用し、糖タンパク質のより疎水性のドメイン(通常はグリコシル化されていない非極性アミノ酸残基からなる)は膜内のリン脂質と相互作用する。

【0019】

オリゴ糖側鎖の追加はアスパラギン(Asn)残基の-NH₂基(N-グリコシル化)またはセリン(Ser)残基の-OH基(O-グリコシル化)で起こる。N-グリコシル化のプロセスは真核細胞の小胞体(ER)で始まり、ゴルジ体で完了する。O-グリコシル化はゴルジだけで起こる。特性が調べられている全ての糖タンパク質のうち、90%はN-グリコシル化されていて、その中にはさらにO-グリコシル化されているものもないものもある。O-グリコシル化だけされているのは僅か10%である(Apweiler, R. 他(1999) *Biochim. Biophys. Acta* 1473:4-8)。約75000のSWISS-PROTタンパク質配列のうちで、殆ど3分の2は推定上のN-グリコシル化部位を含み、自然に置けるこのタンパク質修飾の重要性を示唆している(Apweiler 他, 前出)。N-グリコシル化に関与する生化学的ステップはよく調べられており、以下はその概説である。

【0020】

N-グリコシル化

糖質のタンパク質へのN結合はAsn-X-Ser または Asn-X-Thr (Ser = セリン, Thr = トレオニン, および X = プロリン以外の任意のアミノ酸)という一次アミノ酸配列状況においてアスパラギン(Asn)残基側鎖の窒素原子を通じて起こる。N結合型オリゴ糖の組成はきわめて多様であるが、グリコシル化を担う経路の最初のステップは共通である。2つのN-アセチルグルコサミン(GlcNAc)、9つのマンノース、3つのグルコース残基からなる14残基のコアオリゴ糖が1つのユニットとしてドリコールリン酸供与分子から標的タンパク質の受容Asn残基の-NH₂基に転移される。通常、コアオリゴ糖の3つのグルコース残基はグルコシダーゼIおよびIIによって除去されて、「高マンノースオリゴ糖(high mannose oligosaccharides)」側鎖となる。これらの部分的にプロセッシングされたN結合型糖タンパク質は順次ERからゴルジのcis, medial, および trans層板を通じて輸送される(Bonay, P. 他(1996) *J. Biol. Chem.* 271:37193726)。ゴルジマンノシダーゼI(cis-層板), N-アセチル-グルコサミントランスフェラーゼ(GluNAcT; medial-層板)およびゴルジマンノシダーゼII(trans-層板)を使ってさらにコアマンノース残基を除去するためにオリゴ糖の修飾がさらに行われることもある。ゴルジマンノシダーゼIによってマンノース残基の一部を除去した後、コアオリゴ糖の残りのマンノース残基をゴルジマンノシダーゼIIで除去するためにはGluNAcTによって単一のGlcNAcを付加する必要がある。

【0021】

マンノース-1-リン酸グアニルトランスフェラーゼはタンパク質グリコシル化の早期ステップに関与している。この酵素は糖代謝に関与していて、酵素反応産物は糖タンパク質の合成に向けられる。マンノース-1-リン酸グアニルトランスフェラーゼ1(MPG1)はNDP-ヘキソースピロホスホリラーゼとも呼ばれており、マンノース代謝においてGTPと α D-マンノース1-リン酸の二リン酸とCDP-エタノールアミンへの変換を触媒する。この酵素はCDP-グルコースピロホスホリラーゼと非常に良く似ていて、細胞周期の進行の調節に関与している可能性もある。GTP: α D-マンノース-1-リン酸グアニルトランスフェラーゼ1(MPG1)のcDNAコードが最近、糸状菌(*Trichoderma reesei*)株のcDNAから単離された(Kruszewska, J.S. 他(1998) *Curr. Genet.* 33:44550)。この1.6 kb cDNAのヌクレオチド

配列は、364アミノ酸のタンパク質をコードするORFを示した。配列比較により、出芽酵母 (*Saccharomyces cerevisiae*) グアニルトランスフェラーゼ遺伝子1 (MPG1) と70%の同一性、そして分裂酵母 (*Schizosaccharomyces pombe*) の相同体とは75%の同一性を示した。

【0022】

部分的にプロセッシングされた糖タンパク質の残りのコア糖部分にN-アセチル-グルコサミン、N-アセチルノイラミン酸(シアル酸)、ガラクトースおよびその他の糖部分を付加することにより、複雑なオリゴ糖側鎖ができてくる。これらの修飾はtrans-層板およびtrans-ゴルジネットワーク(TGN)で起こり、N-アセチル-グルコサミルトランスフェラーゼI (GlcNAcTsI)、シアリルトランスフェラーゼ(ST)およびガラクトシルトランスフェラーゼ(GalT)を含む数多くの酵素が関与している。これらの酵素の多くについて、オリゴ糖側鎖の構成糖残基間の特異的な α または β 、1,2-、1,3-、1,4-または1,6-二糖結合をもたらす複数のアイソフォームが存在する。糖質側鎖内の結合の立体化学的性質とタイプが、出来てくる糖タンパク質の全体的な構造と機能に寄与している(Lehninger, A. L. 他(1993) *Principles of Biochemistry*, Worth Publishers, New York NY, 931ページ; Lewin, B. (1997) *GenesVI*, Oxford University Press, New York NY, 1030-1033ページ)。

10

【0023】

ガラクトシルトランスフェラーゼ類は、グリコシルトランスフェラーゼ(糖転移酵素)のサブセットであり、溶液に遊離している糖脂質または糖タンパク質の一部であるN末端アセチルグルコサミン(GlcNAc)オリゴ糖鎖にガラクトース(Gal)を転移する(Kolbinger, F.他(1998) *J. Biol. Chem.* 273:433-440; Amado, M.他(1999) *Biochim. Biophys. Acta* 1473:35-53)。ガラクトシルトランスフェラーゼは細胞表面および可溶性細胞外タンパク質として見出され、ゴルジ体にも存在する。 β 1,3-ガラクトシルトランスフェラーゼは、Gal(β 1-3)GlcNAc結合を有するI型糖鎖を形成する。 β 1,3-ガラクトシルトランスフェラーゼは、短いサイトゾルドメイン、1つの膜貫通ドメイン、および8つの保存された領域を持つ1つの触媒ドメインを有すると考えられる(Kolbinger他、前出; Hennet, T.他(1998) *J. Biol. Chem.* 273:58-65)。マウスUDP-ガラクトース: β -N-アセチルグルコサミン β 1,3-ガラクトシルトランスフェラーゼ-Iでは、領域1がアミノ酸残基の78-83、領域2がアミノ酸残基の93-102、領域3がアミノ酸残基116-119、領域4がアミノ酸残基147-158、領域5がアミノ酸残基の172-183、領域6がアミノ酸残基の203-206、領域7がアミノ酸残基の236-246、領域8がアミノ酸残基の264-275に位置する。マウスUDP-ガラクトース: β -N-アセチルグルコサミン β 1,3-ガラクトシルトランスフェラーゼ-Iの領域8内に存在する1配列の1変異体が、細菌ガラクトシルトランスフェラーゼ類にも存在することから、この配列が、或るガラクトシルトランスフェラーゼ配列モチーフを規定することを示唆する(Hennet, T. 前出)。近年の研究は、brainiacタンパク質は β 1,3-ガラクトシルトランスフェラーゼであることを示唆している(Yuan, Y.他(1997) *Cell* 88:9-11; Hennet 他、前出)。

20

30

【0024】

UDP-Gal:GlcNAc-1,4-ガラクトシルトランスフェラーゼ(-1,4-GalT)が、Gal(β 1-4)GlcNAc結合を有するII型炭化水素鎖の形成を触媒する(Sato, T.他(1997) *EMBO J.* 16:1850-1857)。この酵素の可溶型は膜結合型の切断によって形成される。 β 1,4-ガラクトシルトランスフェラーゼ類に保存されているアミノ酸としては、ジスルフィド結合で連結された2つのシステインと、触媒ドメインにおける推定上のUDP-ガラクトース結合部位とを含む(Yadav, S.及びBrew, K. (1990) *J. Biol. Chem.* 265:14163-14169; Yadav, S. P.及びBrew, K. (1991) *J. Biol. Chem.* 266:698-703; Shaper, N. L.他(1997) *J. Biol. Chem.* 272:31389-31399)。 β 1,4-ガラクトシルトランスフェラーゼは、糖タンパク質または糖脂質上に糖質鎖を合成するのに加えて、いくつかの特定の役割を有する。 β 1,4-ガラクトシルトランスフェラーゼは α -ラクトアルブミンと共にヘテロ二量体の

40

50

一部として、乳腺ラクトースの生成に作用する。精子の表面上の β 1,4-ガラクトシルトランスフェラーゼは、卵を特異的に認識する受容体として働く。細胞表面 β 1,4-ガラクトシルトランスフェラーゼ類はまた、細胞接着、細胞認識、細胞/基底膜相互作用、正常な細胞遊走、および転移性細胞遊走において作用する(Shur, B. (1993) *Curr. Opin. Cell Biol.* 5:854-863; Shaper, J. (1995) *Adv. Exp. Med. Biol.* 376:95104; Masri, K.A. 他 (1988) *Biochem. Biophys. Res. Commun.* 157:657-663)。

【0025】

合成酵素は細胞の適正な機能に重要な役割を果たす糖質修飾酵素のもう一つのクラスである。たとえば、シアリル化複合糖質の生成には、シチジン5'-モノリン酸 N-アセチルノイラミン酸 (CMP-Neu5Ac または CMP-シアル酸)の合成が必要であるが、この反応はCMP-Neu5Ac合成酵素によって触媒される(Munster, A.K. 他 (1998) *Proc. Natl. Acad. Sci. USA* 95:9140-9145)。細胞表面の糖タンパク質および糖脂質のシアル酸はさまざまな組織の適正な構造と機能に寄与している。シアリルトランスフェラーゼ(ST)はグリコシルトランスフェラーゼのサブセットであり、シアル酸を(CMP-シアル酸から)糖タンパク質および糖脂質の糖質基に転移する反応を触媒する。ゴルジの中にはさまざまなタイプIIのこれらの膜タンパク質が存在する。この遺伝子ファミリーのクローンされたメンバーは、N末端の細胞質尾部領域、膜貫通領域および大(L)、小(S)および極小(VS)と名づけられた3つのシアリルモチーフを含む大きな管腔領域を共通してもっている。L-シアリルモチーフは供与基質の結合に寄与し、高度に保存された48~49アミノ酸からなる一続きの中に8つの不変な残基を持っている。23アミノ酸からなるS-シアリルモチーフは供与基質および受容基質の両方の結合に寄与する(Datta, A. 他 (1997) *Indian J. Biochem. Biophys.* 34:157-65)。代表的なシアリルトランスフェラーゼST3GalI (~350 残基)の場合、L、SおよびVS領域はそれぞれアミノ酸138~182, 264~286および309~321に対応する。このファミリーのクローンされたメンバーとしてはその他にST6GalNAcIとST8SialIなどがある。ST6GalNAcIは他の既知のシアリルトランスフェラーゼよりも大きく、500アミノ酸残基からなっている(Tsuiji, S. 他 (1996) *Glycobiology (letter)* 6:vii; Geremia, R. 他 (1997) *Glycobiology (letter)* 7:vvii; Datta, A. 他 (1995) *J. Biol. Chem.* 270:14971500; Datta, A. 他 (1998) *J. Biol. Chem.* 273:96089618; Tsuiji, S. 他 (1998) *J. Biochem.* 120:113)。シアリルトランスフェラーゼは細胞抽出物中に豊富に存在するわけではないが、いくつかはクローンされ、発現されている。少なくとも、1つのインヒビターが合成されている(Horenstein, B. 他 (1996) *J. Am. Chem. Soc.* 118:1037110379)。

【0026】

糖代謝に関与するその他のさまざまな酵素は直接的または間接的にグリコシル化(ERおよびゴルジで起こるイベントの上流にある)に関与する。これらの酵素の多くはもともと細菌と植物で同定されたもので、ヒトについてはそれほどよく調べられていないが、同様の機能を持ったヒトの相同体が存在する可能性がある。たとえば、ADP-グルコースピロホスホリラーゼは細菌と植物の α 1,4グルカン(グリコーゲンまたは澱粉)の生合成の非常に重要なステップを触媒する。すなわちグルコース-1-リン酸とATPから活性化グルコシル供与体であるADP-グルコースを合成するステップである。ADP-グルコースピロホスホリラーゼはアロステリック調節された四量体の酵素である。植物のADP-グルコースピロホスホリラーゼサブユニットには数多くの保存された配列領域がある。さらに、シグネチャパターンであると考えられている領域が3つある。最初の2つの領域はN末端にあり、大腸菌(*Escherichia coli*)の酵素のアロステリック部位および基質結合部位の一部であると提案されている。第3のパターンは酵素の中央部分の保存された領域に対応している。

【0027】

糖質代謝疾患

糖質代謝はいくつかの疾患において変化している。糖尿病は、血糖が異常に高い状態(高血糖)として特徴付けられる。I型糖尿病は膵臓インスリン分泌細胞の自己免疫による損失が原因で発生する。II型糖尿病はグルコースへのインスリン分泌反応の障害およびイン

スリン耐性が原因で発生し、肥満と関連している。低血糖、すなわち血糖が異常に低い状態は薬物使用、糖質代謝酵素の遺伝的欠損、癌、肝臓病、および腎臓病を含むいくつかの原因がある(Berkow, R. 他 (1992) The Merck Manual of Diagnosis and Therapy, インターネット版、8節、91章、Diabetes Mellitus, Hypoglycemia)。

【0028】

タンパク質のグリコシル化に關与する酵素の突然変異はいくつかの疾患を起こす。たとえば、 α マンノシダーゼの突然変異は先天的な赤血球異形成貧血のI型および α Bリソソームマンノシドーシスを引き起こす(Isselbacher, K.J. 他 (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York NY, 2092-2093ページ; および Online Mendelian Inheritance In Man, 224100)。

10

【0029】

グルコシダーゼは、さまざまなグルコシドのグリコシド結合の加水分解を通じて糖質からグルコースを放出する反応を触媒する糖質修飾酵素のもう一つのクラスを代表している。遺伝性疾患のI型ゴーシェ病は血液性異常として特徴付けられているが、白血球の β -グルコシダーゼレベルのアッセイを通じてヘテロ接合性またはホモ接合性の個人からの検出が可能である(Raghavan, S.S. 他 (1980) Am. J. Hum. Genet. 32:158-173)。ゴーシェ病の3つの型全ての患者が、グルコセレブロシドの生分解の最初のステップを触媒するグルコセレブロシダーゼと呼ばれる酵素の欠乏を示す。脳においては、グルコセレブロシドは脳の発生と神経の髄鞘の形成中の複合脂質の代謝回転から生じる。他の組織においては、グルコセレブロシドは主に古い赤血球および白血球の生分解から生じる。

20

【0030】

ガレクチンは、細胞-細胞および細胞-マトリクス間の相互作用に關連する疾患や健康状態で多くの役割を果たす。例えば、ある種のガレクチンは炎症部位に關連し、細胞表面の免疫グロブリンE分子に結合する。更に、ガレクチンが癌転移で重要な役割を果たしている可能性がある。ガレクチンの過剰発現は、ヒトとマウスでの転移潜在性と相関がある。さらに、抗ガレクチン抗体は、細胞凝集や足場非依存性成長などの、細胞形質転換に關連する過程を阻害する。

【0031】

前立腺癌腫瘍抗体1(PCTA-1)としても知られるガレクチン-8は、癌の進行に關与すると考えられる新規なガレクチンである(Su, Z.-Z. 他 (1996) Proc. Natl. Acad. Sci. U S A 93:7252-7257)。ガレクチン-8は浸襲性的前立腺癌と初期段階の前立腺癌で発現されるが、正常な前立腺細胞または良性の前立腺肥大細胞では発現されない。

30

【0032】

糖質代謝の欠陥は癌とも關連付けられている。GAGとプロテオグリカンの発現の減少はヒトの肺癌と關連付けられている(Nackaerts, K. 他 (1997) Int. J. Cancer 74:335-345)。糖質決定基のシアリル-ルイスAとシアリル-ルイスXはしばしばヒトの癌細胞で発現される。これらの決定基、すなわち細胞接着分子E-セレクチンのリガンドは癌細胞の血管内皮への接着に關与し、癌の血行性転移に寄与している(Kannagi, R. (1997) Glycoconj. J. 14:577-584)。細胞表面糖タンパク質のN結合糖質コア構造の改変は、大腸および膵臓の癌に關連付けられている(Schwarz, R.E. 他 (1996) Cancer Lett. 107:285-291)。胃腸の癌において、細胞表面糖脂質および糖タンパク質のSda血液型糖質構造の発現の減少が觀察されている(Dohi, T. 他 (1996) Int. J. Cancer 67:626-631)。

40

【0033】

グリコサミノグリカンレベルの変化はいくつかの自己免疫疾患と關連付けられている。自己免疫性甲状腺疾患および自己免疫性糖尿病の患者にはさまざまなGAGの増加および減少の両方が起こる。GAGに対する抗体が全身性紅斑性狼瘡および自己免疫性甲状腺疾患の患者で見ついている(Hansen, C. 他 (1996) Clin. Exp. Rheum. 14:S59-S67)。グリコサミノグリカンヒアルロン(HA)は腫瘍細胞の接着と遊走を誘導し、HAの小断片は脈管形成を起こす。血清中のHAは肝臓病や慢性関節リウマチのような種々の炎症状態の診断に用いられる。HAの蓄積によって起こる間質性浮腫は、種々の器官で機能障害を起こ

50

す可能性がある (Laurent, T.C. および Fraser, J.R. (1992) FASEB J. 6:2397-2404)。ヒアルロニダーゼは、H A の N - アセチル β - D - グルコサミンと D - グルクロン酸残基の間の 1、4 結合の無作為の加水分解を触媒して H A をオリゴ糖に分解する酵素である。ヒアルロニダーゼは細胞接着、感染、脈管形成、およびシグナル伝達で機能している可能性がある。ヒアルロニダーゼは、生殖、癌、および炎症に関係する。前立腺腫瘍組織でのヒアルロニダーゼ活性は、正常な前立腺と良性の前立腺過形成での活性と比べて有意に高くなっている (Lokeshwar, V.B. 他 (1996) Cancer Res. 56:651-657)。

【 0 0 3 4 】

哺乳類の精巣で発現されるタンパク質でマウスとヒトの精子の原形質膜に存在する PH - 20 は、ヒアルロニダーゼ活性を持っている (Lin, Y. 他 (1994) J. Cell Biol. 125:1157-63)。PH-20によって、精子は哺乳類の卵子の外層を貫通できる。卵子の外層は、H A が多く含まれた細胞外マトリクスに埋め込まれた約 3000 の卵丘細胞から構成されている。この層を貫通することが、受精過程で重要なステップである。PH-20 はまたいくつかの腫瘍細胞でも発現されている。精巣以外からの哺乳動物ヒアルロニダーゼとしては、HYAL1 ヒアルロニダーゼ (ヒトの血清で発現される) とリソソームのヒアルロニダーゼ H Y A L 2 (多くの細胞で発現される) が含まれる (Lepperdinger, G. 他 (1998) J. Biol. Chem. 273:22466-22470)。H Y A L 2 は、一酸化窒素合成酵素のようなシグナル伝達経路に関与する酵素の発現、および脈管形成を誘導する特異な H A 断片を産生する役割をもっている可能性がある。リソソームタイプのヒアルロニダーゼは、サイトカインで制御される過程で肺の線維芽細胞で H A を分解し得る (Sampson, P.M. 他 (1992) J. Clin. Invest. 90:1492-1503)。種々のヘビ、ハチ、スズメバチ、オニダルマオコゼ、カモノハシ、サソリ、およびトカゲなどの多くの動物の毒にはヒアルロニダーゼが含まれている。毒のヒアルロニダーゼは、毒の拡散を助ける因子として作用すると考えられる。

【 0 0 3 5 】

数多くのヒトの病気がタンパク質のグリコシル化の遺伝的または後天的欠損に関連付けられている。糖質欠損糖タンパク質症候群 (CDGS) は数多くの障害および病気におけるグリコシル化の多数の改変を含んでいる。CDGS は遺伝的マルチシステム障害 (hereditary multisystem disorder) (Matthijs, G. 他 (1997) Nat. Genet. 16:88-92) であり、重症の精神運動性遅滞または精神薄弱および血栓症、出血または卒中様発作にみられる凝血異常を引き起こす。CDGS の特徴的な生化学的異常は糖タンパク質の低グリコシル化 (N 結合) である (Freeze, H. および Aebi, M. (1999) Biochim. Biophys. Acta 1455:167-78)。CDGS のタイプに依存するが、糖タンパク質の糖質側鎖は切り詰められるか、タンパク質コアから完全に失われている。CDGS はいくつかの異なるタイプに分類されている。もっとも多く見られる形態である CDGS 1 A 型はホスホマンノムターゼ (PMM1) 欠損が原因で起こる (Matthijs, G. (1998) Am. J. Hum. Genet. 62:54250)。PMM1 は MPG1 (上記) の上流で機能し、D-マンノース-6-リン酸から D-マンノース-1-リン酸への変換を触媒する。後者はタンパク質のグリコシル化の初期ステップに必要である。

【 0 0 3 6 】

CDGS の 2 つ目の形態は CDGS 1 B 型と呼ばれ、これもまた記載されている (Niehues, R. 他 (1998) J. Clin. Invest. 101:1414-1420)。CDGS 1 B 型は精神運動障害と精神薄弱にはならないが、そのかわりにタンパク漏出症、重度の低血糖症、嘔吐、下痢、および先天的な肝線維症を特徴とする胃腸疾患になる。にもかかわらず、CDGS 1 B 型の患者の一部は血栓症および命に関わる出血にかかることがある。ホスホマンノースイソメラーゼ (PMI) の欠損はこの症候群の原因である可能性が最も高いと考えられている。殆どの症状は食事のマンノースサプリメントで抑えることができる (Niehues 他, 前出、Freeze および Aebi, 前出)。この形態の CDGS は利用可能なマンノースの減少を示すヒトの遺伝性代謝異常として最初のものである。

【 0 0 3 7 】

グルコシルトランスフェラーゼ機能の欠陥もいくつかのヒトの病気において重要な役目を果たす。ガラクトシルトランスフェラーゼは自己免疫/炎症疾患に関与しているかもしれ

ない。というのは、自己免疫甲状腺疾患を持つ多くの人々が $\alpha 1,3$ -ガラクトシルトランスフェラーゼの酵素産物に対する抗体を血液内に高レベルで持っているからである(Etienne-Decerf, J. 他 (1987) *Acta Endocrinol.* 115:67-74)。ヒトの卵巣癌細胞株からは異常な切断を受けた可溶性の $\beta 1,4$ -ガラクトシルトランスフェラーゼが分泌される(Uejima, T. 他 (1992) *Cancer Res.* 52:6158-6163)。 $\beta 1,4$ -GalT-欠損形質転換マウスはある実験において成長遅滞を示した(Asano, M. 他 (1997) *EMBO J.* 16:18501857)。その一方で別の研究においてマウスの $\alpha 1,4$ -GalTの標的不活性化は通常致死的であった(Furukawa, K. 他 (1999) *Biochim. Biophys. Acta.* 1473:5466)。3つ目の研究においては、形質転換マウスにおける $\alpha 1,3$ ガラクトシルトランスフェラーゼの構成的過剰発現は尿中のタンパク質分泌の増加、体重減少、毛髪成長への部分的な損傷、および早死につながった(Ikematsu, S. 他 (1999) *Glycoconj. J.* 1999 16:7376)。ガラクトシルトランスフェラーゼは、細胞成長、発生、および分化の調節にも関係があると言われていて、胚形成および腫瘍の発生に重要な役割を果たしている可能性がある。膜結合型のタンパク質分解性切断により、いくつかの場合で得られた分泌性ガラクトシルトランスフェラーゼは、結合糖質に結合することで細胞表面受容体をトリガする可能性があり、また調節されたやり方で細胞表面分子上の糖質を修飾する可能性がある。細胞外糖質部分は発生的に調節されていて、細胞遊走の調節に関与している可能性が高い(Shur, B. 他 (1984) *Mol. Cell. Biochem.* 61:143158; Paulson, J. および Colley, K. (1989) *J. Biol. Chem.* 264:17615-17618)。 $\beta 1,6$ GlcNAcを持つN結合型糖タンパク質の発現は、ヒトの乳癌および大腸癌の腫瘍進行のマーカーとして使用されてきており、ヒトの神経膠腫試料からの星状細胞は正常個人の星状細胞と比較してこれらのタイプの糖タンパク質のレベルが増えていることが見出されている(Yamamoto, H. 他 (2000) *Cancer Res.* 2000 60:134142)。これらの知見は、グルコシルトランスフェラーゼのもう一つのアイソフォームである $\beta 1,6$ GlcNAcTも腫瘍の形成または侵襲性において役割を果たしている可能性を示唆している。

【0038】

シアリルトランスフェラーゼもヒトの病気に関係があるとされている。ヒトの絨毛癌組織で2,6-シアリルトランスフェラーゼのレベルが上がっていることが見出された(しかし2,3-シアリルトランスフェラーゼではレベルが上がっていなかった)。これは転写レベルでの上方制御の結果であると考えられている(Fukushima, K. (1998) *Cancer Res.* 58:43014306)。2,6-シアリルトランスフェラーゼを過渡的にヒトの発癌性神経膠腫細胞株に形質移入すると細胞の侵襲性が低減する(Yamamoto, H. (1997) *J. Neurochem.* 68:25662576)。1ヶ月以上にわたってカロリーの少なくとも3分の1をアルコールから得ていたラットの肝臓において、慢性のアルコール(エタノール)消費により、Gal- $\beta 1,4$ -GlcNAc- $\alpha 2,6$ シアリルトランスフェラーゼ($\alpha 2,6$ -ST)の活性が減少する。肝臓の $\alpha 2,6$ -ST活性は1週間の間アルコールを絶つことにより正常に戻った。核Run OnアッセイとmRNA安定性アッセイによると、 $\alpha 2,6$ -STの減少は $\alpha 2,6$ -ST mRNAの半減期が50%減少した結果であった(Rao, M. (1999) *Metabolism* 48:797803)。臨床的鬱病の患者グループでは血漿 $\alpha 2,6$ -シアリルトランスフェラーゼ活性の有意な減少も見られた。この種の鬱病はグルコルチコイド受容体(GR)機能の変化に起因すると考えられている。これらの知見は、 $\alpha 2,6$ -ST酵素または活性レベルが臨床的鬱病に寄与しているか、あるいは少なくともコルチゾル受容体機能障害の有用な生化学的マーカーであることを示唆している(Maguire, T. 他 (1997) *Biological Psychiatry* 41:11311136)。

【0039】

グリコシル化の欠陥およびその欠陥を引き起こす酵素欠陥にかかわるその他のヒトの病気としては、(i) アスパルチルグルコサミン尿症(精神薄弱を引き起こすアスパルチルグルコサミニダーゼ欠損)、(ii) GM₁ および GM₂ ガングリオシド蓄積症、(それぞれ、 β -ガラクトシダーゼおよび β Nアセチルヘキソサミニダーゼの欠損)(糖脂質貯蔵疾患の原因)、(iii) α マンノシドーシスおよび β -マンノシドーシス(それぞれ α -マンノシダーゼおよび β -マンノシダーゼの欠損に起因する)(神経機能障害の原因)、および(iv) ノイラミン酸欠損に起因するシアリドーシス(肝脾腫大症および神経発生障害が特徴)、な

どがある。

【0040】

発現プロファイル

アレイ技術は、単一の多型遺伝子の発現や、多数の関連遺伝子または無関係の遺伝子の発現プロファイルを探求する簡単な方法を提供し得る。単一遺伝子の発現を試験するときは、アレイを用いて或る特定遺伝子又はその変異体の発現を検出する。発現プロファイルを試験するときは、アレイは次のような試験を行うプラットフォームを提供する。即ちどの遺伝子が組織特異的か、ハウスキーピング機能を行うか、シグナル伝達カスケードの一部であるか、又は、特定の遺伝的素因や、条件、疾患、または障害に、特異的に関連する遺伝子であるかの試験である。

10

【0041】

遺伝子発現プロファイル作製の潜在的応用は特に、疾患の、診断、予後診断、および治療の向上に関わる。例えば、大腸癌患者に由来する組織に発現されるレベルと配列との双方を、正常組織に発現されるレベルおよび配列と比較しうる。

【0042】

結腸癌は遺伝子と環境の両者と因果関係がある。幾つかの分子経路が結腸癌の発生と関係があり、これらの経路いずれかの鍵遺伝子の発現は先天性または後天性突然変異或いは高メチル化により失われる可能性がある。発現における変化が結腸癌または結腸癌発症の素因の早期の指標となりうる遺伝子の同定が特に必要とされている。

【0043】

たとえば、異常なパターンのDNAメチル化がヒトの腫瘍に一貫して発生し、広範なゲノムの低メチル化部分と、局在的にメチル化の増加した部分が同時に含まれることは周知である。特に結腸癌では、結腸癌に先行する前癌性ポリープなど、腫瘍進行においてこれらの変化が早期に発生することが知られる。実際、DNAメチル化を行う酵素であるDNAメチルトランスフェラーゼは、結腸癌または癌に先行する良性ポリープの患者の組織学的に正常な粘膜で著しく増加する。そしてこの増加は結腸腫瘍の進行中に継続する(Wafik 他 (1991) Proc. Natl. Acad. Sci. USA 88:3470-3474)。DNAメチル化の増加は、「CpG島」と呼ばれるゲノムDNAのG+Cリッチ領域に生じ、この「CpG島」は遺伝子の周囲の「開いた」転写構造を維持するために重要であり、これらの領域が過度にメチル化すると、遺伝子転写を止める「閉じた」構造になる。このようなCpG島の異常メチル化による分化遺伝子の静止または下方制御によって不死化細胞の分化を防ぎ得ることが示唆されている(Antequera, F. 他 (1990) Cell 62:503-514)。

20

【0044】

家族性大腸腺腫症(FAP)は大腸癌に先行するまれな常染色体優性症候群であり、大腸腺腫性ポリポーシス(APC)遺伝子の先天性突然変異によって起こる。FAPは平均年令44才で癌に進行する多発性結腸直腸腺腫の早期発症を特徴とする。APC遺伝子は、APC-β-カテニン-Tcf (T-細胞因子)経路の一部である。この経路の障害により、結腸上皮細胞の規則的な複製化、接着、遊走が失われ、ポリープの成長につながる。APC-β-カテニン-Tcf 経路が活性化された後、一連の他の遺伝的变化が続き、正常結腸粘膜から転移性癌への移行を伴う。これらの変化にはK-Ras癌原遺伝子の突然変異、メチル化パターンの変化、および癌抑制遺伝子p53 と Smad4/ DPC4の突然変異または喪失が含まれる。突然変異したAPC遺伝子の遺伝は稀な現象であるが、APCの喪失または突然変異や、APC-β-カテニン-Tcf経路への結果的効果は、一般的集団の大部分の大腸癌患者にとって中心的となると考えられる。

40

【0045】

遺伝性非ポリポーシス結腸直腸癌 (HNPCC) はFAPほどよく定義されていない表現型の、別の先天性常染色体優性症候群である。結腸直腸癌症例の約2%を占めるHNPCCは癌の早期発症の傾向および、他の癌、特に子宮内膜、尿路、胃および胆道系に関連する癌の発症によって区別される。HNPCCは、DNAミスマッチ修復(MMR)経路の1つか複数の遺伝子の突然変異から発生する。2つのヒトMMR遺伝子MSH2およびMLH1の突然変異は、これまでに同定されたHNPCCファミリーの大部分に見出される。DNA MMR経路は複製中のDNAポリメラーゼの

50

作用から生じるエラーを同定し、修復する。さらに、MMR活性が喪失されると、アポトーシスを調節するBAX 遺伝子および細胞増殖を制御するTGF β 受容体II遺伝子の喪失などのような他の遺伝子の突然変異および欠失の蓄積による、癌の進行を生じる。DNA MMR欠損患者では、DNAへの修復できない損傷の可能性があるため、癌への進行は通常より急速である。

【0046】

潰瘍性腸炎は大腸癌に対する原因としては小さいが、罹病した患者は癌を発症する危険性が約20倍になる。進行は、組織学的に正常な組織においても現れ得る、p53 遺伝子の早期の喪失を特徴とする。潰瘍性大腸炎から中間のポリープ状態を経ないで異形成/癌へ疾患が進行することは、大腸粘膜における増殖細胞の大腸内容物への曝露から生じた変異原活性が高度であることを示唆する。

10

【0047】

ほとんどすべての大腸癌は、エストロゲン受容体(ER)遺伝子が沈黙した細胞から発生する。ER遺伝子転写の不活性化は年齢に関係しており、ER遺伝子の過剰メチル化に関連する(Issa, J.P. 他 (1994) Nat. Genet. 7:536-540)。培養した結腸癌細胞に外因性ER遺伝子を導入することにより、著しく成長が抑制される。結腸上皮細胞におけるERタンパク質の喪失とその結果の癌発症との間のつながりは確立されていない。

【0048】

大腸癌に関連、また、本疾患の発生と進行に、特に遺伝子の下方制御または欠失に関連する、多数の遺伝子変容が存在することは明らかであり、これら変容は癌発生の早期指標を提供する可能性があり、疾患進行のモニターにまたは有望な治療標的の提供にも使用される可能性がある。大腸癌の或る症例に影響した特定の遺伝子群は、この疾患の分子的進行に依存する。大腸癌および前癌状態に関連する更なる遺伝子を同定することにより、本疾患の発症および進行に関連する、より信頼できる診断的パターンが得られるであろう。

20

【0049】

新規の糖結合タンパク質、およびそれらをコードするポリヌクレオチドの発見により、新規の組成物を提供することで当分野の要望に応えることができる。この新規の組成物は、糖質代謝、細胞増殖異常、自己免疫/炎症性の疾患、生殖系疾患、遺伝疾患、輸送疾患、神経系疾患および癌の診断・治療・予防において有用であり、また、糖結合タンパク質の核酸配列及びアミノ酸配列の発現における外来性化合物の影響についての評価にも有用である。

30

【発明の開示】

【発明の効果】

【0050】

本発明は、総称して「CHOP」、個別には「CHOP-1」、「CHOP-2」、「CHOP-3」、「CHOP-4」、「CHOP-5」、「CHOP-6」、「CHOP-7」、「CHOP-8」、「CHOP-9」および「CHOP-10」と呼ばれるような、実質的に精製されたポリペプチドである糖質結合タンパク質を提供する。或る実施態様において本発明は、(a) SEQ ID NO:1-10からなる群から選択したアミノ酸配列からなるポリペプチド、(b) SEQ ID NO:1-10からなる群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を含むポリペプチド、(c) SEQ ID NO:1-10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、または(d) SEQ ID NO:1-10からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を含む群から選択した単離されたポリペプチドを提供する。一実施態様では、SEQ ID NO:1-10のアミノ酸配列を含む単離されたポリペプチドを提供する。

40

【0051】

また、本発明は(a) SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を持つ或る天然アミノ酸配列を有するポリペプチド、(c) SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチドの生物学的活性断片、または(d) SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリヌク

50

レオチドの免疫原性断片、からなる群から選択した或るポリペプチドをコードする、単離されたポリヌクレオチドを提供する。一実施態様では、該ポリヌクレオチドは、SEQ ID NO:1-10からなる群から選択した或るポリペプチドをコードする。別の実施態様では、ポリヌクレオチドはSEQ ID NO:11-20からなる群から選択される。

【0052】

本発明は更に、(a)SEQ ID NO:1-10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:1-10からなる群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を含むポリペプチド、(c)SEQ ID NO:1-10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、または(d)SEQ ID NO:1-10からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片から構成される群から選択されたポリペプチドをコードするポリヌクレオチドと機能的に連結したプロモーター配列を含む組換えポリヌクレオチドを提供する。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生物体を提供する。

10

【0053】

また、本発明は、(a)SEQ ID NO:1-10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:1-10からなる群から選択したアミノ酸配列と少なくとも90%が同一である天然のアミノ酸配列を含むポリペプチド、(c)SEQ ID NO:1-10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:1-10からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片から構成される群から選択したポリペプチドを製造する方法を提供する。製造方法は、(a)組換えポリヌクレオチドを用いて形質転換した細胞をポリペプチドの発現に適した条件下で培養する過程と、(b)そのように発現したポリペプチドを回収する過程とを有し、組換えポリヌクレオチドはポリペプチドをコードするポリヌクレオチドに機能的に連結したプロモーター配列を有する。

20

【0054】

本発明は更に、(a)SEQ ID NO:1-10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:1-10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:1-10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:1-10からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片から構成される群から選択されたポリペプチドに特異結合するような単離された抗体を提供する。

30

【0055】

本発明は更に、(a)SEQ ID NO:11-20からなる群から選択したポリヌクレオチド配列を有するポリヌクレオチド、(b)SEQ ID NO:11-20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。

40

【0056】

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:11-20からなる群から選択したポリヌクレオチド配列を有するポリヌクレオチド、(b)SEQ ID NO:11-20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、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の配列を有する。検出方法は、(a)サンプル中の前記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

50

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。

【0057】

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:11-20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:11-20からなる群から選択したポリヌクレオチド配列と少なくとも90%の同一性を有する天然のポリヌクレオチド配列を含むポリヌクレオチド、(c)(a)のポリヌクレオチドに相補的なポリヌクレオチド、(d)(b)のポリヌクレオチドに相補的なポリヌクレオチド、および(e)(a)~(d)のRNA等価物からなる群から選択された配列のポリヌクレオチドを有する。検出方法は、(a)ポリメラーゼ連鎖反応増幅を用いて標的ポリヌクレオチドまたはその断片を増幅する過程と、(b)前記の増幅した標的ポリヌクレオチドまたはその断片の存在・不存在を検出し、該標的ポリヌクレオチドまたはその断片が存在する場合にはオプションでその量を検出する過程を含む。

10

【0058】

本発明は更に、或る有効量のポリペプチドと薬物として許容し得る或る賦形剤とからなる、或る組成物を提供する。有効量のポリペプチドは、(a)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する天然のアミノ酸配列を含むポリペプチド、(c)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を含むポリペプチドの生物学的活性断片、および(d)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択される。一実施態様では、この組成物はSEQ ID NO:1-10からなる一群から選択されたアミノ酸配列を持つ。更に、本発明は、機能的CHOPの発現の低下に関連した疾患や症状の治療を必要とする患者への、この組成物の投与を含む方法を提供する。

20

【0059】

本発明はまた、(a)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する天然アミノ酸配列を持つポリペプチド、(c)SEQ ID NO:1-10からなる群から選択したアミノ酸配列を持つポリペプチドの生物学的活性断片、および(d)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択したポリペプチドのアゴニストとしての有効性を確認するために、或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)該ポリペプチドを有するサンプルを或る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からなる。別法では、本発明は、この方法によって同定されたアゴニスト化合物と好適な医薬用賦形剤とを含む組成物を提供する。更なる別法では、本発明は、この組成物の患者への投与を含む、機能的CHOPの発現の低下に関連した疾患やその症状の治療方法を提供する。

30

40

【0060】

本発明は更に、(a)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b)SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列との少なくとも90%の同一性を有する天然アミノ酸配列を持つポリペプチド、(c)SEQ ID NO:1-10からなる群から選択したアミノ酸配列を持つポリペプチドの生物学的活性断片、および(d)SEQ ID NO:1-10からなる群から選択したアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択したポリペプチドのアンタゴニストとしての有効性につき、或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)該ポリ

50

ペプチドを含むサンプルを或る化合物に曝す過程と、(b) サンプル中のアンタゴニスト活性を検出する過程からなる。或いは本発明は、この方法で同定したアンタゴニスト化合物と薬物として許容し得る賦形剤とを有する組成物を提供する。更なる別法では、本発明は、機能的CHOPの過剰な発現に関連した疾患やその症状の治療方法を提供し、その内にはそのような治療を必要とする患者にこの組成物を投与することが含まれる。

【0061】

本発明は更に、(a) SEQ ID NO:1-10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:1-10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:1-10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、または(d) SEQ ID NO:1-10からなる群から選択したアミノ酸配列を有するポリペプチドの免疫原性断片を含む群から選択されたポリペプチドに特異結合する化合物をスクリーニングする方法を提供する。スクリーニング方法は、(a) ポリペプチドを適切な条件下で少なくとも1つの試験化合物と混合させる過程と、(b) 試験化合物とのポリペプチドの結合を検出し、それによってポリペプチドに特異結合する化合物を同定する過程とを含む。

10

【0062】

本発明は更に、(a) SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチド、(b) SEQ ID NO:1-10からなる群から選択したアミノ酸配列との少なくとも90%の同一性を有する或る天然アミノ酸配列を含むポリペプチド、(c) SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチドの生物学的活性断片、および(d) SEQ ID NO:1-10からなる群から選択した或るアミノ酸配列を持つポリペプチドの免疫原性断片、からなる群から選択した或るポリペプチドの活性を調節する或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) 該ポリペプチドの活性にとり許容し得る条件下で、該ポリペプチドを少なくとも1つの試験化合物と混合する過程と、(b) 該ポリペプチドの活性をこの試験化合物の存在下で算定する過程と、(c) この試験化合物の存在下での該ポリペプチドの活性をこの試験化合物の不存在下での該ポリペプチドの活性と比較する過程からなり、この試験化合物の存在下での該ポリペプチドの活性の変化は、該ポリペプチドの活性を調節する化合物を標示する。

20

【0063】

更に本発明は、SEQ ID NO:11-20からなる群から選択した或るポリヌクレオチド配列を持つ標的ポリヌクレオチドの発現を改変する効果につき、或る化合物をスクリーニングする方法を提供する。この方法は、(a) この標的ポリヌクレオチドを有するサンプルを或る化合物に曝露する過程と、(b) この標的ポリヌクレオチドの発現の改変を検出する過程と、(c) 可変量のこの化合物の存在下でのこの標的ポリヌクレオチドの発現と、この化合物の不在下での発現とを比較する過程とからなる。

30

【0064】

本発明は更に、試験化合物の毒性の算定方法を提供する。この方法には、以下の過程がある。(a) 核酸群を有する生体サンプルを試験化合物で処理する過程。(b) 処理済み生体サンプルの核酸群をハイブリダイズする過程。この過程には、次のようなプローブを用いる。(i) SEQ ID NO:11-20からなる群から選択した或るポリヌクレオチド配列を含むポリヌクレオチド、(ii) SEQ ID NO:11-20からなる群から選択した或るポリヌクレオチド配列との少なくとも90%の同一性を有する天然ポリヌクレオチド配列を含むポリヌクレオチド、(iii) (i) に相補的な配列を持つポリヌクレオチド、(iv) (ii) のポリヌクレオチドに相補的なポリヌクレオチド、(v) (i) ~ (iv) のRNA等価物、からなる群から選択した或るポリヌクレオチドの少なくとも20の連続したヌクレオチドからなるプローブである。ハイブリダイゼーションは、前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、前記標的ポリヌクレオチドは、(i) SEQ ID NO:11-20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、(ii) SEQ ID NO:11-20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%が同一である天然のポリヌクレオチド配列

40

50

を含むポリヌクレオチド、(iii)(i)のポリヌクレオチドに相補的なポリヌクレオチド、(iv)(ii)のポリヌクレオチドに相補的なポリヌクレオチド、および(v)(I)~(iv)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との差は、試験化合物の毒性を意味する。

【発明を実施するための最良の形態】

【0065】

10

(本発明の記載について)

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。

【0066】

請求の範囲及び明細書中で用いている単数形の「或る」及び「その(この)」の表記は、文脈から明らかにそうでないとされる場合を除いて複数のも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。

20

【0067】

本明細書中で用いる全ての技術用語および科学用語は、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して用い得る、細胞株、プロトコル、試薬およびベクターについて説明およ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。

30

【0068】

(定義)

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

【0069】

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

40

【0070】

「対立遺伝子変異体/変異配列」は、CHOPをコードする、別の形態の遺伝子である。対立遺伝子変異体は、核酸配列における少なくとも1つの突然変異から作製し得る。また、変異RNAまたはポリペプチドからも作製し得る。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。或る遺伝子は、その天然型の対立遺伝子変異配列を全く持たない場合もあり、1個以上持つこともある。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一般に、ヌクレオチドの自然な欠失、付加または置換による。これら各種の変化は、単独であるいは他の変化と共に、或る配列内で1回以上、生じ得る。

【0071】

50

CHOPをコードする「変容した／改変された」核酸配列としては、様々なヌクレオチドの欠失、挿入、或いは置換がありながら、CHOPと同じポリペプチド、或いはCHOPの機能特性の少なくとも1つを備えるポリペプチドをもたらす配列もある。この定義には、CHOPをコードするポリヌクレオチド配列の正常な染色体の遺伝子座ではない位置での対立遺伝子変異配列との不適当或いは予期しないハイブリダイゼーション、並びにCHOP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。コードされるタンパク質も「変容する／改変される」ことがあり、また、サイレント変化を生じた結果、機能的に等価なCHOPとなるような、アミノ酸残基の欠失、挿入または置換を持ち得る。意図的なアミノ酸置換は、生物学的或いは免疫学的にCHOPの活性が保持される範囲で、残基の極性、電荷、溶解度、疎水性、親水性、及び／または両親媒性についての類似性に基づいて成され得る。例えば、負に帯電したアミノ酸にはアスパラギン酸及びグルタミン酸があり、正に帯電したアミノ酸にはリジン及びアルギニンがある。親水性値が近似している非荷電極性側鎖を有するアミノ酸には、アスパラギンとグルタミン、セリンとトレオニンがある。親水性値が近似している非荷電側鎖を有するアミノ酸には、ロイシンとイソロイシンとバリン、グリシンとアラニン、フェニルアラニンとチロシンが含まれうる。

10

【0072】

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

20

【0073】

用語「増幅」は、核酸配列の複製物を作製することに関連する。増幅は通常、当業者によく知られたポリメラーゼ連鎖反応（PCR）技術を用いて行う。

【0074】

「アンタゴニスト」の語は、CHOPの生物学的活性を阻害或いは弱める分子を指す。アンタゴニストは、CHOPに直接相互作用するか、或いはCHOPが関与する生物学的経路の成分と作用して、CHOP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。

30

【0075】

「抗体」の語は、抗原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分子やその断片、例えばFab₁、F(ab')₂ 及びFv断片を指す。CHOP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、好みに応じてキャリアータンパク質に抱合することも可能である。通常用いられるキャリアーであってペプチドと化学結合するものは、ウシ血清アルブミン、サイログロブリン及びスカシガイのヘモシアニン（KLH）等がある。その結合ペプチドは、動物を免疫化するために用いる。

40

【0076】

用語「抗原決定基」は、特定の抗体と接触する分子の領域（即ちエピトープ）を指す。タンパク質またはタンパク質断片を用いて宿主動物を免疫化する場合、タンパク質の多数の領域が、抗原決定基（タンパク質の特定の領域または3次元構造）に特異結合する抗体の産生を誘導し得る。抗原決定基は、抗体に結合するための無損傷抗原（即ち免疫応答を誘導するために用いられる免疫原）と競合し得る。

【0077】

用語「アプタマー（aptamer）」は、核酸またはオリゴヌクレオチド分子であって、特定の分子ターゲットに結合する分子を指す。アプタマーはin vitroでの進化プロセスに由来する（例えば、SELEX（Systematic Evolution of Ligands by EXponential Enrichmentの

50

略、試験管内選択法)、米国特許第5,270,163号に記述)。これは、大規模な組み合わせライブラリ群から標的特異的アプタマー配列を選択するプロセスである。アプタマー組成は、2本鎖または1本鎖であってもよく、デオキシリボヌクレオチド、リボヌクレオチド、ヌクレオチド誘導体、または他のヌクレオチド様分子を含み得る。アプタマーの各ヌクレオチド成分は修飾された糖基を持つことがあり(例えば、リボヌクレオチドの2'-OH基は2'-Fまたは2'-NH₂によって置換され得る)、これは、ヌクレアーゼへの抵抗性または血中でのより長い寿命など、所望の特性を向上し得る。循環系からアプタマーが除去される速度を遅くするために、アプタマーを高分子量キャリア等の分子に抱合させることができる。アプタマーは、たとえば架橋剤の光活性化によって各々のリガンドと特異的に架橋させることができる(Brody, E.N. および Gold, L. (2000) J. Biotechnol. 74:5-13)

10

【0078】

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

【0079】

「spiegelmer」の語はL-DNA、L-RNAその他の左旋性ヌクレオチド誘導体またはヌクレオチド様分子を含むアプタマーを指す。左旋性ヌクレオチドを含むアプタマーは、右旋性のヌクレオチドを含む基質に通常作用する天然の酵素による分解に耐性がある。

20

【0080】

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

30

【0081】

用語「生物学的に活性」は、天然分子の構造的、調節的、或いは生化学的な機能を有するタンパク質を指す。同様に、「免疫学的に活性」または「免疫原性」は、天然、組換え体、または合成のCHOP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。

【0082】

「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、「5'-AGT-3'」は、その相補配列「3'-TCA-5'」との対を形成する。

40

【0083】

「所定のポリヌクレオチド配列を含む組成物」または「所定のアミノ酸配列を含む組成物」は広い意味で、所定のポリヌクレオチド配列若しくはアミノ酸配列を含む任意の組成物を指す。この組成物には、乾燥製剤または水溶液が含まれ得る。CHOP若しくはCHOPの断片をコードするポリヌクレオチド配列を含む組成物は、ハイブリダイゼーションプローブとして使用され得る。このプローブは、凍結乾燥状態で保存可能であり、糖質などの安定化剤と結合させることが可能である。ハイブリダイゼーションにおいては、塩(例えばNaCl)、界面活性剤(例えばドデシル硫酸ナトリウム; SDS)及びその他の構成成分(例え

50

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。

【0084】

「コンセンサス配列」は、不要な塩基を分離するためにDNA配列の解析を繰り返し行い、X L-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オーバーラップするcDNAやEST、またはゲノムDNA断片から構築された核酸配列を指す。伸長及びアセンブリの両方を行ってコンセンサス配列を作製する配列もある。

10

【0085】

用語「保存的なアミノ酸置換」は、元のタンパク質の特性を殆ど変えない置換を指す。即ち、置換によってそのタンパク質の構造や機能が大きくは変わらず、そのタンパク質の構造、特にその機能が保存される。下表は、タンパク質中で元のアミノ酸と置換できて、保存アミノ酸置換と認められるアミノ酸を示している。

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

20

30

【0086】

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βシートやα螺旋構造、(b)置換部位における分子の電荷または疎水性、及び/または(c)側鎖の大部分を保持する。

40

【0087】

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

【0088】

用語「誘導体」は、化学修飾されたポリヌクレオチドまたはポリペプチドを指す。例えば、アルキル基、アシル基、ヒドロキシル基またはアミノ基による水素の置換は、ポリヌクレオチド配列の化学修飾に含まれ得る。ポリヌクレオチド誘導体は、天然分子の生物学的または免疫学的機能を少なくとも1つは保持しているポリペプチドをコードする。ポリペプチド誘導体は、グリコシル化、ポリエチレングリコール化(pegylation)、或いは任意

50

の同様なプロセスであって、誘導起源のポリペプチドの少なくとも1つの生物学的若しくは免疫学的機能を保持するプロセスによって、修飾されたポリペプチドである。

【0089】

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

【0090】

「差次的発現」は少なくとも2つの異なったサンプルを比較することによって決められる、増加（上方調節）、あるいは減少（下方調節）、または欠損した、遺伝子またはタンパク質の発現を指す。このような比較は例えば、処理済サンプルと不処理サンプル、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。

10

【0091】

「エキソンシャフリング」は、異なるコード領域（エキソン）の組換えを意味する。1つのエキソンがコードされたタンパク質の1つの構造的または機能的ドメインを代表し得るため、安定したサブストラクチャーを再分類することによって、新しいタンパク質が組立てられることが可能であり、新しいタンパク質機能の進化を促進できる。

【0092】

「断片」は、CHOPの又はCHOPをコードするポリヌクレオチドの固有の部分であって、その親配列（parent sequence）と配列は同一であるが親配列より長さが短いものを指す。断片は、定義された配列の全長から1ヌクレオチド/アミノ酸残基を差し引いた長さよりも短い長さを有し得る。例えば或る断片は、5～1000の連続したヌクレオチドまたはアミノ酸残基を有し得る。プローブ、プライマー、抗原、治療用分子として、或いはその他の目的のために用いられる断片は、少なくとも5、10、15、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選択的に選択し得る。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。

20

【0093】

SEQ ID NO:11-20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:11-20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:11-20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:11-20に関連ポリヌクレオチド配列から区別する類似の方法に有用である。SEQ ID NO:11-20の断片の正確な長さ及び断片に対応するSEQ ID NO:11-20の領域は、断片に対する意図した目的に基づき当業者が慣例的に決定することが可能である。

30

【0094】

SEQ ID NO:1-10のある断片は、SEQ ID NO:11-20のある断片によってコードされる。SEQ ID NO:1-10のある断片には、SEQ ID NO:1-10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:1-10のある断片は、SEQ ID NO:1-10を特異認識する抗体を産出するための免疫原性ペプチドとして有用である。SEQ ID NO:1-10のある断片の正確な長さ、及びその断片に対応するSEQ ID NO:1-10の領域は、断片に対する意図した目的に基づき当業者が慣例的に決定することが可能である。

40

【0095】

「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。

【0096】

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

【0097】

50

ポリヌクレオチド配列についての用語「一致率」または「一致性%」とは、標準化されたアルゴリズムを用いてアラインメントされる、2つ以上のポリヌクレオチド配列間の一致する残基の百分率のことである。このようなアルゴリズムは、2配列間のアラインメントを最適化するために、比較する配列において、標準化された再現性のある方法でギャップを挿入しうるので、2つの配列をより有意に比較できる。

【0098】

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み込まれているようなCLUSTAL Vアルゴリズムのデフォルトのパラメータを用いて決定できる。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およびSharp (1989) (CABIOS 5:151-153)、Higgins他 (1992) (CABIOS 8:189-191)に記載されている。ポリヌクレオチド配列を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。「weighted(重み付けされた)」残基重み付け表が、デフォルトとして選択される。一致率は、アラインメントされたポリヌクレオチド配列間の「percent similarity(類似性パーセント)」としてCLUSTAL Vによって報告される。

10

【0099】

或いは、一般的に用いられ且つ自由に入手できる配列比較アルゴリズム一式が、国立バイオテクノロジー情報センター(NCBI) Basic Local Alignment Search Tool (BLAST) から提供されており(Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410)、これはメリーランド州ベセスダにあるNCBI及びインターネット(<http://www.ncbi.nlm.nih.gov/BLAST/>)を含む幾つかの情報源から入手可能である。このBLASTソフトウェア一式には、既知のポリヌクレオチド配列と様々なデータベースの別のポリヌクレオチド配列とのアラインメントに用いられる「blastn」を含む、様々な配列分析プログラムが含まれる。「BLAST 2 Sequences」と呼ばれるツールが入手可能であり、2つのヌクレオチド配列の直接のペアワイズで比較するために用いられる。「BLAST 2 Sequences」は、<http://www.ncbi.nlm.nih.gov/gorf/bl2.html>にアクセスして、対話形式で利用ができる。「BLAST 2 Sequences」ツールは、blastn および blastp(以下に記載)の両方に用い得る。BLASTプログラムは、一般的には、ギャップ(gap)などのパラメータをデフォルト設定にセットして用いる。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。

20

30

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 及び Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

40

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

【0100】

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

50

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。

【0101】

ポリペプチド配列に用いられる用語「一致率」または「一致性%」とは、標準化されたアルゴリズムを用いてアラインメントされる2つ以上のポリペプチド配列間の一致する残基の百分率のことである。ポリペプチド配列アラインメントの方法は公知である。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、置換部位の電荷および疎水性を保存するので、ポリペプチドの構造を(したがって機能も)保存する。

【0102】

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

【0103】

或いは、NCBI BLASTソフトウェア式を用いてもよい。例えば、2つのポリペプチド配列をペアワイズで比較する場合、「BLAST 2 Sequences」ツールVersion 2.0.12(2000年4月21日)のblastpをデフォルトパラメータに設定して用い得る。デフォルトパラメータの設定例を以下に示す。

Matrix: BLOSUM62

Open Gap: 11 及び Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

一致率は、定義された(例えば特定の配列番号で定義された)ポリペプチド配列全体の長さと比較して測定し得る。或いは、より短い長さ、例えばより大きな定義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。

【0104】

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

【0105】

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

【0106】

「ハイブリダイゼーション」とは、所定のハイブリダイゼーション条件下で、ある一本鎖ポリヌクレオチドがある相補的な一本鎖と塩基対を形成するアニーリングのプロセスである。特異的ハイブリダイゼーションは、2つの核酸配列が高い相補性を共有することを示すものである。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

10

20

30

40

50

合)が減少する。核酸配列のアニール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。アニール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。

【0107】

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の条件下で、完全に一致するプローブに標的配列の50%がハイブリダイズする温度である。T_mを計算する式及び核酸のハイブリダイゼーション条件はよく知られており、Sambrook 他、1989(Molecular Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NYに記載されており、特に2巻の9章)を参照されたい。

10

【0108】

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0.1%のSDSの存在の下、約68℃で1時間の洗浄過程を含む。別法では、約65℃、60℃、55℃、または42℃の温度で行う。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のせん断した変性サケ精子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、ヌクレオチド群及びヌクレオチドにコードされるポリペプチド群について、類似の役割を強く示唆している。

20

【0109】

用語「ハイブリダイゼーション複合体」は、相補的な塩基間の水素結合によって、形成された2つの核酸配列の複合体を指す。ハイブリダイゼーション複合体は、溶解状態で形成し得る(C₀tまたはR₀t解析等)。或いは、一方の核酸配列が溶解状態で存在し、もう一方の核酸配列が固体支持体(例えば紙、膜、フィルタ、チップ、ピンまたはガラススライド、或いは他の適切な基板であって細胞若しくはその核酸が固定される基板)に固定されているような2つの核酸配列間に形成され得る。

30

【0110】

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

【0111】

「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。

40

【0112】

「免疫原性断片」は、生物(例えば哺乳類)に導入すると免疫応答を引き起こし得る、CHOP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、CHOPの任意のポリペプチド断片またはオリゴペプチド断片をも指す。

【0113】

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

50

【0114】

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

【0115】

用語「モジュレート」または「活性を調節」は、CHOPの活性を変化させることを指す。例えば、調節によって、CHOP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。

【0116】

「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。

10

【0117】

「機能的に連結した」は、第1の核酸配列と第2の核酸配列が機能的な関係にある状態を指す。例えば、プロモーターがコード配列の転写または発現に影響を及ぼす場合には、そのプロモーターはそのコード配列に機能的に連結している。同一のリーディングフレーム内で2つのタンパク質コード領域を接続する必要がある場合、一般に、機能的に連結したDNA配列は非常に近接するか、或いは連続的に隣接し得る。

【0118】

「ペプチド核酸(PNA)」は、末端がリシンで終わるアミノ酸残基のペプチドのバックボーンに結合した、少なくとも約5ヌクレオチドの長さのオリゴヌクレオチドを含む、アンチセンス分子または抗遺伝子剤を指す。末端のリジンは、この組成に溶解性を与える。PNAは、相補的一本鎖DNAまたはRNAに優先的に結合して転写の伸長を停止するものであり、ポリエチレングリコール化して細胞におけるPNAの寿命を延長し得る。

20

【0119】

CHOPの「翻訳後修飾」としては、脂質化、グリコシル化、リン酸化、アセチル化、ラセミ化、タンパク質分解切断、及びその他の当分野で既知の修飾を含み得る。これらのプロセスは、合成的或いは生化学的に生じ得る。生化学的修飾は、CHOPの酵素環境に依存し、細胞の種類によって異なり得る。

30

【0120】

「プローブ」とは、核酸配列の内、CHOPやそれらの相補配列、またはそれらの断片をコードし、同一配列や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る配列を指す。プローブは、単離されたオリゴヌクレオチドまたはポリヌクレオチドであって、検出可能な標識またはレポーター分子に結合したものである。典型的な標識には、放射性アイソトープ、リガンド、化学発光試薬及び酵素がある。「プライマー」とは、相補的な塩基対を形成して標的のポリヌクレオチドにアニーリング可能な、通常はDNAオリゴヌクレオチドである短い核酸である。プライマーは次に、DNAポリメラーゼ酵素によって標的DNA鎖に沿って延長し得る。プライマー対は、例えばポリメラーゼ連鎖反応(PCR)による核酸配列の増幅(及び同定)に用い得る。

40

【0121】

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。

【0122】

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他、1989 (Molecular

50

Cloning: A Laboratory Manual, 第2版, 1-3巻, Cold Spring Harbor Press, Plainview NY)、Ausubel 他, 1987(Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY)、Innis他、1990(PCR Protocols, A Guide to Methods and Applications Academic Press, San Diego CA等)を参照されたい。PCRプライマー対を既知の配列から得るには、例えば、そ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)を用い得る。

【0123】

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既知の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キロベ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(Whitehead Institute/MIT Center for Genome Research(マサチューセッツ州ケンブリッジ)より入手可能)を用いれば、プライマー結合部位として避けたい配列を指定できる「非プライミングライブラリ(mispriming library)」を入力できる。Primer3は特に、マイクロアレイのためのオリゴヌクレオチドの選択に有用である。(後二者のプライマー選択プログラムのソースコードは、それぞれの情報源から得てユーザー固有のニーズを満たすように変更してもよい。)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、固有であって保存されたオリゴヌクレオチド及びポリヌクレオチドの断片の同定に有用である。上記選択方法のいずれかによって同定したオリゴヌクレオチド及びポリヌクレオチドの断片は、ハイブリダイゼーション技術において、例えばPCRまたはシーケンシングプライマーとして、マイクロアレイエレメントとして、或いは核酸のサンプルにおいて完全または部分的相補的ポリヌクレオチドを同定する特異プローブとして有用である。オリゴヌクレオチドの選択方法は、上記の方法に限定されるものではない。

【0124】

本明細書における「組換え核酸」は天然の配列ではなく、2つ以上の配列の離れたセグメントを人工的に組み合わせた配列である。この人為的組合せはしばしば化学合成によって達成するが、より一般的には核酸の単離セグメントの人為的操作によって、例えばの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、単に核酸の一部が付加、置換または欠失により改変された核酸も含む。しばしば組換え核酸には、プロモーター配列に機能的に連結した核酸配列が含まれる。このような組換え核酸は、例えばある細胞を形質転換するために使用されるベクターの一部とすることが可能である。

【0125】

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

【0126】

「調節因子」は、通常は遺伝子の非翻訳領域に由来する核酸配列であり、エンハンサー、プロモーター、イントロン及び5'及び3'の非翻訳領域(UTR)を含む。調節エレメント

10

20

30

40

50

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

【0127】

「レポーター分子」は、核酸、アミノ酸または抗体の標識に用いられる化学的または生化学的な部分である。レポーター分子には、放射性核種、酵素、蛍光剤、化学発光剤、発色剤、基質、補助因子、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。

【0128】

DNA配列に対する「RNA等価物」とは、基準となるDNA配列と同じ直鎖の核酸配列から構成されるが、窒素性塩基のチミンがウラシルで置換され、糖鎖のバックボーンがデオキシリボースではなくリボースからなる。

10

【0129】

用語「サンプル」は、その最も広い意味で用いられている。CHOP、CHOP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。

【0130】

用語「特異的結合」及び「特異的に結合する」は、タンパク質若しくはペプチドと、アゴニスト、抗体、アンタゴニスト、小分子、若しくは任意の天然若しくは合成の結合組成物との間の相互作用を指す。この相互作用は、タンパク質の特定の構造（例えば抗原決定基即ちエピトープ）であって結合分子が認識するものが存在するか否かに依存していることを意味している。例えば、抗体がエピトープ「A」に対して特異的である場合、結合していない標識した「A」及び抗体を含む反応液に、エピトープAを含むポリペプチド或いは結合していない無標識の「A」が存在すると、抗体と結合する標識Aの量が減少する。

20

【0131】

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

【0132】

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

30

【0133】

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

【0134】

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

40

【0135】

「形質転換(transformation)」とは、外来DNAが受容細胞に導入されるプロセスのことである。形質転換は、本技術分野で知られている種々の方法に従って自然条件または人工条件下で生じ得るものであり、外来性の核酸配列を原核宿主細胞または真核宿主細胞に挿入する任意の既知の方法を基にし得る。形質転換の方法は、形質転換する宿主細胞の種類によって選択する。限定するものではないが形質転換方法には、ウイルス感染、電気穿孔法(エレクトロポレーション)、熱ショック、リポフェクション及び微粒子銃を用いる方法がある。「形質転換された細胞」には、導入されたDNAが自律的に複製するプラスミドとして或いは宿主染色体の一部として複製可能である安定的に形質転換された細胞が含まれる。さらに、限られた時間に一過的に導入DNA若しくは導入RNAを発現する細胞も含まれる

50

。

【0136】

ここで用いる「遺伝形質転換体」とは任意の生物体であり、限定するものではないが動植物を含み、生物体の1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。或る実施例において、核酸はレンチウイルスベクターのような組換えウイルスベクターを用いた感染によって導入することができる(Lois, C. 他(2002) Science 295:868-872)。遺伝子操作の語は、古典的な交雑育種或いは試験管内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(transconjugation)によって宿主に導入することができる。本発明のDNAをこのような生物体に移入する技術はよく知られており、前出のSambrook他(1989)等の参考文献に与えられている。

10

【0137】

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「多型性」変異体として説明し得る。スプライス変異体は参照分子とかなりの相同性を有し得るが、mRNAプロセッシング中のエキソンの選択的スプライシングによって通常、より多くまたはより少数のヌクレオチドを有することになる。対応するポリペプチドは、追加機能ドメインを有するか或いは参照分子に存在するドメインが欠落していることがある。種変異体は、種によって異なるポリヌクレオチド配列である。結果的に生じるポリペプチドは通常、相互にかなりのアミノ酸相同性を有する。多型性変異体は、与えられた種の個体間で特定の遺伝子のポリヌクレオチド配列中での変異である。多型変異配列はまた、ポリヌクレオチド配列の1つのヌクレオチドが異なる「1塩基多型性」(SNP)も含み得る。SNPの存在は、例えば特定の個体群、病状または病状性向を示し得る。

20

30

【0138】

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定義される。定義づけには、デフォルトパラメータに設定した「BLAST 2 Sequences」ツールVersion 2.0.9(1999年5月7日)を用いてblastpを実行する。このようなポリペプチド対は、一方のポリペプチドの或る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。

40

【0139】

(発明)

本発明は、新規の糖質結合タンパク質(CHOP)の発見、CHOPをコードするポリヌクレオチド、および、糖質代謝、細胞増殖異常、自己免疫/炎症性の疾患、生殖系疾患、遺伝疾患、輸送疾患、神経系疾患および癌の診断、治療、ならびに予防に対するこれらの組成の使用に基づいている。

【0140】

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

50

番号 (IncyteプロジェクトID) と相関する。各ポリペプチド配列は、ポリペプチド配列識別番号 (ポリペプチドSEQ ID NO) とIncyteポリペプチド配列番号 (IncyteポリペプチドID) によって表示した。各ポリヌクレオチド配列は、ポリヌクレオチド配列識別番号 (ポリヌクレオチドSEQ ID NO) とIncyteポリヌクレオチドコンセンサス配列番号 (IncyteポリヌクレオチドID) によって表示した。

【0141】

表2は、GenBankタンパク質 (genpept) データベースとPROTEOMEデータベースとに対するBLAST分析で同定した、本発明のポリペプチド群に相同な配列群を示す。列1および列2はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号 (ポリペプチドSEQ ID NO:) とそれに対応するIncyteポリペプチド配列番号 (IncyteポリペプチドID) を示す。列3は、最も近いGenBank相同体のGenBank識別番号 (GenBank ID NO:) を示す。列4は、各ポリペプチドとその相同体との間の一致について確率スコアを示す。列5は、GenBankとPROTEOMEデータベースの相同体の注釈を示し、更に該当箇所には関連する引用文献も示す。これら全ては引用することを以て本明細書の一部とする。

10

【0142】

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

20

【0143】

表2及び表3は共に、本発明の各々のポリペプチドの特性を要約しており、それら特性が請求の範囲に記載されたポリペプチドが糖結合タンパク質であることを確立している。例えばSEQ ID NO:1は、残基Q170から残基W909までが、マウスの α -グルコシダーゼIIの α サブユニット (GenBank ID g2104689) に対して54%同一であると、Basic Local Alignment Search Tool (BLAST) で判定された (表2参照)。BLAST確率スコアは $1.4e-258$ であり、これは観察されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:1はまた、1つのグリコシルヒドロラーゼファミリー1ドメインを有するが、これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された (表3参照)。BLIMPS及びMOTIFS解析よりのデータは、SEQ ID NO:1がグリコシルヒドロラーゼであることをさらに確証する証拠を提供する。別の例において、SEQ ID NO:4はウシの α 1,3-ガラクトシルトランスフェラーゼ (GenBank ID g163124) と残基E20からT276まで55%の同一性を有することがBasic Local Alignment Search Tool (BLAST) によって決定された (表2参照)。BLAST確率スコアは $5.8e-77$ であり、これは観察されたポリペプチド配列アラインメントが偶然に得られる確率を示している。BLAST解析よりのデータは、SEQ ID NO:4が α 1,3-ガラクトシルトランスフェラーゼである、さらに実証的な証拠を提供する。また他の例として、SEQ ID NO:7は、残基M2から残基F402まで、マウスの β 1,6-N-アセチルグルコサミニルトランスフェラーゼB (GenBank ID g9650954) と77%同一であることがBasic Local Alignment Search Tool (BLAST) によって示された (表2参照)。BLAST確率スコアは $2.1e-177$ であり、これは観察されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:7はまた、1つのコア2/1分岐酵素 (core-2/1-branching enzyme) ドメインを有するが、これは、隠れマルコフモデル (HMM) を基にした保存されたタンパク質ファミリードメイン群のPFAMデータベースにおいて、統計的に有意な一致を検索して決定された (表3参照)。更なるBLAST解析から得られたデータは、SEQ ID NO:7が β 1,6-N-アセチルグルコサミニルトランスフェラーゼBであることを裏づける証拠を提供

30

40

50

する。また他の例として、SEQ ID NO:9はヒトのインテレクチン (GenBank ID g8096221) と残基M8 から残基Q200まで81%の同一性を有することがBasic Local Alignment Search Tool (BLAST)によって示された(表2参照)。BLAST確率スコアは 2.1×10^{-83} であり、これは観察されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:9はインテレクチンであることが、PROTEOMEデータベースを使ったBLAST解析によって示された。SEQ ID NO:9はまた、フィブリノーゲンドメインを有するが、これは、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリドメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された(表3参照)。PRODOMデータベースのBLAST解析からのデータは、SEQ ID NO:9がレクチンである、さらに実証的な証拠を提供する。また他の例として、SEQ ID NO:10はヒトのカーゴ選択タンパク質(cargo selection protein)(マンノース-6-リン酸受容体結合タンパク質)(GenBank ID g14043157)と残基T35から残基E387まで36%同一であることがBasic Local Alignment Search Tool (BLAST)によって示された(表2参照)。BLAST確率スコアは 1.1×10^{-51} であり、これは観察されたポリペプチド配列アラインメントが偶然に得られる確率を示している。SEQ ID NO:10はまた、エンドソーム小胞に局在するタンパク質(マンノース-6-リン酸受容体のエンドソームからゴルジ装置への輸送を仲介する)であり、また47 kDaのエンドソーム/エンドソーム小胞細胞質尾部相互作用タンパク質およびペリリピン(perilipin)タンパク質であるタンパク質に相同性を持つことがPROTEOMEデータベースを使ったBLAST解析によって示された。SEQ ID NO:10はまた、ペリリピンドメインを有するが、これは、隠れマルコフモデル(HMM)を基にした保存されたタンパク質ファミリドメインのPFAMデータベースにおいて、統計的に有意な一致を検索して決定された(表3参照)。BLAST_PRODOMデータベースを用いるBLAST解析よりのデータは、SEQ ID NO:10がペリリピンの相同体あることをさらに確証する証拠を提供する。SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:5、SEQ ID NO:6およびSEQ ID NO:8については、同様の方法で分析し、注釈を付けた。SEQ ID NO:1-10の解析のためのアルゴリズム及びパラメータが表7で記述されている。

10

20

30

40

50

【0144】

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アセンブリした。列1は本発明の各ポリヌクレオチドに対するポリヌクレオチド配列識別番号(ポリヌクレオチドSEQ ID NO)および対応するIncyteポリヌクレオチドコンセンサス配列番号(Incyte ID)、および塩基対の各ポリヌクレオチド配列の長さを示している。列2は、本発明の全長ポリヌクレオチド配列を構築するために使われたcDNA配列および/またはゲノム配列のヌクレオチド開始位置(5')と終了位置(3')を示し、またSEQ ID NO:11-20を同定するための、またはSEQ ID NO:11-20と関連するポリヌクレオチド配列とを区別するための技術(例えば、ハイブリダイゼーション技術または増幅技術)で有用なポリヌクレオチド配列の断片のヌクレオチド開始位置(5')と終了位置(3')を示す。

【0145】

表4の列2で記述されたポリヌクレオチド断片は、具体的にはたとえば組織特異的なcDNAライブラリまたはプールされたcDNAライブラリに由来するIncyte cDNAを指す場合がある。或いは列2の識別番号は、完全長ポリヌクレオチド配列のアセンブリに寄与するGenBank cDNAまたはESTを指す場合もある。さらに、列2のポリヌクレオチド断片は、ENSEMBL(The Sanger Centre、英国ケンブリッジ)データベースから由来した配列(即ち「ENST」命名を含む配列)を同定し得る。或いは、列2で記述されたポリヌクレオチド断片は、NCBI RefSeq Nucleotide Sequence Records データベースから由来する場合もあり(即ち「NM」または「NT」の命名を含む配列)、またNCBI RefSeq Protein Sequence Recordsから由来する場合もある(即ち「NP」の命名を含む配列)。または列で記述されたポリヌクレオチド断片は、「エキソンスティッチング(exon-stitching)」アルゴリズムにより結び合わせたcDNA及びGenscan予想エキソンの両方からなるアセンブリ体を意味する場合がある。例えば、FL_XXXXXX_N1_N2_YYYYY_N3_N4 として同定されるポリヌクレオチド配列はア

ルゴリズムが適用される配列のクラスターの識別番号がXXXXXXであり、アルゴリズムにより生成される予測の番号がYYYYYであり、（もし存在すれば） $N_{1,2,3...}$ が解析中に手動で編集された可能性のある特定のエキソンであるような「縫合された」配列である（実施例5参照）。または、列2のポリヌクレオチド断片は「エキソンストレッチング（exon-stretching）」アルゴリズムにより結び合わせたエキソンの集合を指す場合もある。例えば、FLXXXXXX_gAAAAA_gBBBBB_1_Nとして同定されるポリヌクレオチド配列は「ストレッチ」配列である。ここでXXXXXXはIncyteプロジェクト識別番号、gAAAAAは「エキソンストレッチング」アルゴリズムを適用したヒトゲノム配列のGenBank識別番号、gBBBBBは一番近いGenBankタンパク質相同体のGenBank識別番号またはNCBI RefSeq 識別番号である。（実施例5を参照。）あるRefSeq配列が「エキソンストレッチング」アルゴリズムのためのタン

10

【0146】

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

接頭コード	解析タイプやプログラムの例
GNN 、 GFG 、 ENST	例えば GENSCAN (Stanford University, CA, USA) または FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK)を用いた、ゲノム配列からのエキソン予測
GBI	手動で編集されたゲノム配列の解析
FL	スティッチまたはストレッチゲノム配列（ <u>実施例5</u> 参照）
INCY	ゲノムへの EST 配列群のマッピングからの、完全長転写物とエキソンとの予測。エキソンと生じる転写物を予測するために、ゲノム位置と EST 構成とのデータが組み合わされる。

20

30

【0147】

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

40

【0148】

表5は、Incyte cDNA配列を用いてアセンブリ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およびベクターを表5に示し、表6で説明している。

【0149】

本発明はまた、CHOPの変異体も含む。好適なCHOPの変異体は、CHOPアミノ酸配列に対して、少なくとも約80%、あるいは少なくとも約90%、さらには少なくとも約95%のアミノ酸

50

配列同一性を有し、CHOPの機能的または構造的特徴を少なくとも1つ含む変異体である。

【0150】

本発明はまた、CHOPをコードするポリヌクレオチドを提供する。特定の実施態様で本発明は、CHOPをコードする、SEQ ID NO:11-20 からなる群から選択した1配列を持つポリヌクレオチド配列を提供する。SEQ ID NO:11-20のポリヌクレオチド配列には、配列表に示したように等価RNA配列をも含むが、そこでは窒素塩基チミンの出現はウラシルに置換され、糖のバックボーンはデオキシリボースではなくリボースで構成される。

【0151】

本発明は、また、CHOPをコードするポリヌクレオチド配列の変異配列をも含む。詳細には、このようなポリヌクレオチド配列の変異配列は、CHOPをコードするポリヌクレオチド配列と少なくとも70%のポリヌクレオチド配列同一性、或いは少なくとも85%のポリヌクレオチド配列同一性、更には少なくとも95%ものポリヌクレオチド配列同一性を有する。本発明の特定の実施形態では、SEQ ID NO:11-20からなる一群から選択された1核酸配列との少なくとも約70%のポリヌクレオチド配列同一性、或いは少なくとも約85%のポリヌクレオチド配列同一性、更には少なくとも約95%ものポリヌクレオチド配列同一性を有する、SEQ ID NO:11-20からなる一群から選択された1配列を持つポリヌクレオチド配列の変異配列を提供する。上記したポリヌクレオチド変異配列は何れも、CHOPの機能的或いは構造的特徴の少なくとも1つを有するアミノ酸配列をコードし得る。

【0152】

さらに、あるいは別法において、本発明のポリヌクレオチド変異体はCHOPをコードするポリヌクレオチド配列のスプライス変異体である。或るスプライス変異配列はCHOPをコードするポリヌクレオチド配列との顕著な配列同一性を持つ部分複数を有し得るが、mRNAプロセッシング中のエキソン群の選択的スプライシングによって生ずる、配列の数ブロックの付加または欠失により、通常、より多数またはより少数のポリヌクレオチドを有することになる。或るスプライス変異配列には、約70%未満、または約60%未満、あるいは約50%未満のポリヌクレオチド配列同一性が、CHOPをコードするポリヌクレオチド配列との間で全長に渡って見られるが、このスプライス変異配列のいくつかの部分には、CHOPをコードするポリヌクレオチド配列の各部との、少なくとも約70%、あるいは少なくとも約85%、または少なくとも約95%、なおまたは100%の、ポリヌクレオチド配列同一性を有することとなる。上記したスプライス変異配列は何れも、CHOPの機能的或いは構造的特徴の少なくとも1つを有するアミノ酸配列をコードする。

【0153】

遺伝暗号の縮重により、CHOPをコードする種々のポリヌクレオチド配列が作り出され、中には、既知のいかなる天然遺伝子のポリヌクレオチド配列群とも最小の類似性しか有しない配列もあることは、当業者には理解されよう。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あらゆる可能なポリヌクレオチド配列のバリエーションを網羅し得る。これらの組み合わせは、天然のPP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みなす。

【0154】

CHOP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CHOP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CHOP或いはその誘導体をコードするヌクレオチド配列を作るとは有利となり得る。宿主が特定のコドンを利用する頻度に基づいて、特定の真核宿主又は原核宿主に発生するペプチドの発現率を高めるようにコドンを選択することが可能である。コードされたアミノ酸配列を変えないで、CHOP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。

【0155】

本発明はまた、CHOP及びその誘導体をコードするDNA配列またはそれらの断片を完全に合成化学によって作り出すことも含む。作製後、当分野で公知の試薬類を用いて、この合成配列を任意の多くの入手可能な発現ベクター及び細胞系中に挿入し得る。更に、合成化学を用いて、CHOPまたはその任意の断片をコードする配列に突然変異を誘発し得る。

【0156】

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:11-20及びそれらの断片とハイブリダイズ可能なポリヌクレオチド配列が含まれる(例えば、Wahl, G.M.及びS.L. Berger (1987) *Methods Enzymol.* 152:399-407; and Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.を参照)。アニーリング及び洗浄条件を含むハイブリダイゼーションの条件は、「定義」に記載されている。 10

【0157】

DNAシーケンシングの方法は当分野では公知であり、本発明のいずれの実施例もDNAシーケンシング方法を用いて実施可能である。DNAシーケンシング方法には酵素を用いることができ、例えばDNAポリメラーゼIのクレノウ断片、SEQUENASE (US Biochemical, Cleveland OH)、Taqポリメラーゼ (Applied Biosystems)、熱安定性T7ポリメラーゼ (Amersham, Biosciences, 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 (Invitrogen, Carlsbad CA)において見られるように、ポリメラーゼと校正エキソヌクレアーゼを併用することができる。好適には、MICROLAB2200液体転移システム (Hamilton, Reno, NV)、PTC200サーマルサイクラー (MJResearch, 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準備を自動化する。次に、ABI 373 或いは 377 DNAシーケンシングシステム (Applied Biosystems)、MEGABACE 1000 DNAシーケンシングシステム (Amersham Biosciences) 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, 7.7ユニット、Meyers, R.A. (1995) *Molecular Biology and Biotechnology*, Wiley VCH, New York NY, 856-853ページ)。 20

【0158】

当分野で周知の、PCR法をベースにした種々の方法と、部分的ヌクレオチド配列とを利用して、CHOP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、上流にある配列を検出し得る。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* 2:318-322)。別の方法にインバース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からなる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増幅する方法に関与している (Lagerstrom, M.他 (1991) *PCR Methods Applic* 1:111-119)。この方法では、PCRを行う前に複数の制限酵素の消化及びライゲーション反応を用いて未知の配列領域内に組換え二本鎖配列を挿入することが可能である。また、未知の配列を検索するために用い得る別の方法については当分野で知られている (Parker, J.D.他 (1991) *Nucleic Acids Res.* 19:3055-3060)。更に、PCR、ネステッドプライマー及びPromoterFinder (商標)ライブラリ (Clontech, Palo Alto CA)を用いてゲノムDNAをウォーキングすることができる。この手順は、ライブラリ類をスクリーニングする必要がなく、イントロン/エキソン接合部の発見に有用である。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 30 40 50

る。

【0159】

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

【0160】

市販の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フルオロフォアと、発光された波長の検出に利用するCCDカメラとを有し得る。出力ノ光の強度は、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。

10

【0161】

本発明の別の実施例では、CHOP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CHOP、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CHOPのクローン化及び発現に利用可能である。

20

【0162】

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

30

【0163】

本発明のヌクレオチドを、MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; 米国特許第5,837,458号; Chang, C.-C. 他 (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. 他 (1999) Nat. Biotechnol. 17:259-264; Cramer, A. 他 (1996) Nat. Biotechnol. 14:315-319)などのDNAシャッフリング技術を用いてシャッフリングして、CHOPの生物学的または酵素的な活性、或いは他の分子や化合物と結合する能力などのCHOP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を介する組換えを用いて遺伝子変異体のライブラリを作製するプロセスである。ライブラリはその後、所望の特性を持つ遺伝子変異体群を同定する、選択またはスクリーニングの手順を経る。続いて、これら好適な変異体をプールし、更に反復してDNAシャッフリングおよび選択/スクリーニングを行い得る。従って、人工的な育種及び急速な分子の進化によって多様な遺伝子が作られる。例えば、ランダムポイント突然変異を持つ単一の遺伝子の断片を組み換えて、スクリーニングし、その後所望の特性が最適化されるまでシャッフリングし得る。或いは、所定の遺伝子を同種または異種のいずれかから得た同一遺伝子ファミリーの相同遺伝子と組み換え、それによって天然に存在する複数の遺伝子の遺伝多様性を、指図された制御可能な方法で最大化させることができる。

40

【0164】

別の実施例によれば、CHOP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

50

体或いは一部が合成可能である (Caruthers, M.H. 他 (1980) Nucl. Acids Res. Symp. Ser. 7:215-223; 及び Horn, T. 他 (1980) Nucl. Acids Res. Symp. Ser. 225-232)。別法として、化学的方法を用いてCHOP自体またはその断片を合成することが可能である。例えば、種々の液相または固相技術を用いてペプチド合成を行うことができる (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, 55-60ページ; Roberge, J.Y. 他 (1995) Science 269:202-204)。自動合成はABI 431Aペプチドシンセサイザ (Applied Biosystems) を用いて達成し得る。更にCHOP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/または他のタンパク質または任意のその一部からの配列との組み合わせにより、天然のポリペプチド配列を有するポリペプチドまたは変異体ポリペプチドを作製することが可能である。

10

【0165】

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)。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)。

【0166】

生物学的に活性なCHOPを発現させるために、CHOP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し得る。好適な発現ベクターとは、好適な宿主に挿入されたコーディング配列の転写および翻訳の制御に必要なエレメント群を持つベクターである。これらのエレメントには、ベクター及びCHOP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。必要なエレメント群は、強度および特異性が様々である。特定の開始シグナルによって、CHOP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。開始シグナルの例には、ATG開始コドンと、コザック配列など近傍の配列とが含まれる。CHOP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、コード配列あるいはその断片のみが挿入された場合は、インフレームATG開始コドンなど外因性の翻訳制御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。外因性の翻訳エレメント群および開始コドン群は、様々な天然物および合成物を起源とし得る。発現の効率は、用いられる特定の宿主細胞系に好適なエンハンサーを

20

30

【0167】

当業者に周知の方法を用いて、CHOP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 (Sambrook 他, 前出, ch. 4, 8, および 16-17; Ausubel, F.M. 他 (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, および 16)。

【0168】

種々の発現ベクター/宿主系を利用して、CHOPをコードする配列を保持および発現し得る。限定するものではないがこのような発現ベクター/宿主系には、組換えバクテリオファージ、プラスミドまたはコスミドDNA発現ベクターで形質転換させた細菌などの微生物等や、酵母菌発現ベクターで形質転換させた酵母菌や、ウイルス発現ベクター (例えばバキュロウイルス) に感染した昆虫細胞系や、ウイルス発現ベクター (例えばカリフラワーモザイクウイルスCaMVまたはタバコモザイクウイルスTMV) または細菌発現ベクター (例えばTiプラスミドまたはpBR322プラスミド) で形質転換させた植物細胞系、動物細胞系がある (前出のSambrook 他; 前出のAusubel; Van Heeke, G. および S.M. Schuster (1989) J. Biol. Chem. 264:55035509; Engelhard, E.K. 他 (1994) Proc. Natl. Acad. Sci. U.S.A. 91:3224-3227; Sandig, V. 他 (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307311; The McGraw Hill Yearbook of Science and Technology (19

40

50

92) McGraw Hill, New York NY, pp. 191-196; Logan, J. および T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; および Harrington, J.J. 他 (1997) Nat. Genet. 15:345-355)。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、ポリヌクレオチド配列を標的器官、組織または細胞集団へ送達することができる (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. および N. Somia (1997) Nature 389:239-242)。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。

10

【0169】

細菌系では、CHOP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および発現ベクターを選択し得る。例えば、CHOP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(Invitrogen)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CHOP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レスキュー、入れ子状態の欠失の生成にも有用であろう例えば抗体類の産生などに多量のCHOPが必要な場合は、CHOPのハイレベル発現を指示するベクター類が使用できる。例えば、強力な誘導SP6 バクテリオファージプロモーターまたは誘導T7バクテリオファージプロモーターを含むベクターが使用できる。

20

【0170】

酵母の発現系を使用してCHOPを産出し得る。 α 因子、アルコールオキシダーゼ、PGHプロモーター等の構成型或いは誘導型のプロモーターを含む多数のベクターが、出芽酵母菌 (*Saccharomyces cerevisiae*) またはピキア酵母 (*Pichia pastoris*) に使用可能である。更に、このようなベクターは、発現したタンパク質を分泌或いは細胞内への保持のいずれかに誘導し、安定した増殖のために外来配列の宿主ゲノムへの組込みを可能にする (前出の Ausubel 他 (1995)、Bitter, G.A. 他 (1987) Methods Enzymol. 153:516-544、Scorer, C.A. 他 (1994) Bio/Technology 12:181-184)。

30

【0171】

植物系もCHOPの発現に用い得る。CHOPをコードする配列の転写は、ウイルスプロモーター、例えば単独あるいはTMV由来のオメガリーダー配列と組み合わせて用いられるようなCaMV由来の35Sおよび19Sプロモーターによって促進される (Takamatsu, N. (1987) EMBO J. 6:3073-311)。或いは、RUBISCOの小サブユニット等の植物プロモーターまたは熱ショックプロモーターを用いてもよい (Coruzzi, G. 他 (1984) EMBO J. 3:1671-1680、Broglie, R. 他 (1984) Science 224:838-843、Winter, J. 他 (1991) Results Probl. Cell Differ. 17:85-105)。これらの構成物は、直接DNA形質転換または病原体を媒介とする形質移入によって、植物細胞内に導入可能である (『マグローヒル科学技術年鑑』(The McGraw Hill Yearbook of Science and Technology) (1992) McGraw Hill New York NY, 191-196ページ)。

40

【0172】

哺乳類細胞においては、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CHOPをコードする配列を結合し得る。非必須のE1またはE3領域へウイルスのゲノムを挿入し、宿主細胞でCHOPを発現する感染ウイルスを得ることが可能である。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。

50

【0173】

ヒト人工染色体 (HAC) を用いて、プラスミドに含まれ得る断片やプラスミドから発現し得る断片より大きなDNAの断片群を送達することもできる。治療のために約 6 kb ~ 10 Mb のHACを作製し、従来の輸送方法 (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 (前出のHarrington 他)。

【0174】

哺乳動物系内の組換えタンパク質の長期にわたる産生のためには、細胞株におけるCHOP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、CHOP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、細胞を増殖させ得る。選択可能マーカーの目的は選択剤への抵抗性を与えることであり、選択可能マーカーの存在により、導入した配列をうまく発現するような細胞の成長およ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。

10

【0175】

任意の数の選択系を用いて、形質転換細胞株を回収できる。限定するものではないがこのような選択系には、 tk^{-} 細胞のために用いられる単純ヘルペスウイルスのチミジンキナーゼ遺伝子と、 apr^{-} 細胞のために用いられる単純ヘルペスウイルスのアデニンホスホリボシルトランスフェラーゼ遺伝子がある (Wigler, M. 他 (1977) Cell 11:223-232、Lowy, I. 他 (1980) Cell 22:817-823)。また、選択の基礎として代謝拮抗物質、抗生物質ある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はアミノグリコシド系ネオマイシン及びG-418に対する耐性を与え、alsはクロルスルフロンに対する耐性を、patはホスフィノトリシンアセチルトランスフェラーゼに対する耐性を各々与える (Wigler, M. 他 (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570、Colbere-Garapin, F. 他 (1981) J. Mol. Biol. 150:1-14)。その他の選択可能な遺伝子、例えば、代謝のための細胞の必要条件を変える $trpB$ 及び $hisD$ は、文献に記載されている (Hartman, S.C. 及び R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051)。可視マーカー、例えばアントシアニン、緑色蛍光タンパク質 (GFP; Clontech)、 β グルクロニダーゼ及びその基質 β 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、形質転換体を同定するだけでなく、特定のベクター系に起因する一過性或いは安定したタンパク質発現を定量することも可能である (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131)。

20

30

【0176】

マーカー遺伝子発現の有無によって目的の遺伝子の存在が示唆されても、その遺伝子の存在および発現の確認が必要な場合もある。例えば、CHOP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、CHOP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同定可能である。または単一プロモーター制御下で、CHOPをコードする配列とタンデムにマーカー遺伝子を配置することも可能である。誘導または選択に応答したマーカー遺伝子の発現は通常、タンデム遺伝子の発現も示す。

40

【0177】

一般に、CHOPをコードする核酸配列を含み且つNAAP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と、PCR増幅とがあり、また、核酸配列或いはタンパク質配列の検出、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もある。

【0178】

特異的ポリクローナル抗体または特異的モノクローナル抗体を用いてCHOPの発現の検出と計測とを行うための免疫学的方法は、当分野で既知である。このような技術の例としては

50

、酵素結合イムノソルベントアッセイ (ELISA)、ラジオイムノアッセイ (RIA)、蛍光活性化細胞選別 (FACS) などが挙げられる。CHOP上の2つの非干渉エピトープに反応するモノクローナル抗体を用いた、2部位のモノクローナルベースイムノアッセイ (two-site, monoclonal-based immunoassay) が好ましいが、競合の結合アッセイも用いることもできる。これらのアッセイ及びこれ以外のアッセイは、当分野で公知である (Hampton, R. 他 (1990) Serological Methods, a Laboratory Manual. APS Press, St Paul, MN, Sect. IV, Coligan, J. E. 他 (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY, Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ)。

【0179】

多岐にわたる標識方法及び抱合方法が、当業者に知られており、様々な核酸アッセイおよびアミノ酸アッセイにこれらの方法を用い得る。CHOPをコードするポリヌクレオチドに関連する配列を検出するための、標識されたハイブリダイゼーションプローブ或いはPCRプローブを産出する方法には、オリゴ標識化、ニックトランスレーション、末端標識化、または標識されたヌクレオチドを用いるPCR増幅法がある。別法として、CHOPをコードする配列、またはその任意の断片をmRNAプローブの生成の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などの好適なRNAポリメラーゼおよ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 Biosciences、Promega (Madison WI)、U.S. Biochemicalなど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あるいは標識には、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、発色剤などがある。

【0180】

CHOP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およ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、あるいはその両者に依存する。CHOP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CHOP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。

【0181】

更に、宿主細胞株の選択は、挿入した配列の発現をモジュレートする能力、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、脂質化およびアシル化がある。タンパク質の「プレプロ」または「プロ」形を切断する翻訳後のプロセッシングを利用して、タンパク質の標的への誘導、折りたたみ及び/または活性を特定することが可能である。翻訳後の活性のための固有の細胞装置及び特徴のある機構を有する種々の宿主細胞 (例えばCHO、HeLa、MDCK、MEK293、WI38等) は、American Type Culture Collection (ATCC, Manassas, VA) から入手可能であり、外来タンパク質の正しい修飾及び処理を確実にするように選択し得る。

【0182】

本発明の別の実施例では、CHOPをコードする天然の核酸配列、修飾された核酸配列、または組換えの核酸配列を上記した任意の宿主系の融合タンパク質の翻訳となる異種配列に連結させ得る。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CHOPタンパク質が、CHOP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および異種ペプチド部分も、市販されている親和性マトリックス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

10

20

30

40

50

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および赤血球凝集素（HA）は、これらのエピトープ標識を特異的に認識する市販のモノクローナル抗体およびポリクローナル抗体を用いた、融合タンパク質の免疫親和性精製を可能にする。また、CHOP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、CHOP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および精製方法は、前出のAusubel他（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および精製を促進することもできる。

【0183】

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

【0184】

本発明のCHOPまたはその断片を用いて、CHOP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、CHOP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。試験化合物としては、抗体、オリゴヌクレオチド、タンパク質（例えばリガンドまたは受容体）または小分子が挙げられる。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NTT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）。別の実施例においては、こうして同定された化合物は受容体CHOPの天然のリガンドである（たとえば、Howard, A.D. 他（2001）*Trends Pharmacol. Sci.* 22:132-140; Wise, A. 他（2002）*Drug Discovery Today* 7:235-246）。

【0185】

別の実施例においては、その化合物は、CHOP が結合する天然受容体、或いは少なくとも受容体のある断片、リガンド結合部位または結合ポケットの全体または一部を含む受容体のある断片に密接に関連し得る。たとえば、この化合物はシグナルを伝播する能力を持つCHOPの受容体であるかもしれないし、あるいはシグナルを伝播する能力のないCHOPのおとり受容体であるかもしれない（Ashkenazi, A. および V.M. Divit（1999）*Curr. Opin. Cell Biol.* 11:255-260; Mantovani, A. 他（2001）*Trends Immunol.* 22:328-336）。この化合物は既知の技法を用いて合理的に設計することができる。こうした技法の例としては、化合物エタネルセプト（etanercept）（ENBREL; Immunex Corp., Seattle WA）の作製に用いられた技法がある。この化合物はヒトのリューマチ性関節炎の治療に効果的である。エタネルセプトは遺伝子操作されたp75腫瘍壊死因子（TNF）受容体二量体であり、ヒトのIgG₁のFc部分に連結されている（Taylor, P.C. 他（2001）*Curr. Opin. Immunol.* 13:611-616）。

【0186】

一実施例では、CHOPへの特異的な結合、刺激または阻害を行う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CHOP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、または大腸菌からの細胞が含まれる。CHOPを発現する細胞またはCHOPを含有する細胞膜断片を試験化合物と接触させて、CHOPまたは化合物のいずれかの結合、刺激または阻害を分析する。

【0187】

或るアッセイは、単に試験化合物をポリペプチドに実験的に結合させ、フルオロフォア（蛍光標識試薬）、放射性同位体、酵素抱合体、または他の検出可能な標識により、その結合を検出できる。例えば、このアッセイは、少なくとも1つの試験化合物を、溶液中のあ

10

20

30

40

50

るいは固体支持物に固定されたCHOPと結合させるステップと、CHOPとこの化合物との結合を検出するステップを含む場合がある。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、無細胞再構成標本、化学ライブラリまたは天然の生成混合物を用いて実施することができ、試験化合物は、溶液中で遊離させるか固体支持体に固定させる。

【0188】

ある化合物が天然のリガンドに結合したり天然のリガンドがその天然の受容体に結合するのを阻害する能力を検定するためのアッセイを使うことができる。こうしたアッセイの例としては、米国特許第5,914,236号および第6,372,724号に記載されたような放射ラベルアッセイを含む。関連した実施態様では、1つ以上のアミノ酸置換が或るポリペプチド化合物(受容体など)に導入され、その天然リガンドに結合する能力を向上または改変しうる(Matthews, D.J. および J.A. Wells. (1994) Chem. Biol. 1:25-30)。もう一つの関連する実施例においては、ポリペプチド化合物(たとえばリガンド)に1つ以上のアミノ酸置換を行うことにより、天然受容体への結合能力を改善または改変することができる(Cunningham, B.C. および J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. 他 (1991) J. Biol. Chem. 266:10982-10988)。

10

【0189】

本発明のCHOPまたはその断片を用いて、CHOPの活性をモジュレートする化合物をスクリーニングし得る。このような化合物には、アゴニスト、アンタゴニスト、部分的アゴニスト、または逆アゴニスト等が含まれる。ある実施態様では、CHOPの活性が許容される条件下でアッセイを実施し、そのアッセイでは少なくとも1つの試験化合物をCHOPと混合し、試験化合物の存在下でCHOPの活性を試験化合物不在下でのCHOPの活性と比較する。試験化合物の存在下でのCHOPの活性の変化は、CHOPの活性をモジュレートする化合物の存在を示唆する。別の実施態様において、試験化合物をCHOPの活性に適した条件下でCHOPを含む *in vivo* 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、CHOPの活性を調節する試験化合物は間接的に結合する場合があります、その際は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。

20

【0190】

別の実施態様では、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、CHOPまたはその哺乳類相同体をコードするポリヌクレオチドを「ノックアウト」する。このような技術は当技術分野において周知であり、ヒト疾患動物モデルの作製に有用である(米国特許第5,175,383号及び第5,767,337号)。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)などのマーカー遺伝子で破壊した、目的の遺伝子を持つ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。或いは、Cre-loxP系を用いて相同組換えを行い、組織特異的または発生段階特異的に目的遺伝子をノックアウトする(Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. 他 (1997) Nucleic Acids Res. 25:4323-4330)。形質転換したES細胞を同定し、例えばC57BL/6マウス系等から採取したマウス細胞胚盤胞に微量注入する。胚盤胞を偽妊娠メスに外科的に導入し、得られるキメラ子孫の遺伝形質を決め、これを交配させてヘテロ接合性系またはホモ接合性系を作製する。このようにして作製した遺伝子組換え動物は、潜在的な治療薬や毒性薬物で検査されうる。

30

40

【0191】

CHOP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)。

50

【0192】

CHOPをコードするポリヌクレオチドを用いて、ヒト疾患をモデルとした「ノックイン」ヒト化動物（ブタ）または遺伝子組換え動物（マウスまたはラット）を作製することもできる。ノックイン技術を用いて、CHOPをコードするポリヌクレオチドのある領域を動物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GCRECを乳汁内に分泌するなどCHOP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）。

【0193】

10

（治療）

CHOPのある領域と糖結合タンパク質の領域との間に、例えば配列およびモチーフの内容における化学的および構造的類似性が存在する。さらに、CHOPを発現する組織の例が、表6および実施例11に見つけられる。従って、CHOPは、糖質代謝、細胞増殖異常、自己免疫/炎症性の疾患、生殖系疾患、遺伝疾患、輸送疾患、神経系疾患および癌においてある役割を果たすと考えられる。CHOPの発現若しくは活性の増大に関連する疾患の治療においては、CHOPの発現または活性を低下させることが望ましい。また、CHOP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、CHOPの発現または活性を増大させることが望ましい。

【0194】

20

従って、一実施例において、CHOP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CHOPまたはその断片や誘導体を投与することが可能である。限定されるものではないが、そのような疾患として、先天性II型異常赤血球産生症性貧血、糖尿病、インスリン依存型真性糖尿病、非インスリン依存型真性糖尿病、フルクトース-1,6-ジホスファターゼ欠乏症、ガラクトース血症、グルカゴノーマ、遺伝性フルクトース不耐症、低血糖症、マンノシドーシス、ノイラミニダーゼ欠乏症、肥満、ガラクトースエピメラーゼ欠乏症、糖原病、リソソーム蓄積症、果糖尿、五炭糖尿症、糖質欠損糖タンパク質症候群（1A型および1B型CDGS）、自己免疫性甲状腺疾患、アスパルチルグルコサミン尿症、GM₁ ガングリオシド蓄積症、GM₂ ガングリオシド蓄積症、β-ガタクトシダーゼβ-N-アセチルヘキソサミニダーゼ欠乏症、糖脂質蓄積症、神経機能不全、シアリドーシス、肝脾腫大症、および遺伝性ピルビン酸代謝異常が含まれ； 細胞増殖異常には日光性角化症、動脈硬化、アテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病（MCTD）、骨髓線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症が含まれ； 自己免疫/炎症性疾患には、後天性免疫不全症候群（AIDS）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血症、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、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー皮膚炎、皮膚筋炎、真性糖尿病、肺気腫、リンパ球毒素性一時的リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、骨粗しょう症、膵炎、多発性筋炎、乾癬、ライター症候群、慢性関節リウマチ炎、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、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ； 生殖系疾患としては、プロラクチン産生障害、不妊症、卵管疾患、排卵欠損症、子宮内膜症、性周期障害、月経周期障害、多嚢胞性卵巣症候群、卵巣過刺激症候群、子宮内膜癌、卵巣癌、子宮筋腫、自己免疫疾患、子宮外妊娠、催奇形、乳がん、繊維嚢胞性乳房疾患、乳漏症、精子形成破壊、精子異常生理、精巣癌、前立腺癌、良性前立腺肥大、前立腺炎、パイロニー病、性交不能、男性乳癌、女性化乳房、高ゴナドトロピン性腺機能低下症、低ゴナドトロピン性腺機能低下症、偽性半陰陽、無精子症、早発性卵巣不全症、アクロシン欠乏症

30

40

50

、思春期遅延、逆行性射精、無射精anejaculation、および血管芽細胞腫、嚢胞褐色細胞腫(cystsphaeochromocytomas)、傍神経節腫、副睾丸の嚢腺腫、および内リンパ嚢腫瘍が含まれ； 遺伝疾患としては副腎脳白質ジストロフィー、Alport症候群、コロイデレミア、デュシェンヌおよびベッカー型筋ジストロフィー、ダウン症候群、嚢胞性線維症、慢性肉芽腫症、ゴーシェ病、ハンチントン病、Marfan症候群、筋ジストロフィー、筋強直性ジストロフィー、ピクノディスオストーシス、Refsum症候群、網膜芽細胞腫、鎌形赤血球貧血症、サラセミア、ウェルナー症候群、von Willebrand症候群、ウィルムス腫瘍、ゼルウィガー症候群、ペルオキシソームアシルCoAオキシダーゼ欠乏症、ペルオキシソームチオラーゼ欠乏症、ペルオキシソーム二価性タンパク質欠乏症、ミトコンドリアカルニチンパルミトイルトランスフェラーゼおよびカルニチン欠乏症、ミトコンドリア長鎖アシルCoAデヒドロゲナーゼ欠乏症、ミトコンドリア中鎖アシルCoAデヒドロゲナーゼ欠乏症、ミトコンドリア短鎖アシルCoAデヒドロゲナーゼ欠乏症、ミトコンドリア電子伝達フラビンタンパク質および電子伝達フラビンタンパク質：ユービキノンオキシドレダクターゼ欠乏症、ミトコンドリア三価性タンパク質欠乏症およびミトコンドリア短鎖3ヒドロキシアシルCoAデヒドロゲナーゼ欠乏症が含まれ； 輸送疾患としては、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症、顔面神経麻痺、シャルコー・マリー・トゥース病、尿崩症、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、悪性高体温症、多剤耐性、重症筋無力症、ジストニー、末梢神経障害； 輸送に関連する心臓疾患として例えば狭心症、徐脈性不整脈、頻脈性不整脈、高血圧、QT延長症候群、心筋炎、心筋症、ネマリンミオパシー、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、感染性筋炎、多発性筋炎； 輸送に関連する神経疾患として例えばアルツハイマー病、記憶喪失、双極性障害、痴呆、鬱病、癲癇、トゥーレット病、パラノイド精神病、および精神分裂病(統合失調症)； 輸送に関連するその他の疾患として例えば神経線維腫症、ヘルペス後神経痛、三叉神経障害、サルコイドーシス、鎌状赤血球貧血、ウィルソン病、白内障、不妊症、肺動脈狭窄、感音常染色体難聴、高血糖症、低血糖症、グレーヴズ病、甲状腺腫、クッシング病、アジソン病、グルコース・ガラクトース吸収不良症候群、グリコーゲン貯蔵病、高コレステロール血症、副腎脳白質ジストロフィー、Menkes病、後角症候群、von Gierke病、シスチン尿症、イミノグリシン尿症、Hartup病、およびファンコーニ病が含まれ； 神経の疾患の中には、虚血性脳血管障害、脳卒中、大脳新生物、ピック病、ハンチントン病、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、プリオン病(クールー、クロイツフェルト-ヤコブ病、ゲルストマン症候群、及びGerstmann-Straussler-Scheinker症候群を含む)、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、ダウン症候群を含む中枢神経系性の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、脊髄障害、筋ジストロフィー及びその他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎、遺伝性、代謝性、内分泌性、及び中毒性の筋疾患、重症筋無力症、周期性四肢麻痺、精神病(気分性、不安性の障害)、季節性障害(SAD)、静座不能、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、帯状疱疹後神経痛、進行性核上麻痺、大脳皮質基底核部変性症、及び家族性の前頭側頭性健忘症とが含まれ； 癌としては腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、および奇形癌など、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、および子宮の癌が含まれる。

10

20

30

40

【0195】

別の実施態様では、CHOP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して

50

、上記しだけに限られるものではないが疾患を含むCHOP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。

【0196】

さらに別の実施態様では、実質的に精製されたCHOPを含む組成物を好適な医薬用キャリアと共に患者に投与して、限定するものではないが上記した疾患を含むCHOP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。

【0197】

更に別の実施例では、CHOP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CHOP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。

10

【0198】

更なる実施例では、CHOP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CHOP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、上記した細胞増殖異常、自己免疫/炎症性の疾患、生殖系疾患、および神経系疾患が含まれる。一実施態様では、CHOP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CHOP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送達機構として間接的に用いられ得る。

【0199】

別の実施態様では、限定するものではないが上に列記した疾患を含む、CHOP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、CHOP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被験者に投与し得る。

20

【0200】

別の実施態様では、本発明の任意のタンパク質、アンタゴニスト、抗体、アゴニスト、相補配列、またはベクターを、別の好適な治療薬と組み合わせて投与することもできる。併用療法で用いる好適な治療薬は、当業者が従来の医薬原理に従って選択し得る。治療薬と組合せることにより、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。

【0201】

CHOP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CHOP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CHOPと特異的に結合するものの同定が可能である。CHOPへの抗体も、本技術分野で公知の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、Fab断片、およびFab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。一本鎖抗体（例えば由来はラクダまたはラマ）は強力な酵素阻害剤である可能性があり、ペプチド擬態の設計および免疫吸着剤およびバイオセンサの開発に有用である可能性がある(Muyldermans, S. (2001) J. Biotechnol. 74:277-302)。

30

【0202】

抗体の産生のためには、ヤギ、ウサギ、ラット、マウス、ラクダ、ヒトコブラクダ、ラマ、ヒト及びその他のものを含む種々の宿主が、CHOP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、プルロニックポリオール、ポリアニオン、ペプチド、油性乳剤、スカシガイのヘモシニアン、ジニトロフェノール等の界面活性剤とがある。ヒトに用いられるアジュバントの中では、BCG（カルメット - ゲラン杆菌）及びコリネバクテリウム - パルバム（*Corynebacterium parvum*）が特に好ましい。

40

【0203】

50

CHOPに対する抗体を誘導するために用いるオリゴペプチド、ペプチドまたは断片は、少なくとも約5個のアミノ酸からなるアミノ酸配列を持つものが好ましく、一般的には約10個以上のアミノ酸からなるものとなる。これらのオリゴペプチド、ペプチドまたは断片は、天然のタンパク質のアミノ酸配列の一部と同一であることが望ましい。CHOPの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。

【0204】

CHOP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EBV-ハイブリドーマ技術がある (Kohler, G. 他 (1975) Nature 256:495-497, Kozbor, D. 他 (1985) J. Immunol. Methods 81:31-42, Cote, R.J. 他 (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030, Cole, S.P. 他 (1984) Mol. Cell Biol. 62:109-120)。

10

【0205】

更に、ヒト抗体遺伝子にマウス抗体遺伝子をスプライシングするなどの、「キメラ抗体」作製のために発達した技術が、好適な抗原特異性及び生物学的活性を備える分子を得るために用いられる (Morrison, S.L. 他 (1984) Proc. Natl. Acad. Sci. 81: 4851-4855; Neuberger, M.S. 他 (1984) Nature 312:604-608; Takeda, S. 他 (1985) Nature 314:452, 454)。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、CHOP特異性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)。

20

【0206】

抗体の産生は、リンパ球集団における *in vivo* 産生の誘導によって、或いは文献に開示されているように非常に特異的な結合試薬の免疫グロブリンライブラリまたはパネルのスクリーニングによっても行い得る (Orlandi, R. 他 (1989) Proc. Natl. Acad. Sci. USA 86: 3833-3837, Winter, G. 他 (1991) Nature 349:293-299)。

【0207】

CHOPに対する特異結合部位を持つ抗体断片をも産生し得る。例えば、限定するものではないが、このような断片には、抗体分子のペプシン消化によって作製される $F(ab')_2$ 断片と、 $F(ab')_2$ 断片のジスルフィド架橋を還元することによって作製される Fab断片とがある。或いは、Fab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 Fab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 (Huse, W.D. 他 (1989) Science 246:1275-1281)。

30

【0208】

種々の免疫学的検定 (イムノアッセイ) を用いてスクリーニングすることにより、所望の特異性を有する抗体を同定し得る。確立された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる競合結合アッセイ、または免疫放射線アッセイ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、CHOPとその特異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 CHOP エピトープに対して反応性のモノクローナル抗体を用いる、2部位モノクローナルベースのイムノアッセイが一般に利用されるが、競合的結合アッセイも利用することができる (Pound、前出)。

40

【0209】

ラジオイムノアッセイ技術と共に Scatchard 分析などの様々な方法を用いて、CHOPに対する抗体の親和性を評価し得る。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 CHOP 抗体複合体のモル濃度を遊離抗体と遊離抗原のモル濃度で除して得られる値である。多数の CHOP エピトープに対して親和性が不均一なポリクローナル抗体医薬の K_a は、CHOP に対する抗体の平均親和性または結合活性を表す。モノクローナル抗体は或る特定の CHOP

50

エピトープに対して単一特異的であり、モノクローナル抗体の或る製剤について判定した K_a は、親和性の真の測定値を表す。 K_a 値が約 $10^9 \sim 10^{12}$ L/mol の範囲にあるような高親和性抗体試薬は、CHOP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 $10^6 \sim 10^7$ L/mol の低親和性抗体試薬は、CHOPが抗体から最終的に活性化状態で解離する必要がある免疫精製 (immunopurification) 及び類似の処理に用いるのが好ましい。(Catty, D. (1988) Antibodies, Volume I: A Practical Approach. IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

【0210】

ポリクローナル抗体製剤の抗体価およ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およ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製剤は一般に、CHOP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 (Catty、前出、Coligan 他、前出)。

【0211】

本発明の別の実施態様では、CHOP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を、治療目的で使用する。ある実施態様では、CHOPをコードする遺伝子のコーディング領域や調節領域に相補的な配列やアンチセンス分子 (DNA及びRNA、修飾ヌクレオチド) を設計して遺伝子発現を変更することができる。このような技術は当分野では周知であり、センスまたはアンチセンスオリゴヌクレオチドまたは大きな断片が、CHOPをコードする配列の制御領域から、またはコード領域に沿ったさまざまな位置から設計可能である (Agrawal, S., 編集 (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ 等を参照)。

【0212】

治療に用いる場合、アンチセンス配列を適切な標的細胞に導入するのに好適な、任意の遺伝子送達系を用いることができる。アンチセンス配列は、転写時に標的タンパク質をコードする細胞配列の少なくとも一部に相補的な配列を発現する発現プラスミドの形で細胞内に送達することが可能である (Slater, J.E. 他 (1998) *J. Allergy Clin. Immunol.* 102 (3):469-475 及び Scanlon, K.J. 他 (1995) 9(13):1288-1296)。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)。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システムが含まれる (Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R. J. 他 (1998) *J. Pharm. Sci.* 87(11):1308-1315; Morris, M.C. 他 (1997) *Nucleic Acid Res.* 25(14):2730-2736)。

【0213】

本発明の別の実施態様では、CHOP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得る。遺伝子治療を行うことにより、(i) 遺伝子欠損症 (例えば X 染色体鎖遺伝 (Cavazzana-Calvo, M. 他 (2000) *Science* 288:669-672) により特徴付けられる重度の複合型免疫欠損 (SCID)-X1 の場合)、先天性アデノシンデアミナーゼ (ADA) 欠損症に関連する重度の複合型免疫欠損 (Blaese, R.M. 他 (1995) *Science* 270:475-480; Bordignon, C. 他 (1995) *Science* 270:470-475)、嚢胞性繊維症 (Zabner, J. 他 (1993) *Cell* 75:207-216; Crystal, R.G. 他 (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. 他 (1995) *Hum. Gene Therapy* 6:667-703)、サラセミア (thalassaemia)、家族性高コレステロール血症、第 VIII 因子若しくは第 IX 因子欠損に起因する血友病 (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. および N. Somia (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、Poeschla, E. 他 (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399)、B型若しくはC型肝炎ウイルス(HBV、HCV)、Candida albicans及びParacoccidioides brasiliensis等の真菌寄生虫、並びにPlasmodium falciparum及びTrypanosoma cruzi等の原虫寄生体に対する防御機能を有するタンパク質を発現させることができる。CHOPの発現若しくは調節に必要な遺伝子の欠損が疾患を引き起こす場合、導入した細胞の好適な集団からCHOP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。

【0214】

本発明の更なる実施態様では、CHOPの欠損による疾患や障害を、CHOPをコードする哺乳類発現ベクターを作製し、これらのベクターを機械的手段によってCHOP欠損細胞に導入することによって治療する。in vivoあるいは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) リボソームを介した形質移入、(iv) 受容体を介した遺伝子導入、および(v) DNAトランスポゾンの使用がある(Morgan, R.A. および W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217、Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. および H. Recipon (1998) Curr. Opin. Biotechnol. 9:445-450)。

【0215】

CHOPの発現に効果的である発現ベクターには、限定するものではないが、PCDNA 3.1、EPI TAG、PRCCMV2、PREP、PVAX、PCR2-TOPOTA ベクター (Invitrogen, Carlsbad CA)、PCM VSCRIPT、PCMV-TAG、PEGSH/PERV (Stratagene, La Jolla CA) および PTETOFF、PTETON、PTRE2、PTRE2LUC、PTKHYG (Clontech, Palo Alto CA) が含まれる。CHOPを発現させるために、(i) 恒常的に活性なプロモーター (例えば、サイトメガロウイルス (CMV)、ラウス肉腫ウイルス (RSV)、SV40ウイルス、チミジンキナーゼ (TK)、若しくは β -アクトシン遺伝子等)、(ii) 誘導性プロモーター (例えば、市販されているT-REXプラスミド (Invitrogen) に含まれている、テトラサイクリン調節性プロモーター (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen, M. 他 (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456))、エクジソン誘導性プロモーター (市販されているプラスミドPVGRXR及びPINDに含まれている: Invitrogen)、FK506/ラパマイシン誘導性プロモーター、またはRU486/ミフェプリストーン誘導性プロモーター (Rossi および Blau, 前出)、または (iii) 正常な個体に由来するGCREC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。

【0216】

市販のリボソーム形質転換キット (例えば Invitrogen 社の PERFECT LIPID TRANSFECTION KIT) を用いれば、当業者は経験にそれほど頼らないでもポリヌクレオチドを培養中の標的細胞に導入することが可能になる。別法では、リン酸カルシウム法 (Graham, F.L. 及び A.J. Eb (1973) Virology 52:456-467) 若しくは電気穿孔法 (Neumann, B. 他 (1982) EMBO J. 1:841-845) を用いて形質転換を行う。初代培養細胞にDNAを導入するためには、標準化された哺乳動物の形質移入プロトコルの修飾が必要である。

【0217】

本発明の別の実施態様では、CHOPの発現に関連する遺伝子欠損によって起こる疾患や障害は、(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調節の下でCHOP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ド配列を伴う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. 他 (1995) Proc. Natl. Acad. Sci. U.S.A. 92:6733-6737) に基づいている。上記データを引用することをもって本明細書の一部とする。ベクターは、好適なベクター産生細胞株 (VPCL) において増殖され、VPCLは、標的細胞上の受容体に対する親和性を有するエンベロープ遺伝子またはVSVg等の

10

20

30

40

50

汎親和性エンベロープタンパク質を発現する (Armentano, D. 他 (1987) J. Virol. 61:1647-1650、Bender, M.A. 他 (1987) J. Virol. 61:1639-1646、Adam, M.A. および A.D. Miller (1988) J. Virol. 62:3802-3806、Dull, T. 他 (1998) J. Virol. 72:8463-8471、Zufferey, R. 他 (1998) J. Virol. 72:9873-9880)。Riggに付与された米国特許第5,910,434号 (「Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant」) において、レトロウイルスパッケージング細胞株を得るための方法が開示されており、引用することをもって本明細書の一部とする。レトロウイルスベクターの増殖、細胞集団 (例えばCD4⁺ T細胞) の形質導入、及び形質導入した細胞の患者への戻しは、遺伝子治療の分野では当業者に公知の方法であり、多数の文献に記載されている (Ranga, U. 他 (1997) J. Virol. 71:7020-7029、Bauer, G. 他 (1997) Blood 89:2259-2267、Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716、Ranga, U. 他 (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206、Su, L. (1997) Blood 89:2283-2290)。

【0218】

別法では、アデノウイルス系遺伝子治療の或る送達系を用いて、CHOPの発現に関連する1つ以上の遺伝子異常を有する細胞群に、CHOPをコードするポリヌクレオチド群を送達する。アデノウイルス系ベクターの作製及びパッケージングについては、当業者に公知である。複製欠損型アデノウイルスベクターは、免疫調節タンパク質をコードする遺伝子を脾臓の無損傷の脾島内に導入するために融通のきくことが証明された (Csete, M.E. 他 (1995) Transplantation 27:263-268)。使用できる可能性のあるアデノウイルスベクターは、Armentanoに付与された米国特許第5,707,618号 (「Adenovirus vectors for gene therapy」) に記載されており、引用することをもって本明細書の一部とする。アデノウイルスベクターについては、Antinozzi 他 (1999) Annu. Rev. Nutr. 19:511-544 及び Vermaおよび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。

【0219】

別法では、ヘルペス系遺伝子治療の送達系を用いて、CHOP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、CHOP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CHOP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. 他 (1999) Exp. Eye Res. 169:385-395)。HSV-1ウイルスベクターの作製についても、DeLucaに付与された米国特許第5,804,413号 (「Herpes simplex virus swains for gene transfer」) に開示されており、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他 (1999) J. Virol. 73:519-532 及び Xu他 (1994) Dev. Biol. 163:152-161も参照されたい。両文献は、引用をもって本明細書の一部とする。クローン化ヘルペスウイルス配列の操作、巨大ヘルペスウイルスのゲノムの異なった部分を含む多数のプラスミドを形質移入した後の組換えウイルスの産生、ヘルペスウイルスの成長及び増殖、並びにヘルペスウイルスの細胞への感染は、当業者に公知の技術である。

【0220】

別法では、或るαウイルス (正の一本鎖RNAウイルス) ベクターを用いて、CHOPをコードするポリヌクレオチド群を標的細胞群に送達する。プロトタイプのαウイルスであるセムリキ森林熱ウイルス (Semliki Forest Virus, SFV) の生物学的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. 及び K.-J. Li (1998) Cun. Opin. Biotech. 9:464-469)。αウイルスRNAの複製中に、通常はウ

イルス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、CHOPをコードする配列をαウイルスゲノムのキャ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CHOPをコードするRNAが産生され、高いレベルでCHOPが合成される。通常はα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、αウイルスの溶解複製を遺伝子治療に適用できるように好適に変更可能であることを示唆している（Dryga, S. A. 他（1997）*Virology* 228 :74-83）。様々な宿主にαウイルスを導入できるので、様々なタイプの細胞にCHOP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。αウイルスの感染性cDNAクローンの処置方法、αウイルスのcDNA及びRNAの形質移入方法及びαウイルスの感染方法は、当業者に公知である。

10

【0221】

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(transcription initiation site)とは例えばスタート部位(start site)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充分に開く、二重らせんの能力を阻害するので有用である。三重らせんDNAを用いる最近の治療の進歩については文献に記載がある（Gee, J.E. 他（1994）in: Huber, B.E.及び B.L. Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co., Mt. Kisco, NY, 163-177ページ）。相補配列またはアンチセンス分子もまた、転写物がリボソームに結合するのを阻止することによってmRNAの翻訳を阻止するべく設計することができる。

20

【0222】

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

30

【0223】

任意の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して標的分子をスキャンすることによって先ず同定される。一度同定されると、切断部位を含む標的遺伝子の領域に対応する15～20リボヌクレオチドの短いRNA配列が、そのオリゴヌクレオチドを機能不全にするような2次構造の特徴をもっていないかを評価することが可能になる。候補標的の適合性の評価も、リボヌクレアーゼ保護アッセイを用いて相補的オリゴヌクレオチドとのハイブリダイゼーションのアクセス可能性をテストすることによって行うことができる。

【0224】

本発明の相補リボ核酸分子およびリボザイムは、核酸分子合成のために当分野でよく知られている任意の方法を用いて作製し得る。任意の方法には、固相フォスフォアミダイト化学合成等のオリゴヌクレオチドを化学的に合成する方法がある。或いは、CHOP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。

40

【0225】

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としては、分子の5'末端、3'末端、あるいはその両方にお

50

いて隣接配列群を追加することや、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ロチオエートまたは2'-O-メチルを使用することが含まれる。この概念は、PNAの産出に固有のものであり、これら全ての分子に拡大することができる。それには、内因性エンドヌクレアーゼによって容易には認識されないアデニン、シチジン、グアニン、チミン、及びウリジンにアセチル -、メチル -、チオ - 及び同様の修飾をしたものの他、非従来型塩基、例えばイノシン、クエオシン (queosine)、ワイプトシン (wybutosine) 等を加えることでできる。

【0226】

本発明の更なる実施例は、CHOPをコードするポリヌクレオチドの発現の改変に有効な化合物をスクリーニングする方法を含む。限定するものではないが特異ポリヌクレオチドの発現変化を起こすのに有効な化合物には、オリゴヌクレオチド、アンチセンスオリゴヌクレオチド、三重らせん形成オリゴヌクレオチド、転写因子その他のポリペプチド転写制御因子、及び特異ポリヌクレオチド配列と相互作用し得る非高分子化学的実体がある。有効な化合物は、ポリヌクレオチド発現のインヒビターまたはエンハンサーのいずれかとして作用することによりポリヌクレオチド発現を変異し得る。従って、CHOP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、CHOP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、CHOP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、CHOPをコードするポリヌクレオチドの発現を特異的に促進する化合物が治療上有用であり得る。

【0227】

或る特定ポリヌクレオチドの発現を改変する際の有効性について、少なくとも1個、または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、既に、市販のまたは私的な、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。CHOP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。CHOPをコードするポリヌクレオチドの発現における改変は、当分野で周知の任意の方法でアッセイする。通常は、CHOP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発現改変に有効な化合物に対して、例えば分裂酵母 (*Schizosaccharomyces pombe*) 遺伝子発現系 (Atkins, D. 他 (1999) 米国特許第5,932,435号、Arndt, G.M. 他 (2000) Nucleic Acids Res. 28:E15) またはHeLa細胞等のヒト細胞株 (Clarke, M.L. 他 (2000) Biochem. Biophys. Res. Commun. 268:8-13) を用いてスクリーニングを実行する。本発明の或る特定の実施態様は、或る特定ポリヌクレオチド配列に対するアンチセンス活性について、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、修飾した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過程に関する (Bruce, T.W. 他 (1997) U.S. Patent No. 5,686,242、Bruce, T.W. 他 (2000) U.S. Patent No. 6,022,691)。

【0228】

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、リポソーム注入またはポリカチオンアミノポリマーによる

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

【0229】

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、サルなどの哺乳類を含めて治療が必要な全ての被験体に適用できる。

【0230】

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

10

【0231】

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

【0232】

肺から投与する組成物は、液状または乾燥粉末状で調製し得る。このような組成物は通常、患者が吸入する直前にエアロゾル化する。小分子 (例えば従来の低分子量有機薬) 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可能にしたことにより、インスリン等の薬剤を実質的に血液循環へ輸送することを可能にした (Patton, J.S. 他, 米国特許第5,997,848号)。肺送達は、針注射なしに投与する点で優れており、有毒な可能性のある浸透エンハンサーの必要性をなくす。

20

【0233】

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

【0234】

CHOPを有する、またはその断片群を有する高分子群を直接、細胞内に送達すべく、特殊な種々の形状の組成物群が調製され得る。例えば、細胞不透過性高分子を含むリポソーム製剤は、細胞融合及び高分子の細胞内送達を促進し得る。別法では、CHOP またはその断片をHIV Tat-1タンパク質の陽イオンN末端部に結合することもできる。このようにして生成された融合タンパク質は、マウスモデル系の脳を含む全ての組織の細胞に形質導入することがわかっている (Schwarze, S.R. 他 (1999) Science 285:1569-1572)。

30

【0235】

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

40

【0236】

治療有効量とは、症状や容態を回復させる、たとえばCHOPまたはその断片、CHOPの抗体、CHOPのアゴニストまたはアンタゴニスト、インヒビターなど、活性成分の量を指す。治療有効性および毒性は、細胞培養または動物実験における標準的な薬学手法によって、例えばED₅₀ (集団の50%の治療有効量) またはLD₅₀ (集団の50%の致死量) 統計を計算するなどして判定できる。毒性効果の治療効果に対する投与量の比は、治療指数であり、LD₅₀/ED₅₀比として表すことができる。高い治療指数を示すような組成物が望ましい。細胞培養アッセイと動物実験とから得られたデータは、ヒトに用いる投与量の範囲の策定に用いられる。このような組成物が含まれる投与量は、毒性を殆ど或いは全く含まず、ED₅₀を含むような血中濃度の範囲にあることが好ましい。用いられる投与形態、患者の感受性及

50

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。

【0237】

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

【0238】

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

【0239】

(診断)

別の実施例では、CHOPに特異的に結合する抗体を、CHOPの発現によって特徴付けられる障害の診断に、または、CHOPやCHOP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得る。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。CHOPの診断アッセイには、ヒトの体液から、あるいは細胞や組織の抽出物から、抗体および標識を用いてCHOP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。

【0240】

CHOPを測定するための様々なプロトコル、例えばELISA、RIA、FACS等が当分野では周知であり、変容した、あるいは異常なレベルのCHOPの発現を診断するための基盤を提供する。正常或いは標準的なCHOPの発現の値は、複合体の形成に適した条件下で、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CHOPに対する抗体とを結合させることによって決定する。標準複合体形成量は、種々の方法、例えば測光法で定量できる。被験者、対照、及び疾患生検組織からの各サンプルのCHOPの発現の量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。

【0241】

本発明の別の実施例では、CHOPをコードするポリヌクレオチドを診断のために用いることもできる。用いられる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、疾患と相関し得るCHOP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、CHOPの存在の有無、更に過剰な発現を調べ、治療中のCHOP値の調節を監視する。

【0242】

一実施形態では、CHOP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、CHOPをコードする核酸配列を同定することが可能である。プローブが高度に特異的な領域(例えば5'調節領域)から作られている、或いはやや特異性の低い領域(例えば保存されたモチーフ)から作られているかにかかわらず、そのプローブの特異性と、ハイブリダイゼーション或いは増幅のストリンジェンシーによって、そのプローブがCHOPをコードする自然界の配列のみを同定するかどうか、或いは対立遺伝子や関連配列コードする自然界の配列のみを

10

20

30

40

50

同定するかどうかが決まるであろう。

【0243】

プローブはまた、関連する配列の検出に利用され、CHOP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。本発明のハイブリダイゼーションプローブはDNAまたはRNAとすることができ、SEQ ID NO:11-20の配列、或いはCHOP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。

【0244】

CHOPをコードするDNA群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としては、CHOPまたはCHOP誘導体をコードするポリヌクレオチド配列を、mRNAプローブ作製用ベクターにクローニングする方法を含む。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、 ^{32}P または ^{35}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。

【0245】

CHOPをコードするポリヌクレオチド配列を、CHOPの発現に関係する疾患の診断に用い得る。限定されるものではないが、そのような疾患として、先天性II型異常赤血球産生症性貧血、糖尿病、インスリン依存型真性糖尿病、非インスリン依存型真性糖尿病、フルクトース-1,6-ジホスファターゼ欠乏症、ガラクトース血症、グルカゴノーマ、遺伝性フルクトース不耐症、低血糖症、マンノシドーシス、ノイラミニダーゼ欠乏症、肥満、ガラクトースエピメラーゼ欠乏症、糖原病、リソソーム蓄積症、果糖尿、五炭糖尿症、糖質欠損糖タンパク質症候群(1A型および1B型CDGS)、自己免疫性甲状腺疾患、アスパルチルグルコサミン尿症、GM₁ ガングリオシド蓄積症、GM₂ ガングリオシド蓄積症、β-ガタクトシダーゼβ-N-アセチルヘキソサミニダーゼ欠乏症、糖脂質蓄積症、神経機能不全、シアリドーシス、肝脾腫大症、および遺伝性ピルビン酸代謝異常が含まれ；細胞増殖異常には日光性角化症、動脈硬化、アテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病(MCTD)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症が含まれ；自己免疫/炎症性疾患には、後天性免疫不全症候群(AIDS)、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血症、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、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー皮膚炎、皮膚筋炎、真性糖尿病、肺気腫、リンパ球毒素性一時的リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性大腸症候群、多発性硬化症、重症筋無力症、心筋または心膜炎、骨関節炎、骨粗しょう症、腓炎、多発性筋炎、乾癬、ライター症候群、慢性関節リウマチ炎、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、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷が含まれ；生殖系疾患としては、プロラクチン産生障害、不妊症、卵管疾患、排卵欠損症、子宮内膜症、性周期障害、月経周期障害、多嚢胞性卵巣症候群、卵巣過刺激症候群、子宮内膜癌、卵巣癌、子宮筋腫、自己免疫疾患、子宮外妊娠、催奇形、乳がん、繊維嚢胞性乳房疾患、乳漏症、精子形成破壊、精子異常生理、精巣癌、前立腺癌、良性前立腺肥大、前立腺炎、パイロニー病、性交不能、男性乳癌、女性化乳房、高ゴナドトロピン性腺機能低下症、低ゴナドトロピン性腺機能低下症、偽性半陰陽、無精子症、早発性卵巣不全症、アクロシン欠乏症、思春期遅延、逆行性射精、無射精anejaculation、および血管芽細胞腫、嚢胞褐色細胞腫(cystsphaeochromocytomas)、傍神経節腫、副睾丸の嚢腺腫、および内リンパ嚢腫瘍が含まれ；遺伝疾患としては副腎脳白質ジストロフィー、Alport症候群、コロイデレミア、デュシェンヌおよびベッカー型筋ジストロフィー、ダウン症候群、嚢胞性線

10

20

30

40

50

維症、慢性肉芽腫症、ゴーシェ病、ハンチントン病、Marfan症候群、筋ジストロフィー、筋強直性ジストロフィー、ピクノディスオストーシス、Refsum症候群、網膜芽細胞種、鎌形赤血球貧血症、サラセミア、ウェルナー症候群、von Willebrand症候群、ウィルムス腫瘍、ゼルウィガー症候群、ペルオキシソームアシルCoAオキシダーゼ欠乏症、ペルオキシソームチオラーゼ欠乏症、ペルオキシソーム二価性タンパク質欠乏症、ミトコンドリアカルニチンパルミトイルトランスフェラーゼおよびカルニチン欠乏症、ミトコンドリア長鎖アシルCoAデヒドロゲナーゼ欠乏症、ミトコンドリア中鎖アシルCoAデヒドロゲナーゼ欠乏症、ミトコンドリア短鎖アシルCoAデヒドロゲナーゼ欠乏症、ミトコンドリア電子伝達フラビンタンパク質および電子伝達フラビンタンパク質：ユービキノンオキシドレダクターゼ欠乏症、ミトコンドリア三価性タンパク質欠乏症およびミトコンドリア短鎖3ヒドロキシアシルCoAデヒドロゲナーゼ欠乏症が含まれ； 輸送疾患としては、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症、顔面神経麻痺、シャルコー・マリー・トゥース病、尿崩症、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、悪性高体温症、多剤耐性、重症筋無力症、ジストニー、末梢神経障害；

輸送に関連する心臓疾患として例えば狭心症、徐脈性不整脈、頻脈性不整脈、高血圧、QT延長症候群、心筋炎、心筋症、ネマリンミオパシー、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、感染性筋炎、多発性筋炎； 輸送に関連する神経疾患として例えばアルツハイマー病、記憶喪失、双極性障害、痴呆、鬱病、癲癇、トゥーレット病、パラノイド精神病、および精神分裂病(統合失調症)； 輸送に関連するその他の疾患として例えば神経線維腫症、ヘルペス後神経痛、三叉神経障害、サルコイドーシス、鎌状赤血球貧血、ウィルソン病、白内障、不妊症、肺動脈狭窄、感音常染色体難聴、高血糖症、低血糖症、グレーブズ病、甲状腺腫、クッシング病、アジソン病、グルコース・ガラクトース吸収不良症候群、グリコーゲン貯蔵病、高コレステロール血症、副腎脳白質ジストロフィー、Menkes病、後角症候群、von Gierke病、シスチン尿症、イミノグリシン尿症、Hartup病、およびファンコーニ病が含まれ； 神経の疾患の中には、虚血性脳血管障害、脳卒中、大脳新生物、ピック病、ハンチントン病、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、プリオン病(クールー、クロイツフェルト-ヤコブ病、ゲルストマン症候群、及びGerstmann-Straussler-Scheinker症候群を含む)、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(ce rebelloretinal hemangioblastomatosis)、脳3又神経血管症候群、ダウン症候群を含む中枢神経系の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、脊髄障害、筋ジストロフィー及びその他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎、遺伝性、代謝性、内分泌性、及び中毒性の筋疾患、重症筋無力症、周期性四肢麻痺、精神病(気分性、不安性の障害)、季節性障害(SAD)、静座不能、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、帯状疱疹後神経痛、進行性核上麻痺、大脳皮質基底核部変性症、及び家族性の前頭側頭性健忘症とが含まれ； 癌としては腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、および奇形癌など、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、および子宮の癌が含まれる。変異CHOP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して、CHOPをコードするポリヌクレオチド配列を、サザン法やノーザン法、ドットプロット法、或いはその他の膜系の技術、PCR法と、ディップスティック(dipstick)、ピン(pin)、およびマルチフォーマットのELISA様アッセイ、及びマイクロアレイに使用することが可能である。このような定性方法または定量方法は、当分野で公知である。

【0246】

10

20

30

40

50

或る特定の態様では、CHOPをコードするヌクレオチド配列群は、関連する障害、特に上記した障害を検出するアッセイ類において有用であり得る。CHOPをコードするヌクレオチド配列は標準法で標識化され、ハイブリダイゼーション複合体の形成に好適な条件下で、患者から採取した体液または組織のサンプルに添加することができる。好適なインキュベーション期間が経過したらサンプルを洗浄し、シグナルを定量して標準値と比較する。患者サンプルのシグナル量が制御サンプルと比べて著しく変化している場合は、サンプル内のCHOPをコードするヌクレオチド配列の変異レベルは関連する疾患の存在を示している。このようなアッセイは、動物実験、臨床試験における特定の治療効果を評価するため、ある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。

【0247】

CHOPの発現に関連する疾患の診断基準を提供するために、発現のための正常あるいは標準プロフィールを確立する。これは、ハイブリダイゼーション或いは増幅に好適な条件下、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、CHOPをコードする配列或いはその断片とを混合させることにより達成され得る。実質的に精製されたポリヌクレオチドを既知量用いて行った実験から得た値を正常な被験者から得た値と比較することにより、標準ハイブリダイゼーションを定量することができる。このようにして得た標準値は、疾患の徴候を示す患者から得たサンプルから得た値と比較することができる。標準値からの偏差を用いて疾患の存在を確定する。

【0248】

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定期的に繰り返し得る。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。

【0249】

癌に関しては、個体からの生体組織における異常な量の転写物（過少発現または過剰発現）の存在は、疾患の発生素質を示したり、実際に臨床的症狀が現れる前に疾患を検出する方法を提供し得る。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。

【0250】

CHOPをコードする配列群から設計されたオリゴヌクレオチド群の、更なる診断への利用には、PCRの利用が含まれ得る。これらのオリゴマーは、化学的に合成するか、酵素により生産するか、或いはin vitroで産出し得る。オリゴマーは、好ましくはCHOPをコードするポリヌクレオチドの断片、或いはCHOPをコードするポリヌクレオチドと相補的ポリヌクレオチドの断片を含み、最適条件下で特定の遺伝子や条件を識別するべく利用される。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。

【0251】

或る特定態様においては、CHOPをコードするポリヌクレオチド配列群に由来するオリゴヌクレオチドプライマー類を用いて、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、挿入および欠失である。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）がある。SSCPでは、CHOP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。差異は非変性ゲル中でのゲル電気泳動法を用いて検出可能である。fSSCPでは、オリゴヌクレオチドプライマーを蛍光性に標識する。それによってDNAシーケンシング機などの高処理機器でアンプリマー（amplimer）の検出が可能になる。更に、インシリコSNP（in silico SNP, isSNP）と呼ばれる配列デー

10

20

30

40

50

データベース分析法は、一般的なコンセンサス配列に配列されるような個々のオーバーラップするDNA断片の配列を比較することにより、多型性を同定し得る。これらのコンピュータベースの方法は、DNAの実験室での調整及び統計モデル及びDNA配列クロマトグラムの自動分析を用いたシーケンシングのエラーに起因する配列の変異をフィルタリングして除去する。別の態様では、例えば高処理MASSARRAYシステム (Sequenom, Inc., San Diego CA) を用いた質量分析によりSNPを検出し、特徴付ける。

【0252】

SNPはヒトの病気の遺伝的基礎を研究するために使える可能性がある。たとえば、少なくとも16個の一般的なSNPがインスリン非依存性の糖尿病と関連付けられている。SNPはまた、嚢胞性線維症、鎌形血球貧血、慢性肉芽腫性疾患等の単一遺伝子病の現れ方の違いを研究するために有用である。たとえば、マンノース結合レクチン (MBL2) の変異体は嚢胞性線維症の肺での有害な現れ方との相関が示されている。SNPはまた、薬理ゲノミックス (命のかかわる毒性など、患者の薬への反応に影響する遺伝的変異体の同定) にも役立つ。たとえば、Nアセチルトランスフェラーゼのある変異体は抗結核薬剤イソニアジドに対する高頻度の末梢神経障害と関連付けられているし、ALOX5のコアプロモーターのある変異体は5リポオキシゲナーゼ経路を標的とする抗喘息薬剤による治療への臨床的反応を弱くする。異なる集団におけるSNP分布の分析は、遺伝的浮動、突然変異、組換え、選択の研究において、また集団の起源と移動の追跡以外にも有用である (Taylor, J.G. 他 (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. および Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. 他 (2001) Curr. Opin. Neurobiol. 11:637-641)。

10

20

【0253】

CHOPの発現を定量するために用いられ得る方法には、ヌクレオチドの放射標識或いはビオチン標識、調節核酸の相互増幅 (coamplification)、及び標準的な曲線に結果が加えられたものが含まれる (Melby, P.C. 他 (1993) J. Immunol. Methods, 159:235-44; Duplaa, C. 他 (1993) Anal. Biochem. 229:236)。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。

【0254】

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

30

【0255】

別の実施例では、CHOP、CHOPの断片、CHOP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。

40

【0256】

或る実施態様は、或る組織または細胞タイプの転写イメージを作製する、本発明のポリヌクレオチドの使用に関連する。転写イメージは、特定の組織または細胞タイプにより遺伝子発現の包括的パターンを表す。包括的遺伝子発現パターンは、所与の条件下で所与の時間に発現した遺伝子の数および相対存在量を定量することにより分析し得る (Seilhamer 他 米国特許第5,840,484号「Comparative Gene Transcript Analysis」は特に引用する

50

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。

【0257】

転写イメージは、組織、細胞株、生検またはその生体サンプルから単離した転写物を用いて作製し得る。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。

10

【0258】

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と併せて使用し得る。全ての化合物は、作用および毒性のメカニズムを標示し、しばしば分子フィンガープリントまたは毒性シグネチャと称される、特徴的な遺伝子発現パターンを惹起する (Nuwaysir, E.F. 他 (1999) *Mol. Carcinog.* 24:153-159、Steiner, S. および N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471、該文献は特に引用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、多数の遺伝子およ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、最も有用且つ正確である。理想的には、ゲノム全域にわたる発現の測定が、最高品質のシグネチャを提供する。たとえ、発現が任意の試験された化合物によって変化しない遺伝子があったとしても、それらの発現レベルを残りの発現データを標準化するために使用できるため、それらの遺伝子は重要である。標準化手順は、異なる化合物で処理した後の発現データの比較に有用である。毒性シグネチャの要素への遺伝子機能を割り当てることは毒性機構の解明に役立つが、毒性の予測につながるシグネチャの統計的な一致には遺伝子機能の知識は必要ではない (2000年2月29日に米国国立環境健康科学研究所 (National Institute of Environmental Health Sciences) より発行された Press Release 00-02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。

20

30

【0259】

或る実施例では、核酸を有する生体サンプルを試験化合物で処理することにより、この試験化合物の毒性を算定する。処理した生物学的サンプル中で発現した核酸は、本発明のポリヌクレオチドに特異的な1つ若しくは複数のプローブでハイブリダイズし、それによって本発明のポリヌクレオチドに対応する転写レベルを定量し得る。処理した生物学的サンプル中の転写レベルを、未処理生物学的サンプル中のレベルと比較する。両サンプルの転写レベルの差は、処理されたサンプル中で試験化合物が引き起こす毒性反応を示す。

【0260】

別の実施態様は、本発明のポリペプチド配列群を用いて或る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銀染色液または蛍光染色液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

40

50

、独自の位置にある点としてゲル中で可視化される。各タンパク質スポットの光学密度は、通常サンプル中のタンパク質レベルに比例する。異なるサンプル、例えば試験化合物または治療薬で処理または未処理のいずれかの生物学的サンプルから得られるタンパク質スポットの光学密度を比較し、処理に関連するタンパク質スポット密度の変化を同定する。スポット内のタンパク質は、例えば化学的または酵素的切断とそれに続く質量分析を用いる標準的な方法を用いて部分的にシーケンシングする。スポット内のタンパク質の同一性は、その部分配列を、好適には少なくとも5個の連続するアミノ酸残基を、本発明のポリペプチド配列と比較することにより決定し得る。場合によっては、決定的なタンパク質同定のための更なる配列が得られる。

【0261】

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

【0262】

プロテオームレベルでの毒性シグネチャも中毒学的スクリーニングに有用であり、転写レベルでの毒性シグネチャと並行に分析するべきである。或る組織の或るタンパク質に対しては、転写とタンパク質の存在量の相関が乏しいこともあるので (Anderson, N.L. 及び J. Seilhamer (1997) *Electrophoresis* 18:533-537)、転写イメージにはそれ程影響しないがプロテオームのプロフィールを変化させるような化合物の分析においてプロテオーム毒性シグネチャは有用たり得る。更に、体液中の転写物の分析はmRNAの急速な分解のために困難なので、プロテオームのプロフィール作成はこのような場合により信頼し得、情報価値があり得る。

【0263】

別の実施例では、タンパク質を含有する生体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、未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。

【0264】

別の実施例では、タンパク質を含有する生体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、未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。

【0265】

マイクロアレイは、本技術分野でよく知られている方法を用いて調製し、使用し、そして分析する (Brennan, T.M. 他 (1995) の米国特許第5,474,796号、Schena, M. 他 (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619, Baldeschweiler 他 (1995) PCT出願第W095/251116号、Shalon, D. 他 (1995) PCT出願第W095/35505号、Heller, R.A. 他 (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155, Heller, M.J. 他 (1997) 米国特許第5,605,662号)。様々なタイプのマイクロアレイが公知であり、詳細については、DNA Microarrays: A Practical Approach, M. Schena, 編集. (1999) Oxford University Press, L

10

20

30

40

50

ondonに記載されている。該文献は、特に引用することを以って本明細書の一部となす。

【0266】

本発明の別の実施態様ではまた、CHOPをコードする核酸配列群を用いて、天然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群を作製し得る。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、コード配列より非コード配列の方が好ましい。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他（1997）Nat Genet. 15:345-355、Price, C.M. (1993) Blood Rev. 7:127-134、Trask, B.J. (1991) Trends Genet. 7:149-154）。一度マッピングすると、本発明の核酸配列を用いて、例えば病状の遺伝と特定の染色体領域やまたは制限酵素断片長多型（RFLP）の遺伝とが関連するような遺伝子連鎖地図を作成可能である（Lander, E.S. 及び D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357）。

10

【0267】

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（Heinz-Ulrich, 他（1995）前出の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CHOPをコードする遺伝子の位置と、特定の疾患との相関性、あるいは特定の疾患に対する素因との相関性は、この疾患と関連するDNAの領域の決定に役立ち得るため、ポジショナルクローニングの作業を促進し得る。

20

【0268】

確定した染色体マーカーを用いた連鎖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、正確な染色体上の遺伝子座が未知でも、関連するマーカー類をしばしば明らかにし得る。この情報は、ポジショナルクローニングその他の遺伝子発見技術を用いて遺伝的疾患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、該領域に対する任意のマッピングにより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.他（1988）Nature 336:577-580）。転座、反転などに起因する、健常者、保有者、罹病者の三者間における染色体位置の相違を検出する場合にも、本発明のヌクレオチド配列を用い得る。

30

【0269】

本発明の別の実施態様では、CHOP、その触媒作用断片あるいは免疫原性断片またはそのオリゴペプチド群を、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群のスクリーニングに用い得る。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。CHOP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

【0270】

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処理能力でスクリーニングするために用いられる（Geysen, 他（1984）PCT出願番号 W084/03564）。この方法においては、多数の異なる小さな試験用化合物を固体基板上で合成する。試験用化合物は、CHOP、或いはその断片と反応してから洗浄される。次に、結合したCHOPを本技術分野で周知の方法で検出する。精製したCHOPはまた、上記した薬剤スクリーニング技術に用いるプレート上に直接コーティングできる。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。

【0271】

別の実施例では、CHOPと特異結合可能な中和抗体がCHOPとの結合について試験用化合物と

50

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、1若しくは数個の抗原決定因子をCHOPと共有するペプチドの存在を検出する。

【0272】

別の実施例では、CHOPをコードするヌクレオチド配列を、将来に開発される分子生物学技術であり、現在知られているヌクレオチド配列の特性（限定はされないが、トリプレット遺伝コード、特異的な塩基対相互作用等を含む）に依存する新技術に用い得る。

【0273】

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。

10

【0274】

本明細書において開示した全ての特許、特許出願及び刊行物、特に米国特許第60/293,768号、第60/309,548号、第60/314,400号、第60/343,706号及び第60/337,999号は、言及することをもって本明細書の一部となす。

【実施例】

【0275】

1 cDNAライブラリの作製

Incyte cDNA群の由来は、LIFESEQ GOLDデータベース（Incyte Genomics, Palo Alto CA）に記載されたcDNAライブラリ群である。幾つかの組織はホモジナイズしてグアニジニウムイソチオシアネート溶液に溶解し、他の組織はホモジナイズしてフェノールにまたは変性剤群の好適な混合液に溶解した。混合液の1例であるTRIZOL（Invitrogen）は、フェノールとグアニジンイソチオシアネートとの単相溶液である。結果として得られた溶解物は、塩化セシウム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。

20

【0276】

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、DNA分解酵素でRNAを処理した。殆どのライブラリでは、オリゴd(T)連結常磁性粒子（Promega）、OLIGOTEXラテックス粒子（QIAGEN, Chatsworth CA）またはOLIGOTEX mRNA精製キット（QIAGEN）を用いて、ポリ(A+) RNAを単離した。別法では、別のRNA単離キット、例えばPOLY(A)PURE mRNA精製キット（Ambion, Austin TX）を用いて組織溶解物からRNAを直接単離した。

30

【0277】

場合によってはStratagene社へのRNA提供を行い、対応するcDNAライブラリをStratagene社が作製することもあった。そうでない場合は、UNIZAPベクターシステム（Stratagene）またはSUPERSCRIPトラスミドシステム（Invitrogen）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, 5.1-6.6ユニット）。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（300～1000 bp）選択は、SEPHACRYL S1000、SEPHAROSE CL2BまたはSEPHAROSE CL4Bカラムクロマトグラフィー（Amersham Biosciences）、或いは調製用アガロースゲル電気泳動法を用いて行った。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（Invitrogen）、PCDNA2.1 プラスミド（Invitrogen, Carlsbad CA）、PBK-CMV プラスミド（Stratagene）、PCR2-TOPOTAプラスミド（Invitrogen）、PCMV-ICISプラスミド（Stratagene）、pIGEN（Incyte Genomics, Palo Alto CA）、pRARE（Incyte Genomics）、pINCY（Incyte Genomics）等およびその誘導体である。組換えプラスミドは、Stratagene社のXL1-Blue、XL1-BlueMRFまたはSOLR、あるいはInvitrogen社のDH5α、DH10BまたはElectroMAX DH10Bなど適格な大腸菌細胞に形質転換した。

40

50

【0278】

2cDNAクローンの単離

UNIZAPベクターシステム (Stratagene) を用いた *in vivo* 切除によって、或いは細胞溶解によって、実施例1のようにして得たプラスミドを宿主細胞から回収した。プラスミドの精製には、下記の少なくとも1つを用いた。すなわちMagicまたはWIZARD Minipreps DNA精製システム (Promega)、AGTC Miniprep精製キット (Edge Biosystems, Gaithersburg MD)、QIAGEN社のQIAWELL 8 Plasmid、QIAWELL 8 Plus PlasmidおよびQIAWELL 8 Ultra Plasmid精製システム、R.E.A.L. Prep 96プラスミド精製キットのいずれかである。プラスミドは、沈殿させた後、0.1mlの蒸留水に再懸濁して、凍結乾燥して或いは凍結乾燥しないで4℃で保管した。

10

【0279】

別法では、高処理フォーマットにおいて直接結合PCR法を用いて宿主細胞溶解物からプラスミドDNAを増幅した (Rao, V.B. (1994) Anal. Biochem. 216:1-14)。宿主細胞の溶解および熱サイクリング過程は、単一反応混合液中で行った。サンプルを処理し、それを384ウェルプレート内で保管し、増幅したプラスミドDNAの濃度をPICOGREEN色素 (Molecular Probes, Eugene OR) 及びFLUOROSKANII蛍光スキャナ (Labsystems Oy, Helsinki, Finland) を用いて蛍光分析的に定量した。

【0280】

3シーケンシング及び分析

実施例2に記載したようにプラスミドから回収したIncyte cDNAを、以下に示すようにシーケンシングした。cDNAのシーケンス反応は、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のシーケンス反応は、Amersham Biosciences 社が提供する試薬、またはABIシーケンシングキット、例えばABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems) の試薬を用いて準備した。cDNAのシーケ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シーケンシングシステム (Applied Biosystems) か、標準ABIプロトコル及び塩基呼び出し (base calling) ソフトウェアを用いるABI PRISM 373または377シーケンシングシステム (Applied Biosystems) か、或いはその他の本技術分野で既知の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, 7.7ユニットに概説) を用いて決定した。cDNA配列の幾つかを選択して、実施例8に記載した方法で配列を伸長させた。

20

30

【0281】

Incyte 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)配列を除去し、あいまいな塩基対をマスクすることによって有効性を確認した。その際、BLAST、動的プログラミング及び隣接ジヌクレオチド頻度分析に基づくアルゴリズム及びプログラムを用いた。次に、IncyteのcDNA配列、またはその翻訳を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) と、ヒト、ラット、マウス、線虫、出芽酵母 (*Saccharomyces cerevisiae*)、分裂酵母 (*Schizosaccharomyces pombe*)、および驚口瘡カンジダ (*Candida albicans*) からの配列を含むPROTEOMEデータベース (Incyte Genomics, Palo Alto CA)、及びPFAM等隠れマルコフモデル (HMM) に基づいたタンパク質ファミリーデータベース並びに、SMART (Schultz, J. 他 (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864, Letunic, I. 他 (2002) Nucleic Acids Res. 30:242-244) のようなHMMに基づいたタンパク質ドメインデータベースから選択したデータベースに対して問い合わせた。(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) Cuff. Opin. Struct. Biol. 6:361-365等を参照)。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配列を構築

40

50

し、完全長のポリヌクレオチド配列を産出した。あるいは、GenBank cDNA群、GenBank EST群、スティッチされた配列群、ストレッチされた配列群、またはGenscan予測コード配列群（実施例4および5を参照）を用い、Incyte cDNAの集団を完全長まで伸長させた。PhredとPhrapとConsedとに基づくプログラムを用いて構築し、GenMarkとBLASTとFASTAとに基づくプログラムを用いて、cDNAの集団を、オープンリーディングフレームについてスクリーニングした。完全長ポリヌクレオチド配列を翻訳し、対応する完全長ポリペプチド配列を引き出した。あるいは、本発明のポリペプチドは、完全長翻訳ポリペプチドの任意のメチオニン残基で開始し得る。完全長ポリペプチド配列群の続いての分析としての問い合わせを、GenBankタンパク質データベース群（genpept）、SwissProt、PROTEOMEデータベース群、BLOCKS、PRINTS、DOMO、PRODOMおよびProsites等のデータベースや、PFAM、INCY、およびTIGRFAM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群、並びにSMART等のHMMベースのタンパク質ドメインデータベース群に対し行った。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）およびLASERGENEソフトウェア（DNASTAR）を用いて分析する。ポリヌクレオチド及びポリペプチド配列アラインメントは、アラインメントした配列と配列の一致率も計算するMEGALIGNマルチシーケンスアラインメントプログラム（DNASTAR）に組み込まれているようなCLUSTALアルゴリズムによって特定されるデフォルトパラメータを用いて作製する。

10

【0282】

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

20

【0283】

完全長ポリヌクレオチド配列およびポリペプチド配列の組み立て及び分析に用いる上記のプログラムは、SEQ ID NO:11-20のポリヌクレオチド配列断片の同定にも利用できる。ハイブリダイゼーション及び増幅技術に有用である約20～約4000ヌクレオチドの断片を表4の列2に示した。

30

【0284】

4 ゲノムDNAからのコード配列の同定及び編集

推定上の糖結合タンパク質は、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、様々な生物からのゲノムDNA配列を分析する汎用遺伝子同定プログラムである（Burge, C. 及び S. Karlin (1997) J. Mol. Biol. 268:78-94、Burge, C. 及び S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354）。プログラムは予測エキソンを連結し、メチオニンから終止コドンに及ぶ構築されたcDNA配列を形成する。Genscanの出力は、ポリヌクレオチドおよびポリペプチド配列のFASTAデータベースである。Genscanが一度に分析する配列の最大範囲は、30kbに設定した。これらのGenscan推定cDNA配列の内、どの配列が糖結合タンパク質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モデルにおいて糖結合タンパク質について問合せて分析した。潜在的な糖結合タンパク質はまた、糖結合タンパク質として注釈が付けられていたIncyte cDNA配列への相同性を基に同定された。こうして選択されたGenscan予測配列は、次にBLAST分析により公共データベースgenpept及びgbpriと比較した。必要であれば、genpeptからのトップBLASTヒットと比較することによりGenscan予測配列を編集し、余分なまたは省略されたエキソンなど、Genscanが予測した配列におけるエラーを補正した。BLAST分析はまた、Genscan予測配列の、いかなるIncyte cDNAまたは公共cDNAカバレッジ（coverage）の発見にも用いられ、したがって転写の証拠を提供した。Incyte cDNAカバレッジが利用できた場合には、この情報を用

40

50

いてGenscan予測配列を補正または確認した。完全長ポリヌクレオチド配列は、実施例 3に記載した構築プロセスを用いて、Incyte cDNA配列および/または公共cDNA配列でGenscan予測コード配列を構築して得た。或いは、完全長ポリヌクレオチド配列は編集または未編集のGenscan予測コード配列に完全に由来する。

【0285】

5 cDNA配列データを使ったゲノム配列データのアセンブリ

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シング変異体を生み出した。区間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。結果として得たステッチ配列群を翻訳し、BLAST分析で公共データベースgenpeptおよびgbpriと比較した。Genscanにより予測された不正確なエキソンは、genpeptからのトップのBLASTヒットと比較することにより修正した。必要な場合には、追加cDNA配列を用いるかゲノムDNAの検査により配列を更に伸長させた。

10

20

30

40

50

【0286】

ストレッチ配列 (Stretched Sequence)

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

【0287】

6 CHOP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:11-20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:11-20と一致するこれらのデータベースの配列を、Phrapなどの構築アルゴリズム (表7) を使用して、連続しオーバーラップする配列のクラスター群に組み入れた。スタンフォード・ヒトゲノムセンター (SHGC)、ホワイットヘッド・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決定した。マッピングされた配列がクラスタに含まれている場合は、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。

【0288】

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、染色体の短腕(p-arm)の末端に関連して測定する。(センチモルガン(cM)は、染色体マーカー間の組換え頻度に基づく計測単位である。平均して、1 cMは、ヒト中のDNAの1メガベース(Mb)にほぼ等しい。尤も、この値は、組換えのホットスポット及びコールドスポットに起因して広範囲に変化する。) cM距離は、各クラス内に配列が含まれる放射線ハイブリッドマーカー類に対して境界を提供するGenethonによってマッピングされた遺伝マーカー群に基づく。NCBI「GeneMap'99」(<http://www.ncbi.nlm.nih.gov/genemap/>)などの公的に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、既に同定されている疾患遺伝子群が、上記した区間内若しくは近傍に位置するかを決定できる。

10

【0289】

7. ポリヌクレオチド発現の分析

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

【0290】

BLASTを適用した類似のコンピュータ技術を用いて、GenBankやLifeSeq (Incyte Genomics) 等のcDNAデータベースにおいて同一または関連分子を検索した。ノーザン分析は、多数の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

20

同的なものとして分類するか否かを決定するため、コンピュータ検索の感度を変更することができる。検索の基準は積スコアであり、次式で定義される。

(BLASTスコア $\times$ 配列一致率)5 $\times$ {長さ(配列1)、長さ(配列2)}の最小値

積スコアは、2つの配列間の類似度と、配列が一致する長さとの両方を考慮している。積スコアは、0 ~ 100の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。BLASTスコアを計算するには、或る高スコアリングセグメント対(HSP)内の一致する各塩基に+5のスコアを割り当て、各不一致塩基に-4を割り当てる。2つの配列は、2以上のHSPを共有し得る(ギャップにより隔離される)。2以上のHSPがある場合には、最高BLASTスコアの塩基対を用いて積スコアを計算する。積スコアは、断片的オーバーラップとBLASTアラインメントの質とのバランスを表す。例えば積スコア100は、比較した2つの配列の短い方の長さ全体にわたって100%一致する場合のみ得られる。積スコア70は、一端が100%一致し、70%オーバーラップしているか、他端が88%一致し、100%オーバーラップしているかのいずれかの場合に得られる。積スコア50は、一端が100%一致し、50%オーバーラップしているか、79%一致し、100%オーバーラップしているかのいずれかの場合に得られる。

30

【0291】

あるいは、CHOPをコードするポリヌクレオチド配列を、由来する組織源について分析する。例えば幾つかの完全長配列は、少なくとも一部は、オーバーラップするIncyte cDNA配列群を用いて構築される(実施例3を参照)。各cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。各cDNA配列は、ヒト組織から作製されたcDNAライブラリに由来する。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、細胞株、発達、炎症、神経性、外傷、心血管、プール、その他の1つに分類される。各カ

40

50

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。演算結果の割合は、CHOPをコードするcDNAの組織特異発現または疾患特異発現を反映する。cDNA配列及びcDNAライブラリ/組織情報については、LIFESEQ GOLDデータベース (Incyte Genomics, Palo Alto CA) で見られる。

【0292】

8 ポリヌクレオチドをコードするCHOPの伸長

完全長のポリヌクレオチド配列はまた、完全長分子の適切な断片から設計したオリゴヌクレオチドプライマー群を用いて該断片を伸長させて産生した。或るプライマーは既知の断片の5'伸長を開始するべく合成し、別のプライマーは既知の断片の3'伸長を開始するべく合成した。開始プライマーは、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。

【0293】

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

【0294】

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 Biosciences)、ELONGASE酵素 (Invitrogen)、Pfu DNAポリメラーゼ (Stratagene)を含む。プライマーの組、PCI AとPCI Bに対して以下のパラメータで増幅を行った。ステップ1: 94 で3分間、ステップ2: 94 で15秒間、ステップ3: 60 で1分間、ステップ4: 68 で2分間、ステップ5: ステップ2、3および4を20回繰り返す、ステップ6: 68 で5分間、ステップ7: 4 で保存。別法では、プライマー対であるT7とSK+と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 , 3分、ステップ2: 94 , 15秒、ステップ3: 57 , 1分、ステップ4: 68 , 2分、ステップ5: ステップ2、3および4を20回反復する。ステップ6: 68 、5分、ステップ7: 4 で保存する。

【0295】

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

【0296】

伸長させたヌクレオチドは、脱塩及び濃縮して384ウェルプレートに移し、CviJIコレイ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、pUC 18ベクター (Amersham Biosciences)への再連結反応前に音波処理またはせん断した。ショットガン・シー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pUC 18ベクター (Amersham Biosciences)に再連結し、Pfu DNAポリメラーゼ (Stratagene)で処理して制限部位のオーバーハングを満たし、大腸菌細胞に形質移入した。形質移入した細胞を選択して抗生物質を含む培地に移し、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で一晩培養した。

【0297】

細胞を溶解して、Taq DNAポリメラーゼ(Amersham Biosciences)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 , 3分、ステップ2: 94 、 15 秒、ステップ3: 60 、 1 分、ステップ4: 72 、 2 分、ステップ5: ステップ2、3および4を29回反復する。 ステップ6: 72 、 5 分、ステップ7: 4 で保存する。上記のようにPICOGREEN試薬(Molecular Probes)でDNAを定量した。DNAの回収率が低いサンプルは、上記と同一の条件を用いて再増幅した。サンプルは20%ジメチルスルホキシド(1:2, v/v)で希釈し、DYENAMIC エネルギートランスファー シークエンシングプライマー、及びDYENAMIC DIRECT kit(Amersham Biosciences)またはABI PRISM BIGDYE ターミネーターサイクル シークエンシング反応キット(Terminator cycle sequencing ready reaction kit)(Applied Biosystems)を用いてシークエンシングした。

【0298】

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

【0299】

9 CHOP がコードするポリヌクレオチドのSNP(一塩基多型)の同定

一塩基多型(SNP)として知られる一般的なDNA配列変異体がLIFESEQ データベース(Incyte Genomics)を用いてSEQ ID NO:11-20 において同定された。実施例3に記述したように、同じ遺伝子からの配列を一緒にまとめてクラスター化し、構築した。一連のフィルタから成るアルゴリズムが、SNPを他の配列変異体から区別するために用いられた。予備フィルターは最小Phredクオリティスコア15を要求することによって大多数のベースコールエラーを除去し、配列アラインメントエラーとスプライス変異体、キメラおよびベクター配列の不適正なトリミングによるエラーを除去した。高度染色体解析の自動化手順により、オリジナルのクロマトグラムファイルの中の推定上のSNPの近傍を解析した。クローンエラーフィルターは、統計的に生成されたアルゴリズムを使って、実験室の処理中に入ってくる(逆転写酵素、ポリメラーゼまたは体細胞性突然変異などに起因する)エラーを同定した。クラスタリングエラーフィルターは統計的に生成されたアルゴリズムを使って、近い相同体や偽遺伝子のクラスタリングによるエラーおよび非ヒト配列によるコンタミネーションに起因するエラーを同定した。最後のフィルターセットは免疫グロブリンまたはT細胞受容体で見つかる複製とSNPを除去した。

【0300】

4つの異なるヒトの集団における対立遺伝子頻度を解析するために、ある種のSNPを選択して、高スループットMASSARRAYシステム(Sequenom, Inc.)を使った質量分析で特性を測定した。白人集団は92人(男性46人、女性46人)で、83人がユタ州出身、4人がフランス人、3人がベネズエラ、そして2人がアーミッシュだった。アフリカ人は194人(男性97人、女性97人)からなり、その全てがアフリカ系アメリカ人であった。ラテンアメリカ系集団は324人(男性162人、女性162人)からなり、その全てがメキシコのラテンアメリカ系であった。アジア系集団は126人(男性64人、女性62人)からなり、報告された親の内訳は中国人が43%、日本人が31%、韓国人が13%、ベトナム人が5%、その他のアジア系が8%であった。対立遺伝子頻度は最初に白人集団で解析された。この集団内で対立遺伝子変異を示さなかったSNPはその他3つの集団ではテストしない場合もあった。

【0301】

10 個々のハイブリダイゼーションプローブの標識化及び使用

SEQ ID NO:11-20から導き出されたハイブリダイゼーションプローブを用いて、cDNA、mRNA、またはゲノムDNAをスクリーニングする。約20塩基対からなるオリゴヌクレオチドの標識について特に記載するが、より大きなヌクレオチド断片に対しても本質的に同一の手順が用いられる。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、 $[\gamma\text{-}^{32}\text{P}]$ アデノシン 3 リン酸（Amersham Biosciences）250 μCi と、T4ポリヌクレオチドキナーゼ（DuPont NEN, Boston MA）を混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、SE PHADEX G-25超細繊維分子サイズ排除デキストラン ピードカラム（Amersham Biosciences）を用いて実質的に精製する。下記のいずれか 1 つのエンドヌクレアーゼで消化されたヒトゲノムDNAの、典型的な膜ベースのハイブリダイゼーション解析において、毎分 10^7 カウントの標識されたプローブを含むアリコットを用いる。すなわちAse I、Bgl II、Eco RI、Pst I、Xba I、またはPvu II (DuPont NEN)である。

【0302】

各消化物から得たDNAは、0.7%アガロースゲル上で分画してナイロン膜（Nytran Plus, Schleicher & Schuell, Durham NH）に移す。ハイブリダイゼーションは、40 で16時間行う。非特異的シグナルを除去するため、例えば0.1×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。

【0303】

1.1 マイクロアレイ

マイクロアレイの表面上でアレイエレメントの結合または合成は、フォトリソグラフィ、ピエゾ式印刷（インクジェット印刷、前出のBaldeschweiler他等を参照）、機械的マイクロスポッティング技術及びこれらから派生したものをを用いて達成することが可能である。上記各技術において基板は、均一且つ無孔の固体とするべきである（Schna、前出）。推奨する基板には、シリコン、シリカ、スライドガラス、ガラスチップ及びシリコンウエハがある。或いは、ドットプロット法またはスロットプロット法に類似のアレイを利用して、熱的、紫外線的、化学的または機械的結合手順を用いて基板の表面にエレメントを配置及び結合させてもよい。通常のアレイは、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（Schna, M. 他（1995）Science 270:467-470、Shalon, D. 他（1996）Genome Res. 6:639-645、Marshall, A. および J. H odgson（1998）Nat. Biotechnol. 16:27-31）。

【0304】

完全長cDNA、発現配列タグ（EST）、またはその断片またはオリゴマーは、マイクロアレイのエレメントと成り得る。ハイブリダイゼーションに好適な断片またはオリゴマーを、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。アレイエレメント群を、生体サンプル中のポリヌクレオチド群とハイブリダイズする。生体サンプル中のポリヌクレオチドは、検出を容易にするために蛍光標識などの分子タグに抱合させる。ハイブリダイゼーション後、生体サンプルからのハイブリダイズされていないヌクレオチドを除去し、蛍光スキャナを用いて各アレイエレメントでのハイブリダイゼーションを検出する。あるいは、レーザー脱離および質量スペクトロメトリを用いてもハイブリダイゼーションを検出し得る。マイクロアレイ上の或るエレメントにハイブリダイズする各ポリヌクレオチドの、相補性の度合と相対存在度とを算定し得る。一実施態様におけるマイクロアレイの調製および使用について、以下に詳述する。

【0305】

組織または細胞サンプルの準備

グアニジニウムチオシアネート法を用いて組織サンプルから全RNAを単離し、オリゴ(dT)セルロース法を用いてポリ(A)⁺RNAを精製する。各ポリ(A)⁺RNAサンプルを、MLLV逆転写酵素、0.05pg/ μl のオリゴ(dT)プライマー（21mer）、1×第一鎖合成バッファー、0.03unit/ 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 Biosciences）を用いて逆転写する。逆転写反応は、GEMBRIGHTキット（Incyte）を用いてポリ(A)⁺RNA含有の25体積ml内で行う。特異的対照ポリ(A)⁺RNAは、非コード酵母ゲノムDNAからin vitro転写により合成する。

各反応サンプル（１つはCy3、もう１つはCy5標識）は、2.5 mlの0.5 M水酸化ナトリウムで処理し、85℃で20分間インキュベートし、反応を停止させてRNAを分解させる。サンプルは、2つの連続するCHROMA SPIN 30ゲル濾過スピンカラム（CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA）を用いて精製する。混合後、2つの反応サンプルは、1 mlのグリコーゲン（1 mg/ml）、60 mlの酢酸ナトリウム及び300 mlの100%エタノールを用いてエタノール沈殿させる。サンプルは次に、SpeedVAC（Savant Instruments 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。

【0306】

マイクロアレイの準備

本発明の配列を用いて、アレイエレメントを作製する。各アレイエレメントは、クローン化cDNA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インサートに隣接するベクター配列に相補的なプライマーを用いる。30サイクルのPCRによって、1～2 ngの初期量から5 µgを超える最終量までアレイエレメントを増幅する。増幅したアレイエレメントは、SEPHACRYL-400（Amersham Biosciences）を用いて精製する。

10

【0307】

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処理をかけ、蒸留水で非常に良く洗浄する。スライドガラスは、4%フッ化水素酸（VWR Scientific Products Corporation (VWR), West Chester PA）中でエッチングし、蒸留水中で非常に良く洗浄し、95%エタノール中で0.05%アミノプロピルシラン（Sigma）を用いてコーティングする。コーティングしたスライドガラスは、110℃のオーブンで硬化させる。

20

【0308】

米国特許第5,807,522号で説明されている方法を用いて、コーティングしたガラス基板にアレイエレメントを付加する。該特許は、引用を以って本明細書の一部となす。平均濃度が100 ng/µlのアレイエレメントDNA 1 µlを高速機械装置により開放型キャピラリープリンティングエレメント（open capillary printing element）に充填する。装置はここで、スライド毎に約5 nlのアレイエレメントサンプルを加える。

30

【0309】

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。

【0310】

ハイブリダイゼーション

ハイブリダイゼーション反応液は、5 × SSC、0.2% SDSハイブリダイゼーション緩衝液にCy3及びCy5標識したcDNA合成産物を各0.2 µg含む9 µlのサンプル混合体を含めたものである。サンプル混合体は、65℃まで5分間加熱し、マイクロアレイ表面上に等分して1.8 cm²のカバーガラスで覆う。アレイを、顕微鏡用スライドよりわずかに大きい空洞を有する防水チェンバーに移す。チェンバーのコーナーに140 µlの5 × SSCを加えることにより、チェンバー内部を湿度100%に保持する。アレイを含むチェンバーは、60℃で約6.5時間インキュベートする。アレイは、第1洗浄緩衝液中（1 × SSC, 0.1% SDS）において45℃で10分間洗浄し、第2洗浄緩衝液中（0.1 × SSC）において45℃で10分間各々3度洗浄して乾燥させる。

40

【0311】

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488 nm、C

50

y 5の励起のためには632nmでスペクトル線を発生し得るInnova70混合ガス10Wレーザー (Coherent, Inc., Santa Clara CA)を備えた顕微鏡で検出する。励起レーザー光の焦点をアレイ上に置くため、20×顕微鏡対物レンズ (Nikon, Inc., Melville NY)を用いる。アレイを含むスライドを、顕微鏡の、コンピュータ制御のX-Yステージに置き、対物レンズを通してラスタースキャンする。本実施例で用いる1.8cm×1.8cmのアレイは、解像度20μmでスキャンする。

【0312】

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

10

【0313】

スキャンの感度は通常、既知濃度でサンプル混合体に添加されるcDNA対照種により発生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重量比1:100,000でハイブリダイゼーション種と相関させる。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、2つの蛍光色素で較正するcDNAのサンプルを標識し、ハイブリダイゼーション混合液に各々等量を加えることによって較正を行う。

20

【0314】

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および測定する場合には、各蛍光体の発光スペクトルを用いて、データはまず蛍光色素間の光学的クロストーク (発光スペクトルの重なりに起因する)を補正する。

30

【0315】

グリッドが蛍光シグナルイメージ上に重ねられ、それによって各スポットからのシグナルはグリッドの各エレメントに集められる。各エレメント内の蛍光シグナルは統合され、シグナルの平均強度に応じた数値が得られる。シグナル分析に用いるソフトウェアは、GEMT00LS遺伝子発現分析プログラム (Incyte) である。GEMT00LSプログラム (Incyte Genomics) を使って、発現が少なくとも2倍変化したアレイエレメント、SB比が少なくとも2.5であるもの、およびエレメントスポットサイズが少なくとも40%であるものは差次的に発現していると同定された。

40

【0316】

発現

たとえば、SEQ ID NO:19 は、大腸癌患者の大腸組織で、同じドナーのマッチした微視的正常組織と比較して、差次的発現を示したが、これはマイクロアレイ分析で判定した。SEQ ID NO:19の発現は癌の大腸癌組織において少なくとも2倍減少していた。したがって、SEQ ID NO:19は、大腸癌を含む細胞増殖性障害の段階決定および診断アッセイに有用である。

【0317】

12 相補的ポリヌクレオチド

50

CHOPをコードする配列に、或いはその任意の一部に対して相補的な配列は、天然CHOPの発現を、検出し、低下させ、または阻害するために用いられる。約15～30塩基対を含むオリゴヌクレオチドの使用について記すが、これより小さなあるいは大きな配列の断片の場合でも、本質的に同じ手順を用いる。適切なオリゴヌクレオチドを設計するため、Oligo 4.06ソフトウェア (National Biosciences) およびCHOPのコーディング配列を用いる。転写を阻害するためには、最も独特な5'配列から相補的オリゴヌクレオチドを設計し、これを用いて、プロモーターがコーディング配列に結合するのを防止する。翻訳を阻害するためには、相補的なオリゴヌクレオチドを設計して、リボソームがCHOPをコードする転写物に結合するのを阻害する。

【0318】

10

1.3 CHOPの発現

CHOP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CHOPが発現するために、抗生物質耐性遺伝子及び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としては、lacオペレーター調節エレメントと併用するT5またはT7バクテリオファージプロモーター、および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β-Dチオガラクトピラノシド (IPTG) で誘発されるとCHOPを発現する。真核細胞でのCHOP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* 核多角体病ウイルス (AcMNPV) を感染させて行う。バキュロウイルスの非必須ポリヘドリン遺伝子をCHOPをコードするcDNAと置換するには、相同組換えを行うか、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を行う。ウイルスの感染力は維持され、強力なポリヘドリンプロモーターによって高レベルのcDNA転写が行われる。組換えバキュロウイルスは、多くの場合は夜蛾の1種 *Spodoptera frugiperda* (Sf9) 昆虫細胞への感染に用いるが、ヒト肝細胞への感染に用いることもある。後者の感染の場合は、バキュロウイルスの更なる遺伝的変更が必要になる。(前出のEngelhard 他、前出のSandig 他)

20

【0319】

殆どの発現系では、CHOPが、例えばグルタチオンSトランスフェラーゼ (GST)、またはFLAGや6-Hisなどのペプチドエピトープ標識で合成された融合タンパク質となるため、未精製の細胞溶解物からの組換え融合タンパク質の親和性ベースの精製が素早く1回で行うことができる。GSTは日本住血吸虫からの26kDaの酵素であり、タンパク質の活性および抗原性を維持した状態で、固定化したグルタチオン上での融合タンパク質の精製を可能とする (Amersham Biosciences)。精製の後、GST部分を特定の操作部位でCHOPからタンパク質的に切断できる。FLAGは8アミノ酸のペプチドであり、市販されているモノクローナル及びポリクローナル抗FLAG抗体 (Eastman Kodak) を用いて免疫親和性精製を可能にする。6ヒスチジン残基が連続して伸長した6-Hisは、金属キレート樹脂上での精製を可能にする (QIAGEN)。タンパク質の発現および精製の方法は、前出のAusubel他(1995)10章、16章に記載されている。これらの方法で得た精製CHOPを直接用いて以下の実施例17、18、および19の、適用可能なアッセイを行うことができる。

30

40

【0320】

1.4 機能的アッセイ

CHOP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CHOPをコードする配列の発現によって評価する。cDNAを高いレベルで発現する強いプロモーターを含む哺乳動物発現ベクターにcDNAをサブクローニングする。選択されるベクターには、PCMV SPORTプラスミド (Invitrogen, Carlsbad CA) 及びPCR 3.1プラスミド (Invitrogen) が含まれ、どちらもサイトメガロウイルスプロモーターを有する。リボソーム製剤あるいは電気穿孔法を用いて、5～10μgの組換えベクターをヒト細胞株、例えば内皮由来または造血由来の細胞株に、一過的に形質移入する。更に、標識タンパク質をコードする配列を

50

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、形質移入された細胞と形質移入されていない細胞を区別する手段が与えられる。また、標識タンパク質の発現によって、cDNAの組換えベクターからの発現を正確に予想できる。標識タンパク質は、例えば緑色蛍光タンパク質(GFP; Clontech)、CD64またはCD64-GFP融合タンパク質から選択できる。自動化された、レーザー光学に基づく技術であるフローサイトメトリー(FCM)を用いて、GFPまたはCD64-GFPを発現する形質移入された細胞を同定し、それらの細胞のアポトーシス状態や他の細胞特性を評価する。FCMは、細胞死に先行するか或いは同時に発生する現象を診断する蛍光分子の取込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DNA染色によって計測される核DNA内容物の変化、前方散乱光と90°側方散乱光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DNA合成の下方調節、特異抗体との反応性によって計測される細胞表面及び細胞内におけるタンパク質の発現の変化、及び蛍光複合アネキシンVタンパク質の細胞表面への結合によって計測される原形質膜組成の変化とがある。フローサイトメトリー法については、Ormerod, (1994) Flow Cytometry, Oxford, New York NYに記述がある。

10

【0321】

遺伝子発現におけるCHOPの影響は、CHOPをコードする配列とCD64またはCD64-GFPのどちらかが形質移入された高度に精製された細胞集団を用いて評価することができる。CD64またはCD64-GFPは、形質移入された細胞表面で発現し、ヒト免疫グロブリンG(IgG)の保存された複数の領域に結合する。形質移入された細胞と形質移入されない細胞とは、ヒトIgGかCD64に対する抗体かのどちらかで被覆された磁気ビーズを用いて効率的に分離することができる(DYNAL, Lake Success NY)。mRNAは、当業者に周知の方法で細胞から精製することができる。CHOPと、目的とする他の遺伝子とをコードするmRNAの発現は、ノーザン分析やマイクロアレイ技術で分析することができる。

20

【0322】

1.5 CHOP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. 1816 - 3088-495)または他の精製技術で実質的に精製されたCHOPを用いて、標準的なプロトコルで動物(例えば、ウサギ、マウス等)を免疫化して抗体を作り出す。

【0323】

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

30

【0324】

通常は、長さ約15残基のオリゴペプチドを、FMOCケミストリを用いるABI 431Aペプチドシンセサイザ(Applied Biosystems)を用いて合成し、N-マレイミドベンゾイル-N-ヒドロキシスクシンイミドエステル(MBS)を用いた反応によってKLH(Sigma-Aldrich, St. Louis MO)に結合させて、免疫原性を高める(前出のAusubel, 1995)。完全フロイントアジュバントにおいて、オリゴペプチド-KLH複合体を用いてウサギを免疫化する。得られた抗血清の抗ペプチド活性及びCHOP活性を検査するには、ペプチドまたはCHOP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。

40

【0325】

1.6 特異的な抗体を用いる天然CHOPの精製

天然CHOP或いは組換えCHOPを、CHOPに特異的な抗体を用いるイムノアフィニティークロマトグラフィーにより実質的に精製する。イムノアフィニティークラムは、CNBr-活性化SEPHA ROSE(Amersham Biosciences)のような活性化クロマトグラフィー用レジンと抗CHOP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このレ

50

ジンをブロックし、洗浄する。

【0326】

CHOPを有する培養液をイムノアフィニティーカラムに通し、CHOP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で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CHOP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、CHOPを回収する。

【0327】

1.7 CHOPと相互作用する分子の同定

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

【0328】

或いは、CHOPと相互作用する分子は、Fields および Song (1989, Nature 340:245-246) に記載されているような酵母2ハイブリッドシステムを用いて分析するか、またはMATCHMAKERシステム（Clontech）等の2ハイブリッドシステムに基づく市販のキットを用いて分析する。

【0329】

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

【0330】

1.8 CHOP活性の実証

ガラクトシルトランスフェラーゼ活性アッセイ # 1

CHOPの β 1,3-ガラクトシルトランスフェラーゼ活性および β 1,4-ガラクトシルトランスフェラーゼ活性は、放射性アッセイにおけるUDP-ガラクトースからGlcNAc末端オリゴサッカライド鎖へのガラクトースの転移を測定することによって測定できる（Kolbinger, F. 他 (1998) J. Biol. Chem. 273:58-65; Hennet, T. 他 (1998) J. Biol. Chem. 273:58-65）。CHOPのアリコットを、14 μl のアッセイ用ストック溶液（180 mM のカコジル酸ナトリウム、pH 6.5、1 mg/ml のウシ血清アルブミン、0.26 mM のUDP-ガラクトース、2 μl のUDP-[^3H]ガラクトース）、1 μl の MnCl_2 (500 mM) および 2.5 μl の GlcNAc β 0-(CH_2) $_8$ -CO $_2$ Me（ジメチルスルホキシド中に37 mg/ml）で37 にて約1時間 インキュベートする。この反応は1mlの水を加えて停止させ、C18 Sep-Pak カートリッジ（Waters）に搭載し、カラムを2回5mlの水で洗浄して未反応のUDP-[^3H]ガラクトースを除去する。反応生成物の[^3H]ガラクトシル化GlcNAc β 0-(CH_2) $_8$ -CO $_2$ Me は水で洗浄する間はカラムに結合されたままであり、5mlのエタノールで溶出する。溶出された物質の放射活性は液体シンチレーションカウンターで測定され、これはサンプル中のCHOPガラクトシルトランスフェラーゼ活性に比例する。

【0331】

ガラクトシルトランスフェラーゼ活性アッセイ # 2

別方法として、CHOPの β 1,4ガラクトシルトランスフェラーゼ活性を[^{14}C]ガラクトースのDP[^{14}C]Gal からオボアルブミンへの転移の定量化によって求めることができる。約50 μl の反応液は、50 mMの HEPES (pH 7.35)、10 mMの MnCl_2 、1.5 mgのオボアルブミン、50 mMの NaCl、5 μl の UDP[^{14}C]ガラクトース (25 nCi)、および 5 μl の CHOPを含む。このアッセイを60 で 30分間インキュベートし、氷冷した2.5%のリンタングステン酸 (w/v) (1 M HCl中)を加えて停止させた。取り込まれなかったUDP[^{14}C]GalはWhatman GF/Cグラスファイバーフィルタによるろ過で分離した。フィルタは一度2.5%のリンタ

ングステン酸(w/v)で洗浄し、氷冷したエタノールでリンスした。フィルタを乾燥してから、シンチレーションカウンタを使って放射能を測定した。放射能の量はCHOPの活性に比例する(Verdon, B. および Berger, E. (1983) in *Galactosyltransferase. Methods of Enzymatic Analysis*, Bergmeyer, H. 他 (編集), 第III巻、第3版、Verlag Chemie, Weinheim, Deerfield Beach, Basel, 374381ページ; Tilo, S. 他(1996) *J. Biol. Chem.* 271:33983405)。

【0332】

シアリルトランスフェラーゼ活性アッセイ # 1

別方法として、CHOPのシアリルトランスフェラーゼ活性を次の方法でアッセイできる。このアッセイで使用するシアリルトランスフェラーゼ受容体は6(5フルオレセインカルボキ
10
サミド)ヘキサノ酸サクシミジルエステル (FCHASE)と反応させたアミノフェニルグリコシドから得られる。概略すると、10 mgのp-アミノフェニルグリコシドを0.5 mlの0.2 M 酢酸トリエチルアミンバッファ (pH8.2) に溶かす。FCHASE を0.5 ml のメタノールに溶かし、先のアミノフェニルグリコシド溶液に加える。この混合液を室温の暗い場所で約3時間攪拌し、凍結乾燥し、約 200 μ l の 50%アセトニトリルに再び懸濁する。シリカ薄層クロマトグラフィ (TLC) プレートに混合液をスポットして、酢酸エチル/メタノール/水/酢酸溶媒系で展開する。排気チェンバーの中で乾燥した後、黄色い生成物を掻き取り、蒸留水で溶出し、濃縮、脱塩の後、SepPak C18逆相カートリッジに結合させる。カートリッジを数倍量の水で洗浄し、50 %のアセトニトリルで溶出させた後、公知の分光法で定量
20
する。シアリルトランスフェラーゼ受容体を調製した後、20 μ l の容量、37 $^{\circ}$ C の温度で酵素反応を行う。このときの溶液は50 mM MES (pH 6.0)および10 mM $MnCl_2$ のバッファに0.2 または 1.0 mMの標識化受容体、0.2 mMのCMPNeu5Ac供与体およびさまざまな量のCHOPを加えたものである。10 mM NaOH による希釈によって反応を停止し、キャピラリー電気泳動法で分析する。キャピラリー電気泳動法 (CE) はアルゴンイオンレーザ誘導蛍光検出器 (励起 = 488 nm; 発光 = 520 nm) を使って行う。FCHASE2,3シアリルNアセチルラクトサミンからなる生成物ピークを同定し、定量する。これはサンプル中のシアリルトランスフェラーゼ活性に比例する(Wakarchuk, W. 他 (1996) *J. Biol. Chem.* 271:1916619173; Gilbert, M. (1996) *J. Biol. Chem.* 271:2827128276)。

【0333】

シアリルトランスフェラーゼ活性アッセイ # 2

別方法として、CHOPのシアリルトランスフェラーゼ活性を次の方法でアッセイできる。シアリルトランスフェラーゼのアッセイは、10 mMの $MgCl_2$ 、0.3% Triton CF54、100 mMのカコジル酸ナトリウムバッファ (pH 6.0)、0.66 mMの非標識化 CMPNeu5Ac 供与体、4
30
400 dpm/ μ l CMP[14 C]Neu5Ac (トレーサー)、CHOP溶液および基質を含む全容量20~50 μ lの反応混合液で行う。反応混合液を37 $^{\circ}$ C で2時間インキュベートし、500 μ lの水を加えて反応を停止する。生成物をC18 Sep-Pakカートリッジで単離し、エタノール/ピリジン/n-ブタノール/酢酸/水(100:10:10:3:30)の溶媒系を使った薄層クロマトグラフィ (TLC) で分析する (Okajima, T. 他 (1999) *J. Biol. Chem.* 274:1147911486)。放射能標識された生成物は、当業者に知られた標準的なオートラジオグラフィー技法で可視化する。CHOPのシアリルトランスフェラーゼ活性はオートラジオグラムのシグナルに比例する。
40

【0334】

シアリルトランスフェラーゼ活性アッセイ # 3

別方法として、CHOPのシアリルトランスフェラーゼ活性を次の方法でアッセイできる。CHOPをコードするプラスミドにより、ヒトの胚腎細胞 (293) を安定に形質転換する。細胞を225 cm^2 の組織培養フラスコでコンフルエンスまで成長させ、細胞をリン酸緩衝生理食塩水 (PBS) に掻き入れることによって収穫する。細胞をペレット化して、約1 ml の 1% Triton X100、50 mM NaCl、5 mM $MnCl$ および25 mM MES (pH 6.0) に再び懸濁する。ピペット操作とボルテックスを繰り返すことで細胞ペレットを溶かす。この未精製のホモジネートを1000 x gで10分間遠心してクリアにして、そのまま酵素源として使用する。アッセイ混合物は、250,000 cpm の CMP[14 C]Neu5Acをトレーサーとして加えた50 μ M CMPNeu5A
50

c、0.1% Triton CF54、20 mM カコジル酸塩 (pH 6.0) および 10 μ l の CHOP含有抽出液からなり、反応容量は30 μ lである。糖タンパク質および糖脂質生成物はゲルろ過によってCMPNeu5Acから分離し、溶出した分画内の標識の量を定量してサンプル内のCHOP活性の相対量を求める(Sjoberg, E. 他 (1996) J. Biol. Chem. 271:74507459)。

【0335】

0-グルコシルトランスフェラーゼのトランスフェラーゼ活性アッセイ

CHOPの0-グルコシルトランスフェラーゼ活性を次の方法でアッセイできる。使用の直前に、CHOP試薬を氷冷した脱塩バッファ(20 mM Tris pH 8.0、20% グリセロール、0.02% Na₃N₃)で希釈する。受容体としては約3 mMの濃度のペプチド基質(Kreppel, L. 他 (1999) J. Biol. Chem. 274:3201532022)を使用する。反応の中のペプチドは、SPSephadexカラムを 10 使って反応物から分離する。修飾されたペプチドと修飾されないペプチドをSeppak C18にロードする。修飾されないペプチドは50 mMの蟻酸、0.5 M NaClを含む10 ml の50 mM蟻酸、および10 mMの蒸留水で溶出する。修飾されたペプチドは、メタノールを使ってカートリッジから直接シンチレーションバイアルに溶出する。酵素活性は毎分転位されるGlcNAcのマイクロモル数で表されるが、これはサンプル中のCHOP活性のレベルに比例する(Kreppel 他, 前出)。

【0336】

マンノシダーゼ活性のアッセイ

CHOPのマンノシダーゼ活性はMan₉ (GlcNAc)₂オリゴ糖からマンノースを放出させる能力によって測定する(Schweden, J. 他 (1986) Eur. J. Biochem. 157:563-570)。200 mM リン 20 酸バッファ (pH 6.5) と1% Triton X-100に溶かしたCHOPを [¹⁴C](Man₉)(GlcNAc)₂ (2-3 x 10³ cpm)と混合し、最終容量を30 μ lとし、37 °Cで60分反応させる。30 μ lの氷酢酸を加えて反応を停止させる。放出された マンノースの量は2-プロパノール/酢酸/水(容積比29/4/9)によるペーパークロマトグラフィで分析するが、これは開始サンプル内のCHOPの活性に比例する。

【0337】

糖質結合アッセイ

CHOP活性は、 β ガラクトシド糖に結合する能力として実証される。CHOPがラクトシルセファロースカラムにかけられ、そのカラムを0.1Mのラクトースで溶出する。溶出液中のCHO 30 Pの存在は、SDSポリアクリルアミドゲル電気泳動によって検出され、CHOPの β ガラクトシド糖への結合能力を示す。

【0338】

ヒアルロナン加水分解アッセイ

CHOPの活性は、ヒアルロナン(HA)を加水分解する能力によっても測定することができる (Lepperdinger, 前出)。放射能標識されたHAを、1-エチル-3-(3-ジメチルアミノプロピル)カルボジイミドとN-ヒドロキシ サルフォサクシンイミド(sulfosuccinimide)を使って、微量定量プレートに固定化する。CHOPとインキュベーションした後の可溶化された放射能が液体シンチレーションカウンターを使って測定され、その放射能は開始サンプルのCHOP活性と比例する。

【0339】

細胞形質転換アッセイ

別法では、CHOP活性は、NIH3T3マウス線維芽細胞の形質転換を調節する能力によって測定される。CHOPをコードするcDNAは好適な真核生物発現ベクターにサブクローンされる。このベクターは当分野で周知の方法を用いてNIH3T3細胞に形質移入される。形質移入細胞と非形質移入細胞を、発癌性形質転換細胞の定量化できる特徴的性質について比較する。すなわち、接触障害の喪失に関連する培養中の高密度増殖、懸濁液中または軟寒天での増殖、血清必要性の低下、および免疫不全マウスに注入された場合の腫瘍誘発能について評価する。CHOPの活性は、CHOPで形質移入されたNIH3T3細胞と非形質移入細胞の形質転換の程度に比例する。

【0340】

10

20

30

40

50

1 9 CHOP関連 α 1,3ガラクトシルトランスフェラーゼ活性の発現と精製

可溶で α 1,3-ガラクトシルトランスフェラーゼ (α 1,3-GalT) 活性を持つ触媒活性のある CHOP を、CHOP を発現するプラスミドで形質移入した、ほぼコンフルエントなヒトの胚腎 (293) 細胞の培地から精製する。培地を 20 mM HEPES (pH 7.0) と 5 mM MnCl_2 に対して一晩、4 で透析する。そのときバッファを何度か取替え、次に UDP ヘキサノールアミン-セファロース (4 $\mu\text{mol/ml}$ レジン) のアフィニティカラムにかける。溶出液の A_{280} がバックグラウンドレベルになるまで、カラム透析バッファと 0.75 M NaCl を含む透析バッファで洗浄する。結合した酵素は透析バッファに 5 mM UDP を溶かしたもので溶出し、Centricon10 (Millipore) を使って濃縮する。精製された酵素の均一度は、10% SDSPAGE ゲル上の分離により求める (Kim, S. 他 (1997) J. Biol. Chem. 272:13622-13628)。

10

【 0 3 4 1 】

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。分子生物学または関連分野の専門家には明らかな、本明細書に記載する本発明の実施方法の様々な修正は、明確に特許請求の範囲内にあるものとする。

【 0 3 4 2 】

(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。

20

【 0 3 4 3 】

表 2 は、本発明のポリペプチド群の GenBank 識別番号と、最も近い GenBank 相同体の注釈 (annotation) と、PROTEOME データベース識別番号と、PROTEOME データベース相同体群の注釈とを示す。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。

【 0 3 4 4 】

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。

30

【 0 3 4 5 】

表 4 は、本発明のアセンブリされたポリヌクレオチド配列およびポリヌクレオチド配列から選択された断片の SEQ ID NO:、Incyte ID および長さを示す。

【 0 3 4 6 】

表 5 は、本発明のポリヌクレオチドの代表的な cDNA ライブラリを示す。

【 0 3 4 7 】

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

【 0 3 4 8 】

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、参照文献及び閾値パラメータと共に示す。

40

【 0 3 4 9 】

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	ポリヌクレオチド SEQ ID NO	Incyte ポリヌクレオチド ID
7247177	1	7247177CD1	11	7247177CB1
5596327	2	5596327CD1	12	5596327CB1
4179651	3	4179651CD1	13	4179651CB1
7482109	4	7482109CD1	14	7482109CB1
2867618	5	2867618CD1	15	2867618CB1
7488348	6	7488348CD1	16	7488348CB1
5539166	7	5539166CD1	17	5539166CB1
7500341	8	7500341CD1	18	7500341CB1
7500846	9	7500846CD1	19	7500846CB1
3230921	10	3230921CD1	20	3230921CB1

10

表 2				
ポリペプチドIncyte ポリGenBank ID NO:	SEQ ID	ペプチドID	確率スコア	注釈
1	7247177CD1	g2104689	1.4E-258	[マウス] α グルコシダーゼ II、 α サブユニット Arendt, C.W. および Ostergaard, H.L. (1997) J. Biol. Chem. 272:13117-13125.
2	5596327CD1	g2621120	1.9E-17	[メタン生成古細菌 (Methanothermobacter thermotrophicus)] O-結合型 GlcNAc トランスフェラーゼ Smith, D.R. 他 (1997) J. Bacteriol. 179:7135-7155.
3	4179651CD1	g14150450	0.0	[ラット] UDP-GalNAc:ポリペプチド N-アセチルガラクトサミンルトランスフェラーゼ T9 Ten Hagen, K.G. 他 (2001) J. Biol. Chem. 276:17395-17404.
4	7482109CD1	g163124	5.8E-77	[ウシ] α -1,3-ガラクトシルトランスフェラーゼ Joziasse, D.H. 他 (1989) J. Biol. Chem. 264:14290-14297.
5	2867618CD1	g2642187	4.4E-182	[ラット] エンド- α -D-マンノシダーゼ Spiro, M.J. 他 (1997) J. Biol. Chem. 272:29356-29363.
6	7488348CD1	g31179	4.1E-243	[ヒト] エノラーゼ
7	5539166CD1	g9650954	2.1E-177	[マウス] β -1,6-N-アセチルグルコサミンルトランスフェラーゼ B Chen, G.Y. 他 (2000) Glycobiology 10:1001-1011.
8	7500341CD1	g3790363	5.4E-30	[ヒト] DPM2 Maeda, Y. 他(1998) EMBO J. 17:4920-4929. Lennon, G. 他 (1996) Genomics 33:151-152.
		339872[DPM2]	4.8E-31	[ヒト][調節サブユニット; トランスフェラーゼ] [小胞体; 細胞質] ドリコールリン酸マンノースシンターゼの調節サブユニットであり、ドリコールリン酸マンノースの合成を調節。後者は N-結合型オリゴ糖前駆体およびグリコシルホスファチジルイノシトールアンカーのコアにとってマンノシル残基の供与体。

20

30

表 2 (続)				
ポリペプチドIncyte ポリGenBank ID NO: 確率スコア 注釈				
SEQ ID NO:	ID	ペプチドID		
9	7500846CD1	g8096221	2.1E-83	[ヒト] インテレクチン Tsuiji, S. 他 (2001) J. Biol. Chem. 276:23456-23463.
		585107Jltn	4.2E-80	[マウス] インテレクチン。小腸のパネート細胞で発現し、微生物に対する防御に役割を果たしている可能性がある。
10	3230921CD1	g14043157	1.1E-51	[ヒト] カーゴ選択タンパク質 (マンノース-6-リン酸 受容体結合タンパク質) Diaz, E. および Pfeffer, S.R. (1998) Cell 93:433-443.
10 cont.		342866TIP47	9.2E-53	[ヒト][エンドソーム/エンドソーム小胞; 細胞質] 47 kDaの尾部相互作用タンパク質。マンノース-6-リン酸受容体の細胞質内ドメインに結合し、エンドソームからゴルジ装置への輸送を仲介する。子宮頸部癌および正常な妊娠において血清レベルが上昇。 Orsel, J. G., 他 (2000) Proc. Natl. Acad. Sci. USA 97:9047-9051.
		33906QADFP	9.8E-33	[ヒト][小胞体; 細胞質; 脂質粒子] 脂肪分化関連タンパク質。脂肪酸輸送に関与する膜結合タンパク質。脂質蓄積のマーカー。
		583559Adip	2.7E-26	[マウス][小分子結合タンパク質][細胞質; 不特定の膜] 脂肪分化関連タンパク質。脂肪酸およびコレステロール結合および長鎖脂肪酸輸送に関与する膜結合タンパク質。脂質蓄積のマーカー。 Tansey, J. T., 他 Proc. Natl. Acad. Sci. U.S.A. 98:6494-9. (2001)
		337032PLIN	2.8E-16	[ヒト] ペリリピン。脂肪細胞内のタンパク質キナーゼ A によって複数のリン酸化を受ける可能性あり。
		329340Plin	3.5E-16	[ラット][細胞質; 脂質粒子] ペリリピン。脂肪分解を調節。脂肪細胞内のタンパク質キナーゼ A によって複数のリン酸化を受ける。

表 3					
SEQ ID NO:	Incyte リペプチド ID	ポリアミノ酸残基数	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	7247177CD1912		S10 S38 S56 S145 S427 S477 S577 S804 S809 S835 S844 T42 T51 T52 T103 T108 T180 T184 T290 T340 T650 T732 T759 T907 Y246 Y449	N286 N843	グリコシル ヒドロラーゼ ファミリー: P138-T875 HMMER_PFAM 膜貫通ドメイン:W592-F620 N-末端は細胞 TMAP 質内 グリコシル ヒドロラーゼ ファミリー BL00129:BLIMPS_BLOC G353-D398, G464-Y490, R565-W592,KS L603-A646, R667-F702, T783-S820 ヒドロラーゼ グリコシダーゼ 糖タンパクBLAST_PRODO ク質 α-グルコシダーゼ マルターゼ 前駆M 体 シグナル タンパク質 グルコシダーゼ シュクラーゼ イソマルターゼ PD001543: W195-F518, Q523-T874 ヒドロラーゼ グリコシダーゼ α-グルコBLAST_PRODO シダーゼ 糖タンパク質 マルターゼ シュM クラーゼ イソマルターゼ 前駆体 シグナ ル 酸 腸の PD001716: G329-W472, H224-Y302

10

20

30

表 3 (続)						
SEQ ID NO:	Incyte リペプチド ID	ポ 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					シュクララーゼ/イソマルターゼ DM01359 P10253 285-950: H537-L898 DM01359 P23739 278-931: Q523-A906 DM01359 P14410 268-925: N534-I905 DM01359 JC4624 199-864: E522-Q822	BLAST_DOMO
cont.					グリコシルヒドロラーゼファミリー 31 シグネチャ: F507-E514	MOTIFS
2	5596327CD1 691		S80 S137 S198 S291 S327 S510 S615 S658 S664 S674 S684 T99 T541 T561 T671 Y270		ATP シンターゼ α 鎖, C 末端: R128-L142	HMMER_PFAM
					TPR ドメイン: I149-H182, V183-N216, HMMER_PFAM P115-F148, P48-A81, R82-Q114 膜貫通ドメイン: E575-Y591, P595-K617 N-末端は細胞質内でない。 ミトコンドリア 外膜 タンパク質, 70K DM04281 P23231 276-618: I17-Q180	TMAP BLAST_DOMO

10

表 3 (続)						
SEQ ID NO:	Incyte リペプチド ID	ポ 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	4179651CD1 506		S21 S29 S40 S75 S90 S108 S141 S253 S319 S337 S443 S465 S470 T2 T114 T137 T409 T422 T437 T498 Y196 Y238 Y459	N27 N49 N496	グリコシル転移酵素: S51-G236	HMMER_PFAM
3					QXW レクチンリピーター: G417-D455, HMMER_PFAM K456-T495, G364-N402 N-アセチルガラクトサミニルトランスフ ェラーゼ M トランスフェラーゼ ポリペプチド アセ チルガラクトサミニルトランスフェラー ゼ UDP-GALNAC:ポリペプチド グリコ シルトランスフェラーゼ タンパク質 UDP タンパク質 UDP N PD003162-I202-A362 do アセチルガラクトサミニルトランス フェラーゼ; ポリペプチド DM03891 Q07537 32-558: D3-F503 DM03891 P34678 37-600: D8-E493 DM03891 I37405 21-571: Y11-F492	BLAST_PRODO
cont.						

20

30

表 3 (続)						
SEQ ID NO:	Incyte リペプチド ID	ポアミノ酸残基数	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
4	7482109CD1	276	S3 S153 S260 T47 T84 T131 T139 T213 Y70	N74 N259	signal_cleavage: M1-V19 シグナルペプチド: M1-R21, M1-Y22, HMME M1-R24 トランスフェラーゼ α 1,3- ガラクトシBLAST_PRODO ルトランスフェラーゼ N-アセチルラクM ト サ ミ ニ ド (ACETYLLACTOSAMINIDE) ガラク トシルトランスフェラーゼ UDP-ガラク トース: β D-ガラクトシル 1 グループ 4-N-アセチル D-グルコサミニド グリコ シルトランスフェラーゼ PD010022:K41-T276 シグナル-アンカー 膜貫通 (ガラクトシBLAST_DOMO ルトランスフェラーゼ フコシル糖タンパ ク質 α -1 α) DM07533 P14769 6-367: E20-T276 DM07533 P16442 16-353: R42-T276 do ガラクトシルトランスフェラーゼ; α BLAST_DOMO -1; アセチルラクトサミニド; DM08008 A56480 1-376: E20-T276 DM08008 I49698 1-371: R24-T276	SPSCAN
4 cont.						
5	2867618CD1	378	S80 S85 S206 S256 S280 S317 S336 S359 T125 T190 T325 T355 Y369		エンド- α D-マンノシダーゼ PD141586: P3-S378	BLAST_PRODO M

10

20

表 3 (続)						
SEQ ID NO:	Incyte リペプチド ID	ポアミノ酸残基数	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	7488348CD1	458	-{}-S263 S272 S306 S370 S392 S424 T27 T246 T341 T421 Y61 Y279	N76 N108	エノラーゼ: S2-R456 エノラーゼタンパク質 BL00164: F403-A441, R35-D57, KS I153-K202, N229-K271, P302-D316, I331-N366 エノラーゼシグネチャ: Q314-L362 エノラーゼシグネチャ PR00148: V38-E52, G113-G129, S A173-R186, G335-R346, L361-L375, S392-V409 エノラーゼ リアーゼ 解糖 マグネシウムBLAST_PRODO ヒドロラーゼ 2-ホスホグリセリン酸 デM ヒドラターゼ 2-ホスホ-D-グリセリン酸 2-ホスホ-D-グリセリン酸 PD000902: M1-V293 リアーゼ マグネシウム エノラーゼ 解BLAST_PRODO 糖 ヒドロリアーゼ (hydrolyase) デヒドM ラターゼ 2-ホスホグリセリン酸 2-ホス ホ-D-グリセリン酸 2-ホスホ-D-グリセリ ン酸 PD003216: D275-E440	HMMER_PFAM
6 cont.						

30

40

表 3 (続)						
SEQ ID NO:	Incyte リペプチド ID	ポ 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
					エノラーゼ リアーゼ 解糖 マグネシウム ヒドロラーゼ 2-ホスホグリセリン酸 ヒドラターゼ 2-ホスホ-D-グリセリン酸 2-ホスホ-D-グリセリン酸 PD000948: V408-R456 エノラーゼ DM00487 S22071 2-455: S2-R456 DM00487 S52858 1-431: M1-P455 DM00487 JC1039 1-431: M1-P455 DM00487 P15429 1-430: I7-R456 エノラーゼシグネチャ: L361-S374	BLAST_PRODO BLAST_DOMO MOTIFS

7	5539166CD1 402		S4 S89 S154 S158 S295 S329 T67 T68 T73 T87 T199 T247 Y206	N41 N316 N390	signal_cleavage: M1-N25	SPSCAN
7 cont.					コア-2/I - 分枝酵素: T73-L382 サイトゾル内ドメイン: M1-K6 膜貫通ドメイン: H7-W29 非サイトゾル内ドメイン: E30-F402 タンパク質 β 1-トランスフェラーゼ グ リコシルトランスフェラーゼ 酵素 コア M 6-N-アセチルグルコサミニルトランスフ ェラーゼ 3-ガラクトシルグリコシル糖タ ンパク質 6-N-アセチルグルコサミニルト ランスフェラーゼ PD005410: L171-H344 PD003538: T73-D170	HMMER_PFAM TMHMMER

表 3 (続)						
SEQ ID NO:	Incyte リペプチド ID	ポ 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
					β 1-トランスフェラーゼ グリコシルト ランスフェラーゼ 酵素 コア 6-N-アセチ ルグルコサミニルトランスフェラーゼ 3- ガラクトシルグリコシル 糖タンパク質 6-N-アセチルグルコサミニルトランスフ ェラーゼ 分岐 PD011484: G345-F402 管腔ドメイン DM07544 Q06430 8-399: L9-F402 DM07544 Q02742 29-427: C74-R385	BLAST_PRODO BLAST_DOMO
8	7500341CD1 109		T3 T73		signal_cleavage: M1-S67 サイトゾル内ドメイン: M1-V8, P69-D109 膜貫通ドメイン: V9-L31, P46-F68 非サイトゾル内ドメイン: P32-L45	SPSCAN TMHMMER
9	7500846CD1 225		S30 S52 S62 S96 T80 T107 T142		signal_cleavage: M1-A26 シグナルペプチド: M1-A27, M1-A28, HM M4-A26, M4-A28, M8-C24, M8-A26, M8-A28, M8-A29 フィブリノーゲン β および γ 鎖, C 末端: HM L49-V94 サイトゾル内ドメイン: M1-C12 膜貫通ドメイン: F13-L35 非サイトゾル内ドメイン: S36-C225 レクチン インテレクチン 皮質 顆粒 前 駆体 糖タンパク質 PD040823: R51-Q200	SPSCAN HMMER TMHMMER M

9 cont.						
------------	--	--	--	--	--	--

表 3 (続)						
SEQ ID NO:	Incyte リペプチド ID	ポリアミノ酸残基数	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
10	3230921	CD1463	S2 S13 S117 S140 S187 S205 S273 S277 S304 T68 T169		signal_cleavage: M1-A38	SPSCAN
ベリリピンファミリー: P10-V369 HMMER_PFAM						
タンパク質 ベリリピン 脂肪細胞分化関連BLAST_PRODO						
ADRP 膜 カargo 選択 (cargo selection) M						
TIP47 A/B PD018256: S148-E384, Q19-G128						

表 4	
ポリヌクレオチド SEQ ID NO/ Incyte ID/配列長	配列断片
11/7247177CB1/ 3150	1-388, 1-394, 1-730, 1-737, 5-358, 5-423, 35-603, 220-394, 228-423, 239-978, 239-980, 285-389, 295-485, 295-905, 297-544, 311-418, 311-709, 454-719, 498-1352, 508-1352, 546-1352, 549-669, 549-751, 549-812, 550-1352, 568-1168, 580-1352, 585-1352, 589-1088, 589-1236, 619-1352, 681-1352, 790-1650, 790-1661, 821-1714, 872-1147, 876-1584, 885-1584, 956-1608, 971-1615, 1002-1621, 1063-1942, 1070-1961, 1075-1277, 1098-1704, 1099-1781, 1099-1864, 1184-1635, 1193-1893, 1224-1901, 1235-1955, 1236-1961, 1250-1955, 1315-1574, 1324-1955, 1345-1955, 1398-1967, 1457-1954, 1496-1721, 1498-1949, 1505-1952, 1505-1955, 1512-1595, 1518-1864, 1651-1864, 1786-2256, 1817-2103, 1827-2567, 1827-2584, 1827-2598, 1827-2614, 1827-2637, 1831-2622, 1865-2227, 1892-2474, 1894-2196, 1918-2032, 1923-2092, 1946-2324, 1975-2840, 1988-2586, 2188-2897, 2200-2897, 2205-2897, 2206-2897, 2208-2897, 2413-2526, 2417-2885, 2447-3062, 2489-3150, 2605-2951, 2641-2804, 2789-2932, 2802-2951, 2869-2951, 3113-3138
12/5596327CB1/ 3290	1-788, 303-1028, 315-1031, 591-798, 662-962, 683-960, 683-1394, 812-1327, 840-1045, 887-1605, 890-1054, 894-1126, 1010-1623, 1031-1302, 1052-1144, 1090-1723, 1140-2052, 1349-1637, 1372-1618, 1524-1761, 1594-2220, 1636-1965, 1659-1956, 1662-1946, 1748-2036, 1919-2156, 1919-2371, 1919-2397, 2062-2282, 2214-2912, 2256-2538, 2256-2569, 2285-2466, 2285-2593, 2397-2603, 2397-2605, 2424-2682, 2424-2698, 2427-2854, 2453-3065, 2462-2745, 2462-3197, 2495-2631, 2543-3211, 2784-3290, 2801-3060, 2924-3185, 2941-3188, 3072-3154, 3072-3221
13/4179651CB1/ 4279	1-695, 2-126, 7-281, 265-4279, 267-440, 301-523, 380-573, 389-1062, 551-1058, 560-1043, 560-1108, 857-1252, 938-1061, 994-1560, 1081-1379, 1169-1391, 1265-1556, 1328-1369, 1373-1993, 1393-1433, 1554-1786, 1554-2093, 1754-2218, 1756-2060, 1793-2052, 1825-2259, 1874-2127, 1875-2460, 1922-2189, 1932-2180, 1967-2451, 1969-2254, 1981-2640, 2042-2534, 2307-2541, 2357-2783, 2389-2670, 2451-2923, 2455-3049, 2464-2724, 2510-3099, 2562-3072, 2619-2856, 2662-3273, 2663-2858, 2666-3246, 2719-2946, 2747-2995, 2747-3269, 2778-3037, 2986-3292, 3139-3416, 3162-3390, 3162-3656, 3186-3469, 3187-3859, 3199-3501, 3210-3493, 3232-3800, 3280-3490, 3296-3818, 3333-3593, 3335-3617, 3340-3605, 3362-3609, 3362-3835, 3362-3843, 3406-3764, 3432-3860, 3525-3751, 3536-3781, 3536-3782, 3641-3905, 3695-3878, 3698-3866, 3698-3870, 3709-3953, 3709-4209, 3933-4202, 3953-4214, 3968-4107, 4033-4279

表 4	
ポリヌクレオチド SEQ ID NO/ Incyte ID/配列長	配列断片
14/7482109CB1/ 1042	1-119, 70-257, 72-831, 677-1042, 678-1042, 680-1041, 680-1042
15/2867618CB1/ 4320	1-641, 1-653, 565-1278, 877-1420, 1055-1618, 1079-1693, 1106-1729, 1142-1922, 1188-1774, 1338-2100, 1358-1974, 1394-2074, 1478-1760, 1481-1878, 1760-2479, 1820-2098, 1841-2531, 1885-2045, 1977-2517, 1986-2165, 2001-2591, 2035-2507, 2057-2531, 2123-2532, 2140-2524, 2145-2747, 2151-2532, 2172-2532, 2221-2547, 2225-2493, 2330-3026, 2652-2874, 2706-3044, 2710-3396, 2856-3130, 2876-3159, 2892-3127, 2892-3451, 2918-3173, 3031-3596, 3070-3504, 3086-3338, 3108-3386, 3157-3493, 3174-3455, 3246-3549, 3255-3463, 3255-3707, 3308-3579, 3319-4013, 3321-4010, 3342-3695, 3526-4170, 3570-4148, 3575-4261, 3615-3869, 3617-3904, 3675-3942, 3701-3871, 3742-4032, 3742-4035, 3758-4296, 3776-4247, 3789-4264, 3814-4265, 3818-4268, 3820-4083, 3824-4102, 3834-4273, 3844-4046, 3844-4103, 3844-4234, 3846-4100, 3846-4320, 3847-4273, 3849-4267, 3858-4263, 3910-4277, 3938-4221, 3991-4222, 3991-4255, 3991-4266, 4002-4292, 4004-4260, 4037-4273, 4064-4272, 4075-4261, 4080-4268, 4085-4273, 4113-4264
16/7488348CB1/ 1733	1-1733, 33-1404
17/5539166CB1/ 2201	1-430, 209-939, 228-476, 239-1055, 262-859, 350-991, 453-1120, 457-1013, 457-1019, 611-1443, 682-869, 756-963, 756-1218, 756-1230, 820-1338, 1079-1530, 1411-1529, 1469-1998, 1469-2107, 1470-1758, 1470-1908, 1470-1943, 1470-2011, 1470-2065, 1470-2066, 1470-2201
18/7500341CB1/ 457	1-126, 1-218, 1-230, 1-457, 2-230, 12-126, 15-164, 16-230, 17-223, 19-230, 20-230, 22-168, 23-230, 28-175, 28-230, 29-230, 31-209, 36-445, 36-451, 37-202, 38-230, 42-202, 50-452, 228-428, 232-445, 310-446

10

20

30

40

50

表 4	
ポリヌクレオチド SEQ ID NO/ Incyte ID/配列長	配列断片
19/7500846CB1/ 1513	1-268, 1-276, 1-495, 1-593, 1-1020, 4-266, 8-578, 13-221, 16-283, 112-643, 214-499, 246-678, 301-1011, 332-998, 384-648, 442-861, 654-803, 660-1023, 669-948, 669-972, 669-1013, 669-1023, 669-1032, 693-1023, 775-1020, 854-1513, 871-1111
20/3230921CB1/ 2084	1-240, 1-261, 1-266, 28-291, 28-650, 33-619, 36-332, 46-717, 47-896, 53-583, 115-354, 115-660, 160-297, 160-599, 160-818, 231-355, 231-369, 231-378, 231-422, 231-438, 238-438, 250-438, 254-438, 343-593, 439-568, 439-576, 439-582, 439-604, 439-613, 439-614, 439-619, 443-619, 444-572, 445-619, 449-619, 453-619, 456-619, 458-619, 466-619, 468-689, 468-795, 468-1004, 472-619, 483-1054, 483-1124, 497-616, 620-889, 620-1000, 620-1058, 620-1118, 641-1123, 646-1110, 653-892, 664-1126, 700-771, 716-1118, 723-1118, 750-1127, 751-1118, 759-1118, 931-1027, 957-1127, 1048-1118, 1068-1411, 1119-1915, 1120-1696, 1128-1677, 1144-1325, 1144-1446, 1144-1483, 1144-1605, 1144-1624, 1164-1408, 1164-1635, 1167-1652, 1200-1583, 1217-1333, 1223-1770, 1250-1893, 1261-1612, 1300-1691, 1304-1569, 1304-1750, 1319-1916, 1325-1920, 1401-1698, 1419-1722, 1419-1921, 1467-1727, 1482-1794, 1485-2084

表 5		
ポリヌクレオチド SEQ ID NO	Incyte プ ロ ジ ェ ク ト I D	代 表 的ライブラリ
11	7247177CB1	MIXDTXE01
12	5596327CB1	OVARNOT09
13	4179651CB1	BRAGNON02
15	2867618CB1	HIPONON01
17	5539166CB1	KIDNFEC01
18	7500341CB1	LNODNOT03
19	7500846CB1	PROSNOT06
20	3230921CB1	PANCNOT23

表 6 (続)		
ライブラリ	ベクター	ライブラリの説明
BRAGNON02 pINCY		このノーマライズした黒質組織ライブラリは黒質組織ライブラリの 4.2×10^7 の独立したクローンから作製した。開始 RNA はアテローム性動脈硬化に起因する出血と胸大動脈破裂で死亡した 81 歳の白人女性から採取された黒質から単離した RNA から調製した。病理学検査では、双方の内部頸動脈に中程度のアテローム性動脈硬化、前頭皮質および海馬に微視的な梗塞、および散在広汎性のアミロイド斑、および神経原線維変化が見られたが、これは年齢と矛盾しない。肉眼で見て、軟膜は上矢状静脈洞に沿った軽度の硝子質化と厚化が見られるのみであった。残りの軟膜は薄く、充血した血管がいくつか見られた。主に前頭極および前頭葉、及び側頭葉(双方)に軽度の萎縮が見られた。微視的には、アルツハイマータイプ II 星状細胞のペアが新皮質の深層にあった。中央前頭皮質の灰白質の奥深くのニューロンの周りでサテリトシス(satellitosis)が増大していた。扁桃はまれなびまん性斑と神経原線維変化を含んでいた。後部海馬は反応性グリオシスに囲まれたヘモジデリン含有マクロファージを伴う微視的な囊胞性の空洞化領域を含んでいた。患者の病歴は、敗血症、胆管炎、術後無気肺、肺炎冠動脈疾患(CAD)、左心室肥大に起因する心肥大、巨脾症、細動脈性腎硬化、結節性コロイド甲状腺腫、気腫、鬱血心不全(CHF)、甲状腺機能低下症および抹消血管疾患を含んでいた。このライブラリは、極めて長時間(48 時間/1 回)の再アニーリングによるハイブリダイゼーションを用いた他は、Soares 他, 1994 (Proc. Natl. Acad. Sci. USA 91:9228-9232)及び Bonaldo 他, 1996 (Genome Res.6:791)を適用した条件を用いて 2 回にわたリノーマライズした。
HIPONON01 PSPORT1		このノーマライズした海馬ライブラリは、或る海馬組織ライブラリの 113 万個の独立クローンから作製した。RNA は、頭蓋内出血で死亡した 72 才白人女性の海馬組織から単離された。患者の病歴には鼻癌、高血圧および関節炎がある。ノーマライズ条件およびハイブリダイズ条件は、Soares 他 1994 (Proc. Natl. Acad. Sci. USA 91:9228-9232)を適用した。
KIDNFEC01 PBLUESCRIPT		ライブラリは、妊娠 19〜23 週間で自然流産した 12 人の白人男子および女子胎児のブルールから採取した腎組織から単離した RNA を使って作製した。
LNODNOT03 pINCY		ライブラリは、気管支鏡法と肺の部分除去中の 67 才白人男性から取り除かれたリンパ腺組織から単離された RNA を用いて作製された。顕微鏡検査では、この組織は 10%の生存腫瘍のある強度な壊死であることが発見された。随伴する腫瘍組織の病理学検査では、グレード 3 から 4 の扁平上皮細胞の侵襲性癌が見られた。患者の病歴には血管腫が含まれる。家族歴には、アテローム硬化型冠動脈疾患、悪性の高血圧、鬱血性心不全症、およびアテローム硬化型冠動脈疾患が含まれる。
MIXDTXE01 PBK-CMV		この 5' 偏向のランダムプライムされたライブラリは 9 人のドナーからのブルールされた cDNA を使って作製された。男性 (ドナー A) の T 細胞から得られた Jurkat 細胞株、1 歳の男子 (ドナー B) の末梢血から得られた THP-1 細胞株、16 歳の黒人男子 (ドナー C) の B リンパ芽球から得られた Daudi 細胞株、29 歳の白人男性 (ドナー D) のリンパ腫から得られた RPMI-1666 細胞株、1 歳の白人男子 (ドナー E) の脾臓、胸腺除去手術中に 21 歳の白人男性 (ドナー F) から除去された胸腺、42 歳の白人女性 (ドナー G) のリンパ節、全胸腺除去手術中に 56 歳の白人女性 (ドナー H) から得られた胸腺腫瘍、およびドナーのブルール (ドナー I) から得られた PBMC か

10

20

30

40

50

表 6 (続)		
ライブラリ	ベクター	ライブラリの説明
OVARNOT09 pINCY		ライブラリは、28 オ白人女子の膣式子宮摘出術、並びに卵管・卵巣摘出時に摘出された卵巣組織から単離された RNA を用いて作製された。病理学的検査は、右および左の卵巣内にサイズが 0.4 から 5cm の多数の卵胞嚢胞、慢性子宮頸管炎および子宮頸管の扁平上皮異形成および弱い増殖相にある子宮内膜を示した。家族歴には、良性高血圧症、高脂血症およびアテローム性冠動脈疾患が含まれている。
PANCNOT23 pINCY		ライブラリは、発砲による頭部の損傷で死亡した 43 オの白人女性から採取した病変した膵臓組織から単離した RNA を用いて作製した。患者の病歴としては、I 型の糖尿病と血清学的に陽性な CMV 抗体などがあった。
PROSNOT06 PSORT1		ライブラリは、57 オ白人男性の根治的前立腺除去、両精巣の除去、および所属リンパ節切除時の罹患前立腺組織から単離された RNA を用いて作製された。病理学検査では腺癌増殖増殖症が見られた。関連する腫瘍組織の病理学検査では、腺癌 (Gleason グレード 3+3) が見られた。患者の病歴には良性の大腸腫瘍、およびタイプ I 糖尿病が含まれる。家族歴には前立腺の悪性新生物、タイプ I 糖尿病が含まれる。

表 7

プログラム	説明	引用文献	パラメータ閾値
ABI FACTURA	ベクター配列を除き、核酸配列のまぎらわしい塩基をマスクするプログラム	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	アミノ酸または核酸配列の比較、注釈に有用なファーストデータファインダ。	Applied Biosystems, Foster City, CA. Paracel Inc., Pasadena, CA.	不一致<50%
ABI オートアセンブラ	核酸配列のアセンブリを行うプログラム。	Applied Biosystems, Foster City, CA.	
BLAST	アミノ酸配列および核酸配列の配列類似性検索に有用なベーシックローカルアラインメント検索ツール (Basic Local Alignment Search Tool)。BLAST には 5 つの機能がある: blastp、blastn、blastx、tblastn および tblastx。	Altschul, S.F. 他 (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. 他 (1997) Nucleic Acids Res. 25:3389-3402.	<i>EST</i> s 確率値=1.0E-8 以下 完全長配列: 確率値= 1.0E-10 以下
FASTA	一群の同じタイプの配列と問合せ配列との間の類似性について検索する Pearson および Lipman アルゴリズム。FASTA は最少 5 つの機能からなる: fasta、tfasta、fastx、tfastx および ssearch。	Pearson, W.R. 及び 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.	<i>EST</i> s: fasta E 値=1.0E-6 アセンブリされた <i>EST</i> s: 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フィンガープリント領域を検索する BLocks IMProved Searcher。	Henikoff, S. 及び J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572、Henikoff, J.G. 及び S. Henikoff (1996) Methods Enzymol. 266:88-105、Attwood, T.K. 他 (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	確率値= 1.0E-3 以下
HMMER	問合せ配列を PFAM のようなタンパク質ファミリー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. 1-350.	PFAM ヒット確率値= 1.0E-3 以下 シグナル ペプチド ヒット: スコア=0 以上

表 7 (続)

10

20

30

プログラム	説明	引用文献	パラメータ閾値
ProfileScan	Prosite で定義された配列パターンと一致するタンパク質配列内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 of Enzymol.183:146-159 、Bairoch, A. 他.(1997) Nucleic Acids Res. 25:217-221.	標準化された質スコア≧特定 の Prosite モチーフに対する GCG 指定「HIGH」値 通常、スコア=1.4〜2.1.
Phred	高度感受性および確立性のある自動シーケンサートレースを調べる塩基コーリングアルゴリズム。	Ewing, B. 他(1998) Genome Res. 8:175-185, Ewing, B 及び P. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的なインプレ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配列のアセンブリに有用である。	Smith, T.F. 及び M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. 及び 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 アセンブリ体の表示および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、分泌シグナルペプチドの存在を調べる重み行列分析プログラム。	Nielson, H. 他(1997) Protein Engineering 10:1-6; Claverie, J.M. 及び S. Audic (1997) CABIOS 12:431-439.	スコア=3.5 以上
TMAP	重み行列を使ってタンパク質配列上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。	Persson, B. および P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. および P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	隠れ Markov モデル(HMM) を使ってタンパク質配列上の膜貫通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., Glasgow 他, 編集, The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, 175-182 ページ.	
Motifs	Prosite で定義されたものと一致するパターンのアミノ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, バージョン 9, M51-59 ページ, Genetics Computer Group, Madison, WI	

10

20

<110> INCYTE GENOMICS, INC.

SWARNAKAR, Anita

GORVAD, Ann E.

HAFALIA, April J.A.

DUGGAN, Brendan M.

EMERLING, Brooke M.

ISON, Craig H.

10

NGUYEN, Dannel B.

LEE, Ernestine A.

YUE, Henry

YUE, Huibin

FORSYTHE, Ian J.

LI, Joana X.

THANGAVELU, Kavitha

20

WALIA, Narinder K.

BURFORD, Neil

MASON, Patricia M.

LAL, Preeti G.

LEE, Sally

BECHA, Shanya D.

TANG, Y. Tom

30

<120> Carbohydrate-Associated Proteins

<130> PF-0982 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/293,768; 60/309,548; 60/314,400; 60/343,706; 60/337,999

<151> 2001-05-25; 2001-08-01; 2001-08-23; 2001-10-19; 2001-12-07

<160> 20

40

<170> PERL Program

<210> 1

<211> 912

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7247177CD1

10

<400> 1

Met Glu Ala Ala Val Lys Glu Glu Ile Ser Val Glu Asp Glu Ala

1 5 10 15

Val Asp Lys Asn Ile Phe Arg Asp Cys Asn Lys Ile Ala Phe Tyr

20 25 30

Arg Arg Gln Lys Gln Trp Leu Ser Lys Lys Ser Thr Tyr Arg Ala

35 40 45

Leu Leu Asp Ser Val Thr Thr Asp Glu Asp Ser Thr Arg Phe Gln

50 55 60

Ile Ile Asn Glu Ala Ser Lys Val Pro Leu Leu Ala Glu Ile Tyr

65 70 75

Gly Ile Glu Gly Asn Ile Phe Arg Leu Lys Ile Asn Glu Glu Thr

80 85 90

Pro Leu Lys Pro Arg Phe Glu Val Pro Asp Val Leu Thr Ser Lys

95 100 105

Pro Ser Thr Val Arg Leu Ile Ser Cys Ser Gly Asp Thr Gly Ser

110 115 120

Leu Ile Leu Ala Asp Gly Lys Gly Asp Leu Lys Cys His Ile Thr

125 130 135

Ala Asn Pro Phe Lys Val Asp Leu Val Ser Glu Glu Glu Val Val

140 145 150

Ile Ser Ile Asn Ser Leu Gly Gln Leu Tyr Phe Glu His Leu Gln

20

30

40

155	160	165
Ile Leu His Lys Gln Arg Ala Ala Lys	Glu Asn Glu Glu Glu Thr	
170	175	180
Ser Val Asp Thr Ser Gln Glu Asn Gln	Glu Asp Leu Gly Leu Trp	
185	190	195
Glu Glu Lys Phe Gly Lys Phe Val Asp	Ile Lys Ala Asn Gly Pro	
200	205	210
Ser Ser Ile Gly Leu Asp Phe Ser Leu	His Gly Phe Glu His Leu	
215	220	225
Tyr Gly Ile Pro Gln His Ala Glu Ser	His Gln Leu Lys Asn Thr	
230	235	240
Gly Asp Gly Asp Ala Tyr Arg Leu Tyr	Asn Leu Asp Val Tyr Gly	
245	250	255
Tyr Gln Ile Tyr Asp Lys Met Gly Ile	Tyr Gly Ser Val Pro Tyr	
260	265	270
Leu Leu Ala His Lys Leu Gly Arg Thr	Ile Gly Ile Phe Trp Leu	
275	280	285
Asn Ala Ser Glu Thr Leu Val Glu Ile	Asn Thr Glu Pro Ala Val	
290	295	300
Glu Tyr Thr Leu Thr Gln Met Gly Pro	Val Ala Ala Lys Gln Lys	
305	310	315
Val Arg Ser Arg Thr His Val His Trp	Met Ser Glu Ser Gly Ile	
320	325	330
Ile Asp Val Phe Leu Leu Thr Gly Pro	Thr Pro Ser Asp Val Phe	
335	340	345
Lys Gln Tyr Ser His Leu Thr Gly Thr	Gln Ala Met Pro Pro Leu	
350	355	360
Phe Ser Leu Gly Tyr His Gln Cys Arg	Trp Asn Tyr Glu Asp Glu	
365	370	375

10

20

30

40

Gln Asp Val Lys Ala Val Asp Ala Gly Phe Asp Glu His Asp Ile
 380 385 390
 Pro Tyr Asp Ala Met Trp Leu Asp Ile Glu His Thr Glu Gly Lys
 395 400 405
 Arg Tyr Phe Thr Trp Asp Lys Asn Arg Phe Pro Asn Pro Lys Arg
 410 415 420
 Met Gln Glu Leu Leu Arg Ser Lys Lys Arg Lys Leu Val Val Ile
 425 430 435
 Ser Asp Pro His Ile Lys Ile Asp Pro Asp Tyr Ser Val Tyr Val
 440 445 450
 Lys Ala Lys Asp Gln Gly Phe Phe Val Lys Asn Gln Glu Gly Glu
 455 460 465
 Asp Phe Glu Gly Val Cys Trp Pro Gly Leu Ser Ser Tyr Leu Asp
 470 475 480
 Phe Thr Asn Pro Lys Val Arg Glu Trp Tyr Ser Ser Leu Phe Ala
 485 490 495
 Phe Pro Val Tyr Gln Gly Ser Thr Asp Ile Leu Phe Leu Trp Asn
 500 505 510
 Asp Met Asn Glu Pro Ser Val Phe Arg Gly Pro Glu Gln Thr Met
 515 520 525
 Gln Lys Asn Ala Ile His His Gly Asn Trp Glu His Arg Glu Leu
 530 535 540
 His Asn Ile Tyr Gly Phe Tyr His Gln Met Ala Thr Ala Glu Gly
 545 550 555
 Leu Ile Lys Arg Ser Lys Gly Lys Glu Arg Pro Phe Val Leu Thr
 560 565 570
 Arg Ser Phe Phe Ala Gly Ser Gln Lys Tyr Gly Ala Val Trp Thr
 575 580 585
 Gly Asp Asn Thr Ala Glu Trp Ser Asn Leu Lys Ile Ser Ile Pro

10

20

30

40

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

10

20

30

40

Gly Glu Leu Tyr Leu Asp Asp Gly His Ser Phe Gln Tyr Leu His
 815 820 825
 Gln Lys Gln Phe Leu His Arg Lys Phe Ser Phe Cys Ser Ser Val
 830 835 840
 Leu Ile Asn Ser Ser Ala Asp Gln Arg Gly His Tyr Pro Ser Lys
 845 850 855
 Cys Val Val Glu Lys Ile Leu Val Leu Gly Phe Arg Lys Glu Pro
 860 865 870
 Ser Ser Val Thr Thr His Ser Ser Asp Gly Lys Asp Gln Pro Val
 875 880 885
 Ala Phe Thr Tyr Cys Ala Lys Thr Ser Ile Leu Ser Leu Glu Lys
 890 895 900
 Leu Ser Leu Asn Ile Ala Thr Asp Trp Glu Gly Ser
 905 910

10

20

<210> 2

<211> 691

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5596327CD1

<400> 2

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

30

40

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

10

20

30

40

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

10

20

30

40

485	490	495
Thr Lys Thr Lys Asp Gly Phe Thr Val Asn Thr Lys Val Pro Ser		
500	505	510
Leu Lys Asp Gln Gly Lys Glu Tyr Asp Gly Phe Thr Ile Thr Ile		
515	520	525
Thr Gly Asp Lys Val Gly Asn Ile Leu Phe Ser Val Glu Thr Gln		
530	535	540
Thr Thr Glu Glu Arg Thr Gln Leu Tyr His Ala Glu Ile Asp Ala		
545	550	555
Leu Tyr Lys Asp Leu Thr Ala Lys Gly Lys Val Leu Ile Leu Ser		
560	565	570
Ser Glu Phe Gly Glu Ala Asp Ala Val Cys Asn Leu Ile Leu Ser		
575	580	585
Leu Val Tyr Tyr Phe Tyr Asn Leu Met Pro Leu Ser Arg Gly Ser		
590	595	600
Ser Val Ile Ala Tyr Ser Val Ile Val Gly Ala Leu Met Ala Ser		
605	610	615
Gly Lys Glu Val Ala Gly Lys Ile Pro Lys Gly Lys Leu Val Asp		
620	625	630
Phe Glu Ala Met Thr Ala Pro Gly Ser Glu Ala Phe Ser Lys Val		
635	640	645
Ala Lys Ser Trp Met Asn Leu Lys Ser Ile Ser Pro Ser Tyr Lys		
650	655	660
Thr Leu Pro Ser Val Ser Glu Thr Phe Pro Thr Leu Arg Ser Met		
665	670	675
Ile Glu Val Leu Asn Thr Asp Ser Ser Pro Arg Cys Leu Lys Lys		
680	685	690
Leu		

10

20

30

40

<210> 3

<211> 506

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4179651CD1

10

<400> 3

Met Thr Asp Ala Glu Arg Val Asp Gln Ala Tyr Arg Glu Asn Gly

1 5 10 15

Phe Asn Ile Tyr Val Ser Asp Lys Ile Ser Leu Asn Arg Ser Leu

20 25 30

Pro Asp Ile Arg His Pro Asn Cys Asn Ser Lys Arg Tyr Leu Glu

35 40 45

Thr Leu Pro Asn Thr Ser Ile Ile Ile Pro Phe His Asn Glu Gly

50 55 60

Trp Ser Ser Leu Leu Arg Thr Val His Ser Val Leu Asn Arg Ser

65 70 75

Pro Pro Glu Leu Val Ala Glu Ile Val Leu Val Asp Asp Phe Ser

80 85 90

Asp Arg Glu His Leu Lys Lys Pro Leu Glu Asp Tyr Met Ala Leu

95 100 105

Phe Pro Ser Val Arg Ile Leu Arg Thr Lys Lys Arg Glu Gly Leu

110 115 120

Ile Arg Thr Arg Met Leu Gly Ala Ser Val Ala Thr Gly Asp Val

125 130 135

Ile Thr Phe Leu Asp Ser His Cys Glu Ala Asn Val Asn Trp Leu

140 145 150

Pro Pro Leu Leu Asp Arg Ile Ala Arg Asn Arg Lys Thr Ile Val

40

	155	160	165
Cys Pro Met Ile Asp Val Ile Asp His Asp Asp Phe Arg Tyr Glu			
	170	175	180
Thr Gln Ala Gly Asp Ala Met Arg Gly Ala Phe Asp Trp Glu Met			
	185	190	195
Tyr Tyr Lys Arg Ile Pro Ile Pro Pro Glu Leu Gln Lys Ala Asp			
	200	205	210
Pro Ser Asp Pro Phe Glu Ser Pro Val Met Ala Gly Gly Leu Phe			
	215	220	225
Ala Val Asp Arg Lys Trp Phe Trp Glu Leu Gly Gly Tyr Asp Pro			
	230	235	240
Gly Leu Glu Ile Trp Gly Gly Glu Gln Tyr Glu Ile Ser Phe Lys			
	245	250	255
Val Trp Met Cys Gly Gly Arg Met Glu Asp Ile Pro Cys Ser Arg			
	260	265	270
Val Gly His Ile Tyr Arg Lys Tyr Val Pro Tyr Lys Val Pro Ala			
	275	280	285
Gly Val Ser Leu Ala Arg Asn Leu Lys Arg Val Ala Glu Val Trp			
	290	295	300
Met Asp Glu Tyr Ala Glu Tyr Ile Tyr Gln Arg Arg Pro Glu Tyr			
	305	310	315
Arg His Leu Ser Ala Gly Asp Val Ala Val Gln Lys Lys Leu Arg			
	320	325	330
Ser Ser Leu Asn Cys Lys Ser Phe Lys Trp Phe Met Thr Lys Ile			
	335	340	345
Ala Trp Asp Leu Pro Lys Phe Tyr Pro Pro Val Glu Pro Pro Ala			
	350	355	360
Ala Ala Trp Gly Glu Ile Arg Asn Val Gly Thr Gly Leu Cys Ala			
	365	370	375

10

20

30

40

Asp Thr Lys His Gly Ala Leu Gly Ser Pro Leu Arg Leu Glu Gly
 380 385 390
 Cys Val Arg Gly Arg Gly Glu Ala Ala Trp Asn Asn Met Gln Val
 395 400 405
 Phe Thr Phe Thr Trp Arg Glu Asp Ile Arg Pro Gly Asp Pro Gln
 410 415 420
 His Thr Lys Lys Phe Cys Phe Asp Ala Ile Ser His Thr Ser Pro
 425 430 435
 Val Thr Leu Tyr Asp Cys His Ser Met Lys Gly Asn Gln Leu Trp
 440 445 450
 Lys Tyr Arg Lys Asp Lys Thr Leu Tyr His Pro Val Ser Gly Ser
 455 460 465
 Cys Met Asp Cys Ser Glu Ser Asp His Arg Ile Phe Met Asn Thr
 470 475 480
 Cys Asn Pro Ser Ser Leu Thr Gln Gln Trp Leu Phe Glu His Thr
 485 490 495
 Asn Ser Thr Val Leu Glu Lys Phe Asn Arg Asn
 500 505

10

20

<210> 4

<211> 276

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7482109CD1

<400> 4

Met Asn Ser Lys Arg Met Leu Leu Leu Val Leu Phe Ala Phe Ser
 1 5 10 15
 Leu Met Leu Val Glu Arg Tyr Phe Arg Asn His Gln Val Glu Glu

30

40

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

Asp Phe Ile Lys Glu Tyr Leu Asn Gly Val Ile His Asp Ile Lys
 245 250 255
 Asn Gly Leu Asn Ser Thr Tyr Glu Lys His Leu Asn Lys Tyr Phe
 260 265 270
 Tyr Leu Asn Lys Pro Thr
 275

<210> 5

10

<211> 378

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2867618CD1

<400> 5

20

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

30

40

110	115	120
Asn Leu Val Pro Thr Ile Leu Asp Lys	Ala His Lys Tyr Asn Leu	
125	130	135
Lys Val Thr Phe His Ile Glu Pro Tyr	Ser Ile Glu Met Ile Lys	
140	145	150
His Val Gln Asn Val Lys Tyr Ile Ile	Asp Lys Tyr Gly Asn His	
155	160	165
Pro Ala Phe Tyr Arg Tyr Lys Thr Lys	Thr Gly Asn Ala Leu Pro	
170	175	180
Met Phe Tyr Val Tyr Asp Ser Tyr Val	Thr Lys Pro Glu Lys Arg	
185	190	195
Ala Asn Leu Leu Thr Thr Ser Gly Ser	Arg Ser Ile Arg Asn Ser	
200	205	210
Pro Tyr Gly Gly Leu Phe Ile Ala Leu	Leu Val Glu Glu Glu His	
215	220	225
Lys Tyr Asp Ile Leu Gln Arg Val Asp	Gly Ile Tyr Thr Tyr Phe	
230	235	240
Ala Thr Asn Gly Phe Thr Tyr Gly Ser	Ser His Gln Asn Gly Ala	
245	250	255
Ser Leu Lys Leu Leu Cys Asp Lys Tyr	Asn Leu Ile Phe Ile Pro	
260	265	270
Ser Val Gly Pro Gly Tyr Ile Asp Thr	Ser Ile Arg Pro Trp Asn	
275	280	285
Thr Gln Asn Thr Arg Asn Arg Ile Asn	Gly Lys Tyr Tyr Glu Ile	
290	295	300
Gly Leu Ser Ala Ala Leu Gln Thr Arg	Pro Ser Leu Ile Ser Ile	
305	310	315
Thr Ser Phe Asn Glu Trp His Glu Gly	Thr Gln Ile Glu Lys Ala	
320	325	330

10

20

30

40

Val Pro Lys Arg Thr Ser Asn Thr Val Tyr Leu Asp Tyr Arg Pro
 335 340 345
 His Lys Pro Gly Leu Tyr Leu Glu Leu Thr Arg Lys Trp Ser Glu
 350 355 360
 Lys Tyr Ser Lys Glu Arg Ala Thr Tyr Ala Leu Asp Arg Gln Leu
 365 370 375
 Pro Val Ser

10

<210> 6

<211> 458

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7488348CD1

<400> 6

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

20

30

40

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

10

20

30

40

Asp Trp Gly Ala Trp Lys Lys Leu Phe Thr Gly Ser Leu Val Gly
 320 325 330
 Ile Gln Val Val Gly Asp Asp Leu Thr Val Thr Lys Pro Glu Ala
 335 340 345
 Arg Ile Ala Lys Ala Val Glu Glu Val Lys Ala Cys Asn Cys Leu
 350 355 360
 Leu Leu Leu Lys Val Asn Gln Ile Gly Ser Val Thr Glu Ser Leu
 365 370 375
 Gln Ala Cys Lys Leu Ala Gln Ser Asn Gly Trp Gly Val Met Pro
 380 385 390
 Val Ser His Arg Leu Ser Gly Glu Thr Glu Asp Thr Phe Met Ala
 395 400 405
 Asp Leu Val Val Gly Leu Cys Thr Gly Gln Ile Lys Thr Gly Pro
 410 415 420
 Thr Cys Arg Ser Glu Arg Leu Ala Lys Tyr Asn Gln Leu Leu Arg
 425 430 435
 Ile Glu Glu Ala Glu Ala Gly Ser Lys Ala Arg Phe Ala Gly Arg
 440 445 450
 Asn Phe Arg Asn Pro Arg Ile Asn
 455

10

20

30

40

<210> 7

<211> 402

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5539166CD1

<400> 7

Met Met Gly Ser Trp Lys His Cys Leu Phe Ser Ala Ser Leu Ile

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

10

20

30

40

Val Gly Arg Thr Lys Tyr Val His Gln Glu Leu Leu Asn His Lys
 230 235 240
 Asn Ser Tyr Val Ile Lys Thr Thr Lys Leu Lys Thr Pro Pro Pro
 245 250 255
 His Asp Met Val Ile Tyr Leu Gly Thr Ala Tyr Val Ala Leu Thr
 260 265 270
 Arg Asp Phe Ala Asn Phe Val Leu Gln Asp Gln Leu Ala Leu Asp
 275 280 285
 Leu Leu Ser Trp Ser Lys Asp Thr Tyr Ser Pro Asp Glu His Phe
 290 295 300
 Trp Val Thr Leu Asn Arg Ile Pro Gly Val Pro Gly Ser Met Pro
 305 310 315
 Asn Ala Ser Trp Thr Gly Asn Leu Arg Ala Ile Lys Trp Ser Asp
 320 325 330
 Met Glu Asp Arg His Gly Gly Cys His Gly His Tyr Val His Gly
 335 340 345
 Ile Cys Ile Tyr Gly Asn Gly Asp Leu Lys Trp Leu Val Asn Ser
 350 355 360
 Pro Ser Leu Phe Ala Asn Lys Phe Glu Leu Asn Thr Tyr Pro Leu
 365 370 375
 Thr Val Glu Cys Leu Glu Leu Arg His Arg Glu Arg Thr Leu Asn
 380 385 390
 Gln Ser Glu Thr Ala Ile Gln Pro Ser Trp Tyr Phe
 395 400

10

20

30

40

<210> 8

<211> 109

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500341CD1

<400> 8

Met Ala Thr Gly Thr Asp Gln Val Val Gly Leu Gly Leu Val Ala

1 5 10 15

Val Ser Leu Ile Ile Phe Thr Tyr Tyr Thr Ala Trp Val Ile Leu

20 25 30

Leu Pro Phe Ile Asp Ser Gln His Val Ile His Lys Tyr Phe Leu

35 40 45

Pro Arg Ala Tyr Ala Val Ala Ile Pro Leu Ala Ala Gly Leu Leu

50 55 60

Leu Leu Leu Phe Val Gly Ser Phe Pro Tyr Pro Asp Thr Ala Pro

65 70 75

Asp Pro Thr Lys Pro Ser Gly Pro Gly Lys Pro Gly Glu Gly Leu

80 85 90

Thr Asp Pro Arg Thr Ser Pro Gly Ala Gln Gly Ser His Pro Gly

95 100 105

Pro Leu Thr Asp

<210> 9

<211> 225

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500846CD1

<400> 9

Met Leu Ser Met Leu Arg Thr Met Thr Arg Leu Cys Phe Leu Leu

1 5 10 15

10

20

30

40

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

⟨210⟩ 10

40

<211> 463

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3230921CD1

<400> 10

Met Ser Glu Glu Glu Ala Ala Gln Ile Pro Arg Ser Ser Val Trp

1 5 10 15

Glu Gln Asp Gln Gln Asn Val Val Gln Arg Val Val Ala Leu Pro

20 25 30

Leu Val Arg Ala Thr Cys Thr Ala Val Cys Asp Val Tyr Ser Ala

35 40 45

Ala Lys Asp Arg His Pro Leu Leu Gly Ser Ala Cys Arg Leu Ala

50 55 60

Glu Asn Cys Val Cys Gly Leu Thr Thr Arg Ala Leu Asp His Ala

65 70 75

Gln Pro Leu Leu Glu His Leu Gln Pro Gln Leu Ala Thr Met Asn

80 85 90

Ser Leu Ala Cys Arg Gly Leu Asp Lys Leu Glu Glu Lys Leu Pro

95 100 105

Phe Leu Gln Gln Pro Ser Glu Thr Val Val Thr Ser Ala Lys Asp

110 115 120

Val Val Ala Ser Ser Val Thr Gly Val Val Asp Leu Ala Arg Arg

125 130 135

Gly Arg Arg Trp Ser Val Glu Leu Lys Arg Ser Val Ser His Ala

140 145 150

Val Asp Val Val Leu Glu Lys Ser Glu Glu Leu Val Asp His Phe

155 160 165

10

20

30

40

Leu Pro Met Thr Glu Glu Glu Leu Ala Ala Leu Ala Ala Glu Ala
 170 175 180
 Glu Gly Pro Glu Val Gly Ser Val Glu Asp Gln Arg Arg Gln Gln
 185 190 195
 Gly Tyr Phe Val Arg Leu Gly Ser Leu Ser Ala Arg Ile Arg His
 200 205 210
 Leu Ala Tyr Glu His Ser Val Gly Lys Leu Arg Gln Ser Lys His
 215 220 225
 Arg Ala Gln Asp Thr Leu Ala Gln Leu Gln Glu Thr Leu Glu Leu
 230 235 240
 Ile Asp His Met Gln Cys Gly Val Thr Pro Thr Ala Pro Ala Arg
 245 250 255
 Pro Gly Lys Val His Glu Leu Trp Gly Glu Trp Gly Gln Arg Pro
 260 265 270
 Pro Glu Ser Arg Arg Arg Ser Gln Ala Glu Leu Glu Thr Leu Val
 275 280 285
 Leu Ser Arg Ser Leu Thr Gln Glu Leu Gln Gly Thr Val Glu Ala
 290 295 300
 Leu Glu Ser Ser Val Arg Gly Leu Pro Ala Gly Ala Gln Glu Lys
 305 310 315
 Val Ala Glu Val Arg Arg Ser Val Asp Ala Leu Gln Thr Ala Phe
 320 325 330
 Ala Asp Ala Arg Cys Phe Arg Asp Val Pro Ala Ala Ala Leu Ala
 335 340 345
 Glu Gly Arg Gly Arg Val Ala His Ala His Ala Cys Val Asp Glu
 350 355 360
 Leu Leu Glu Leu Val Val Gln Ala Val Pro Leu Pro Trp Leu Val
 365 370 375
 Gly Pro Phe Ala Pro Ile Leu Val Glu Arg Pro Glu Pro Leu Pro

10

20

30

40

380	385	390
Asp Leu Ala Asp Leu Val Asp Glu Val Ile Gly Gly Pro Asp Pro		
395	400	405
Arg Trp Ala His Leu Asp Trp Pro Ala Gln Gln Arg Ala Trp Glu		
410	415	420
Ala Glu His Arg Asp Gly Ser Gly Asn Gly Asp Gly Asp Arg Met		
425	430	435
Gly Val Ala Gly Asp Ile Cys Glu Gln Glu Pro Glu Thr Pro Ser		
440	445	450
Cys Pro Val Lys His Thr Leu Met Pro Glu Leu Asp Phe		
455	460	

10

<210> 11

<211> 3150

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7247177CB1

<400> 11

```

aaataaatca ttgtagaaac gctagtttgg gcctgaaaaa ttccaggagc aagagtcaag 60
atttgtcact ccatgagaat ctggagggga ctcccttccc agaaacttga cgatgaagta 120
ctggttgtaa ttttagaaag acaccaatc ggctttttta aaagatcgcc cagggccctt 180
gtcctgagag ctgggagctg gtcggagtga cagagaagcc atggaagcag cagtgaagaa 240
ggaaataagt gttgaagatg aagctgtaga taaaaacatt ttcagagact gtaacaagat 300
cgcatittac aggcgtcaga aacagtggct ttccaagaag tccacctatc gggcattatt 360
ggattcagtc acaacagatg aagacagcac caggttccaa atcatcaatg aagcaagtaa 420
ggttcctctc ctggctgaaa tttatgggtat agaaggaaac attttcaggc ttaaaattaa 480
cgaagagact cctctaaaac ccagatttga agttccggat gtcttcacaa gcaagccaag 540
cactgtaagg ctgatttcat gctctgggga cacaggcagt ctgatattgg cagatggaaa 600

```

30

40

aggagacctg aagtgccata tcacagcaaa cccattcaag gtagacttgg tgtctgaaga 660
 agagggtgtg attagcataa attccctggg ccaattatac tttagagcatc tacagattct 720
 tcacaaacaa agagctgcta aagaaaatga ggaggagaca tcagtggaca cctctcagga 780
 aatcaagaa gatctgggcc tgtgggaaga gaaatttggg aaatttgtgg atatcaaagc 840
 taatggccct tcttctattg gtttggattt ctcttgcac ggatttgagc atctttatgg 900
 gatccacaa catgcagaat cacaccaact taaaaatact ggtgatggag atgcttaccg 960
 tctttataac ctggatgtct atggatacca aatataatgat aaaatgggca tttatggttc 1020
 agtaccttat ctcttggccc acaaactggg cagaactata ggtattttct ggctgaatgc 1080
 ctccgaaaca ctgggtggaga tcaatacaga gcctgcagta gactacacac tgaccagat 1140
 gggcccagtt gctgctaaac aaaaggtcag atctcgcact catgtgcact ggatgtcaga 1200
 gagtggcatc attgatgttt tctgtctgac aggacctaca ccttctgatg tcttcaaaca 1260
 gtactcacac cttacaggca cacaagccat gccccctctt tctcttttgg gataccacca 1320
 gtagccgttg aactatgaag atgagcagga tgtaaaagca gtggatgcag ggtttgatga 1380
 gcatgacatt ccttatgatg ccatgtggct ggacatagag cacactgagg gcaagaggta 1440
 cttcaccitg gacaaaaaca gattccaaa cccaagagg atgcaagagc tgctcaggag 1500
 caaaaagcgt aagcttgtgg tcatcagtga tccccacatc aagattgatc ctgactactc 1560
 agtataatgt aaggccaaag atcagggctt ctttgtgaag aatcaggaag gggaagactt 1620
 tgaaggggtg tgttggccag gtcctctctc ttacctggat ttaccaatc ccaaggtcag 1680
 agagtggat tcaagtcttt ttgctttccc tgtttatcag ggatctacgg acatctcttt 1740
 cctttggaat gacatgaatg agccttctgt ctttagaggg ccagagcaaa ccatgcagaa 1800
 gaatgccatt catcatggca attgggagca cagagagctc cacaacatct acggttttta 1860
 tcatcaaatg gctactgcag aaggactgat aaaacgatct aaagggaagg agagaccctt 1920
 tgttcttaca cgttctttct ttgctggatc acaaaagtat ggtgccgtgt ggacaggcga 1980
 caacacagca gaatggagca acttgaat tcttatccca atgttactca ctctcagcat 2040
 tactgggatc tctttttgcg gagctgacat aggcgggttc attgggaatc cagagacaga 2100
 gctgctagtg cgttgggtacc aggcctggagc ctaccagccc ttcttccgtg gccatgccac 2160
 catgaacacc aagcgacgag cgccctggct ctttggggag gaacacaccc gactcatccg 2220
 agaagccatc agagagcgt atggcctcct gccatatagg tattctctgt tctacatgc 2280
 acacgtggct tccaacctg tcatgaggcc tctgtgggta gagttccctg atgaactaaa 2340

10

20

30

40

gacttttgat atggaagatg aatacatgct ggggagtgca ttattggttc atccagtcac 2400
 agaaccacaaa gccaccacag ttgatgtggt tcttcagga tcaaatgagg tctggatga 2460
 ctataagaca ttgtctcatt ggggaaggagg gtgtactgta aagatcccag tagccttgga 2520
 cactattcca gtgtttcagc gaggtggaag tgtgatacca ataaagacaa ctgtaggaaa 2580
 atccacaggc tggatgactg aatcctccta tggactccgg gtgtctctaa gcactaaggg 2640
 ttcttcagtg ggtgagttat atcttgatga tggccattca tccaatacc tccaccagaa 2700
 gcaatitttg cacaggaagt tttcattctg ttcagtggt ctgatcaata gttctgtga 2760
 ccagaggggt cattatccca gcaagtgtgt ggtggagaag atcttggctt taggcctcag 2820
 gaaggagcca tcttctgtga ctaccacac atctgatggt aaagatcagc ctgtggcttt 2880
 tacgtattgt gccaaaacat ccctctgag cctggagaag ctctcactca acattgccac 2940
 tgactgggag ggttctctaa aaagctctaa aatgaatgca tagtgtgtct tctctcaag 3000
 atgtaaatgt caaggaatga gttggcaaat cagtttaaaa acacaaacta cccaacagaa 3060
 aaatccagaa gtccccatc cagttcgtct ggaagagacc tccggaaata tgcgcttgcc 3120
 agcacactgc ggccgtaacg tgagcggccc 3150

10

20

<210> 12

<211> 3290

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5596327CB1

<400> 12

ggtctaaggt agcgagcggc ttgtctgtct ctgtctgttc tggggcggcg ctgtgccgcg 60
 cggcgccgcc cgggttgtct gctgtgtctg ctgtctgggg tctgtccgcc gggctgcggc 120
 aggcgcccctg gccaccgagc actactcgcc gtctccctgc tcaagcagga gctgcagacc 180
 ggcaagagagg aggccccggg gcggggggct gagcccgagt ccgggggact ggggggacca 240
 gtactctgcc gagtgcggcg agtcattctt ttgaacttc catgactcag actgcgaacc 300
 caagggatca tcacctgtg actccttctt tccctcaac actgagaaga ttctgagcca 360
 ggccaagtct attgcagaac agaagagatt cccgtttgcc actgataatg acagcacaaa 420

30

40

tgaagagtta gctattgctt atgtcttgat tggcagtggt ctgtatgatg aagcaatagc 480
 gcatttttca acaatgcttc aggaggagcc tgatctgggt agtgcaattt atggccgagg 540
 gatagcctat ggaaagaagg gactacatga cattaagaat gctgagcttg ctctgttcga 600
 actgagccga gtaattacct tgggaaccaga tcgtccagag gtatttgagc agcgagcaga 660
 aattcigtcc cctctgggac gaattaatga agcagtgaat gacctacta aagctatcca 720
 actgcagccc tcagcacggc tgtacagaca tcggggaacc ctgtacttca taccagagga 780
 ctaigcaaca gcccataag actttcagca gtcccttagaa ctgaacaaaa accagcctat 840
 agctaigcta tacaaagggt taactttctt tcacagagga cttctgaagg aagctattga 900
 atccttcaaa gaagctttga agcagaaagt tgactttatt gatgcatata aaagtctagg 960
 gcaggcataat agagaactgg gcaattttga agcagccact gagagcttcc aaaaggccact 1020
 gtgtctcaac caaaatcatg tgcaaaccct ccagctccgg ggaatgatgc tctaccacca 1080
 cggcagctta caggaagccc ttaagaactt taagcggigt ctgcagctag agccatataa 1140
 tgaagigtgc cagtatatga aagggtcag ccatgttgcc atgggacagt tttatgaagg 1200
 gataaaaagca caaacaaaag ttaigtctga tgatcctctc ccaggccaga aggctagccc 1260
 tgagtatctt aaagttaaagt atctccgaga gtattctcga tatcttcatg cacaccttga 1320
 tccccccctt acggaatata acattgatgt ggatctgcct ggaagcttta aggaccactg 1380
 ggctaaaaat ttgcttttcc tcatagaaga ctacgaagag cagccagggt tgcaacccca 1440
 cataaaaagat gtgttacatc agaattttga gagttataag cctgaagtac aggagctgat 1500
 ttgtgtggct gatcgitttg gatccctgat gcaatatgaa acacctgggt tcttgccaaa 1560
 caagagaata cacagagcta tgggtttggc cgcattggag gtcattgcaag ccgtgcagcg 1620
 tacatggacc aactcgaaag ttcgaatgaa tgggaagaca cggttgatgc agtggagaga 1680
 catgtttgac attgcagtta aatggagaag gattgctgac ccagaccagc ccgtgctgtg 1740
 gttagatcaa atgccagcac gaagtcttag cagaggtttt aacaaccaca ttaatttaaat 1800
 caggggacag gtgatcaaca tgagataacct agaatatitt gagaaaatac ttcattttat 1860
 taaagacaga attcttgitt atcatggagc taataatcct aaaggattgc tgggaagttcg 1920
 ggaagccctg gaaaaggtag acaaagtaga agaccttctt ccgattatga agtttaatac 1980
 taaaacgaag gatgggttca ccgtgaacac aaaagttccc agccttaaag accaaggga 2040
 ggaatatgat ggattcaca tcacgattac aggagacaaa gttggcaata tattattttc 2100
 tigtgaaact caaaccacgg aagaaaggac acaattatat catgctgaaa tagatgcact 2160

10

20

30

40

ttataaagat ttgacagcaa aaggaaaagt attgattcctt tcatcagaat ttggggaggc 2220
 tgatgctgtc tgcaacttaa tcttatecct agtttattac ttttataatt taatgccact 2280
 cctcgcagga tccagtgtaa ttgcttactc ggcatcgtg ggagcactga tggcaagtgg 2340
 aaaagaagta gcaggaaaaa ttcccaaagg gaagttagtc gactttgaag ctatgacagc 2400
 ccttggttca gaggccttta gcaaagtcgc caaaagctgg atgaacttga aaagtatttc 2460
 accttcttat aagactcttc catcagtttc agaaacgttt ccaacgttaa gatcgatgat 2520
 tgaggigcta aacacagact cttctccacg ttgtcttaag aaactctagt tctgctgctg 2580
 tatttataca agtaaagggc cggacctctt gcttcttaag ttatttttta aaacatggaa 2640
 ttataaagaa attaggctct ctattacttt tgataccaat ttttatagga attgaatact 2700
 tgaaatcatt ttctttactt tctgaacca taatgaaaat catgaaagag gctttcgggg 2760
 ggaaaaagct acttgacttg gagggaaagc gtatgtatta gccggcagcc tgttggttcc 2820
 cacttataaa gtcatataac tcttctctta cagtaaaaaa aattttcttc aacaactaaa 2880
 atataatact gaagatgaac aaaaacaaaa ccaaatagga gigtittgtat tcttttagag 2940
 atctggatat ttttcaaaga cttgcttttt acgcacattg tgatagagca tctttigtgt 3000
 actcacgtaa cctccatgtg tgggtccttt gtittattag aaccatggaa gaaggttcct 3060
 gggccatcag tatgtgttgc agagctcaca gccaggaaca citattttga acaatgcaga 3120
 cagtacttgc cagagtccca aatgctgaat taitgaacca tgtttttttt ttctttgttt 3180
 gaaaaaaaata aaccagggtga tttctataaa tcaccagttt ttggaaaaaa aaaaaaaaag 3240
 gggcccgccg aaaaggagc tctgaaccc gggaataaat ccgggaccgg 3290

10

20

30

40

<210> 13

<211> 4279

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4179651CB1

<400> 13

ggagaagcgg ctctgcagg cggtagcgct ggtgctggcg gccctgggtc tcttgcctaa 60
 cgtggggctt tgggcgctgt accgcgagcg gcagcccgac ggcacccctg tgtgggatcg 120

ggggcgtgtg tgtgtgtgtc cgtgtgtgtg gtgacatggg ctacacatg tcgacaaaat 180
 gaaaacgttt ttcttgttga gatgggcaga agctgaagt actggcatga caatggaggc 240
 catccggaig ggacgctcag cgcgtaggaa atggagaaca aggaagacct taccatga 300
 ccgatgtga gagagtggat caggcatacc gagaaaatgg atttaacatc tacgtcagt 360
 ataaaatctc ctigaatcgc tctctcccag atatccggca cccaaactgc aacagcaagc 420
 gctacctgga gacacttccc aacacaagca tcatcatccc ctccacaac gagggctggt 480
 cctccctcct ccgcaccgtc cacagtgtgc tcaatcgctc gcctccagag ctggctgccc 540
 agattgtact ggtcgacgac ttcagtgtc gagagcacct gaagaagcct ctigaagact 600
 acatggccct ttccccagt gtgaggattc ttgaaccaa gaaacgggaa gggctgataa 660
 ggacccgaat gctgggggccc tcagtggcaa ctggggatgt catcacattc ttggattcac 720
 actiggaagc caatgtcaac tggcttcccc cctgtctga ccgcatgtc cggaaccgca 780
 agaccattgt gtgcccgatg attgatgtaa ttgacatga cgactttcgg tacgagacac 840
 aggcagggga tgcctatcgg ggagcctttg actgggagat gtactacaag cggatccgca 900
 tccctccaga actgcagaaa gctgacccca gcgaccatt tgagtctccc gtgatggccg 960
 gggactgtt cggcgtggat cgggaagtgt tctgggaact cggcgggtat gaccaggct 1020
 tggagatctg gggaggggag cagtatgaaa tctcttcaa ggtgtggatg tgtggggccc 1080
 gcatggagga catccctgc tccagggtgg gccatatcta caggaagtat gtgccctaca 1140
 aggtccccgc cggagtcagc ctggccccga accttaagcg ggtggccgaa gtgtggatgg 1200
 atgagtacgc agagtacatt taccagcgc ggctgaata ccgccacctc tccgtgggg 1260
 atgtcgcagt ccagaaaaag ctccgcagct ccttaactg caagagtctc aagtggttta 1320
 tgacgaagat agctgggac ctgcccatt tctaccacc cgtggagccc ccggctgcag 1380
 ctgggggga gatccgaat gtgggcacag ggctgtgtgc agacacaaag cagggggcct 1440
 tgggtcccc actaaggcta gagggctgcg tccgaggccg tggggaggct gcctggaaca 1500
 acatgcaggt attcacttc acctggagag aggacatccg gcctggagac cccagcaca 1560
 ccaagaagtt ctgcttgat gccatttccc acaccagccc tgtcacgtg tacgactgcc 1620
 acagcatgaa gggcaaccag ctgtggaaat accgcaaaga caagacctg taccacctg 1680
 tcagtggcag ctgcatggac tgcagtgaat gtgacatag gatcttcatt aacacctgca 1740
 acccatctc tctaccag cagtggctgt ttgaacacac caactcaaca gtcttgaaa 1800
 aattcaatag gaactgagcc ctcatgtccc ctggcaggc ccccagggt ctggcactca 1860

10

20

30

40

ctgcagactt cctctttcaa gggaggcagg gccctgtgg gcactagggtg taaaagggtgc 1920
 tggccaaatg gttcagggtg aagagggtc ttgattcagg ggctggggtc tgcctgggtcc 1980
 ttgagccccct gagttgtggg ggtagggtga agagcatac ccacaagagg cccacagggtg 2040
 agcagagact gctttaatcc ctgctgacat cacggaaaag caacagagcc ttttcaactt 2100
 tgtcactatg tccccitgaa cattatgtgg gagaacacca aggtagccta ggccacccaa 2160
 aagttagtcc tgcgaggttg cccagccctc agatggctct cctacatgat ggtgctttag 2220
 aaacaaagggt aaaatttgcc tgtttggggc agcttttagt atcgatgcca ctcatctgca 2280
 gcagaagaga aagaagtcct cttggggctt tttagtttct gccgtcctgg ggggaacatt 2340
 gcagttactg cacagcttct gttctctgtc acaaccccag gtgatttggc ccgggtcaaag 2400
 gccatacttg gggccctaag agtggtcagt attgaatgct gatcagctgc cagggtagga 2460
 gtcagaagag ggagccccc tagacatttc ttgacagcta tggacatgcg ggatactcc 2520
 cctgtctctc tgggtatttg aaatgtcaat tttagcactc tccaggcaca aggacagccc 2580
 agcaccagct ttacagggtc gtgtttcaga tggccctgag cccacggaaa aggccaggta 2640
 gacctccaaa ctagaaatgc tggttgattt gccctgatcc atgcttccat tccccgtct 2700
 ctcttcccca ggcaattact ggcttcaaaa gaggaacaga ggtgtctgca ggtgtctacc 2760
 tcacagagtc tggaggctc caggatcaac tgtgggcaaa gtgcctgcct ctgacctcat 2820
 catggttcta gttctcatac agaactccag aattttttaa gaactctata attggattgc 2880
 aaactaggat gctacatagg attctgggtat tccacatcca atatggattt ctagaatgct 2940
 gtgattaaag gagccagcca ggtgtaatc agtcaaggca gccccagcc tagagacaat 3000
 ctgtgaaatc caaagttagt ggtgttggga aagcaggggg acatgtgtcc ctacagctcag 3060
 cagaggctgt ggtacaacat ggtccttggc gaagacctgc acccctggaa cctcccacca 3120
 tcatcacaac tgtagtctca ttgacagtgg agaaaagaac ccgacgtccc acagccagat 3180
 atacacccag ctccatgcca gcccttcatg ttacctttt gctttgttaa ttacatgtca 3240
 gactcctaga gggcctccag actaatagga agcatttctg taaccaacct gccaccact 3300
 gattcagaaa tggaaatcac attccacaat ctatggcttc caccagctag cccaggaaat 3360
 acttgaaatc agcattccaa ttagtgttga gtctcttgat tgtgtcattt accaattaaa 3420
 taactgagac ctaagtctgg gaacagagcc acgaatctgc ctttgagatg ctggcagatc 3480
 tcaaggccat caattattgg gggaggagg gacaaacact cccaatcacc caccagtcag 3540
 actgaatgtg tagctggcga ggaattactt ccacttctgg cccagcaca gccctgtctt 3600

10

20

30

40

ggccacctgt ctgcaagaga ggccggccct gtgcttgcaa cgcttacgtg ttgatcccag 3660
 tgtccitttc caaatgagtg ctgtagcttt agaagtggcc cttatataga agaagtcaaa 3720
 agatgaggcc ctttctagaa tctaggataa caagagtgtt gacagtitta ggagtcgaat 3780
 tgagattcat catcaaagag caatgcagcg tcgttaaaat aaaaactgtg ctttttaaaa 3840
 agaaaaatgc aaatatagag caaatcccta aacttgaacc atttcctagt gccttgctag 3900
 acaaacataa aggggcaggg ttgtggggag gggaacgttt ttgggtggtg gtgcagttcc 3960
 tactggatga gtgtgtgttt cttaatgtct catttcaagc aggagatgtt gggctcggag 4020
 caagatctga gactgagatg tcccccaagg gtgacaggtc agatttattt cagtcagtgc 4080
 aagtcacact ccaaactaaa ggctgggcag tgcgtttagt ctgtggccta gaagacctca 4140
 tcagcccccg agaggaggcc ttgcattacc tctctgggac ctgtttgggg gccatccctg 4200
 cggtcacccc tctgaacccc cagcatgttt gtctgtttc tcaccattgg gtttaaggggt 4260
 gccgagcact agctagact 4279

10

<210> 14

<211> 1042

20

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7482109CB1

<400> 14

atgaattcta aaagaatgct gttattgggt ttatttgctt ttctactgat gttgggttag 60
 cgttatttca ggaatcacca agtagaagaa cticggctct cagactgggt tcatcctaga 120
 aaacgccctg atgttataac gaaaacagac tggctcgctc ctgtcctatg ggaagggact 180
 ttcgacaggc gggctcctgga aaaacattac agaaggcggg atatcactgt gggcctggcc 240
 gtctttgcta ctggcagggt tgcagaggag tacctgaggc cgttcctaca ctccgcaaat 300
 aagcacttca tgacaggcta ccgagtgatc ttctacatca tgggtggacgc cttcttcaag 360
 ctgccigaca tagagcccag tctctttcga acgttcaaag catttaaagt gggcaccgag 420
 aggtgggtggc tcgatggccc cctgggtgcat gtgaagagcc tgggtgaaca catcgccagt 480
 cacatccagg acgagggtga ctctctcttc agcatggctg ccaaccaggt ctccagaat 540

30

40

gagttcgggg tggagaccct gggcccgttg gtggcccagc tccacgcctg gtggtatttc 600
 agaaacacca agaacttccc ttatgagagg aggccgacct cagcagcttg catcccgttt 660
 ggacagggag atttctatta tggcaacttg atggttgggtg gcacacccca taatatitta 720
 gacttcatca aagaatatct gaacggagtt attcatgaca tcaaaaatgg actcaatagc 780
 acttaigaaa agcaccttaa caaatatttt tacttcaata aaccaccta gctgttatca 840
 ccagcataca gctgggactt tgcattttct cctcctccac agatccaata cgtcaaggtc 900
 gcacatgatt cccagaggaa attatgaatt acgccatgag gtttatgacg aataaatgaa 960
 tcacggcaat ttcttataaa tgagggaagt cttagacct tccatttttt agagacatac 1020
 aaatttacct cctcccaaaa aa 1042

10

<210> 15

<211> 4320

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2867618CB1

<400> 15

attcatttgg ggaaaaattt tgatttccaa agagtgcacg aatcaacagt gaaacgaata 60
 ccaagaattt aaaaagtgtt gaaatcacia tgaacccttc caaggcctct gaacttiaact 120
 tggatgaact accacctctg aacatttata tacatgtatt ttattacagt tggtaatggaa 180
 atccacaatt tgatggtaaa tatatacatt ggaatcatcc agtgtagag cattgggacc 240
 ctagaatagc caagaattat ccacaaggga gacacaaccc tccagatgac attggctcca 300
 gcttttatcc tgaattggga agtiacagtt ctggggaacc ttctgtcata gaaactcaca 360
 tgagacaaat gcgctcagct tcaattgggtg tactagccct ctcttgggtac ccacctgatg 420
 taaatgatga aaatggagaa cctactgata acttgggtacc cactattttg gataaagctc 480
 ataaatataa cctaaagggt acttttcaca tagaaccata tagcatcgag atgatcaaac 540
 atgtacaaaa tgtcaagtat attatagaca aatatggaaa tcatccggcc ttttacaggt 600
 acaagacgaa gactggcaat gctcttccca tgttttatgt ctatgattcc tatgtgacca 660
 agcctgaaaa aagggccaat ctgitaacca cctcagggtc tcggagtatt cgcaattctc 720

30

40

cctaagggtgg actgittatt gcccttctgg tagaagaaga acataagtat gatattcttc 780
aaaggggtga tggaaattac acatattttg ccacaaatgg ctttacttat ggctcatcac 840
atcagaatgg ggctagccta aaattattgt gtgataaata caacttaata tttatcccaa 900
gtgtggggccc aggatacata gataccagca tccgtccatg gaacacgcaa aacactcgga 960
accgaatcaa tgggaagtat tatgaaatg gtctgagtgc cgcacttcag acacgcccc 1020
gcttaatttc tatcacctct tttaatgagt ggcatgaagg aactcagatt gaaaaagctg 1080
ttcccaaaaag aaccagtaat acagtgtacc tagattaccg tctcataaaa ccaggctttt 1140
acctagaact gactcgcaag tggctgaaa aatacagtaa ggaaagagca acttaigcat 1200
tagatcgcca gctgcctgtt tcttaatgca ttgattaaag tttaatagtt atcaaaatca 1260
cctaattttt aaaaatagct ttcgttttga gtcttgaaa gaaaactgtc aaaatcagta 1320
tatactatta gttatattta aaaatatttt tttaaattct ttacagataa tattatactt 1380
gttacccttc acaataccac atgagaaaat atctgagaca aaatgtatac aaatatattc 1440
cctaaggcat aatttattgc atttctgact gaaatcaaaa ttctgatttg atggcaattg 1500
aattttcatt ttacaataga taaatgcttg tgttacctaa agcacttagc acacagttaa 1560
attatattta catcctagac ccaaataaat aggatttgtt gtatatttgg gatattctatt 1620
gaagaaaaaa agaaaacccc ttaaagataa tgtacatgct tcatgtcatg tctttaaaat 1680
aatttaatca acttattgtt cttagtattt agactctgga taactctaca ataatgagga 1740
aattcittaag aataacaaaa tcaactgtacc ttcctctcaa ttttgctgtg aacctgaaat 1800
ggctttaaat taatactctt attttttatt taatttaatt acataaatta aaccttacca 1860
tgaccaaatt gtgttaggaa ggcttgcctat ctacagcaca gtgtgtcatt tgcagatttg 1920
tggttaccta taccacgcta ggtgttttga catgtttagt gtttctgctt tacagtgcctg 1980
aattccatat tttagaagct atgaaagtec ttttatgaaa aagttactga ttgcttctca 2040
gttattagga aaacagttgt ttcacaattia ttatgtagat atgatgccca aatatcattt 2100
ttagtatatc ttgtcgatct ttaagttgtt actatttgtt tattcatgtc tttaaatcag 2160
ataccaaata ttttttagga aagaaaaatg ttattactgt cattaggttg tcttttaata 2220
ctttaagtta ttttgacgaa aagtaataga gaaaatttac ttagcatttt agattctaga 2280
gacatggaaa tgaaaattat tttatgtctt gagtaggtcc tgaagtttgg ctttacatta 2340
agtttagcac tgtatcagaa tgaagaaact aatattttac ataaaaacta atactttcaa 2400
ttttttatat agtaatatcc ccattttgtt aatgttagac ttttatcata cctgtaagtt 2460

10

20

30

40

aaaatacttg ttatcaataa ctgtcatgt tgtgacaaat tgatcacttg tgtacgaaaa 2520
 ataaaatctcc ttaaaaacta aataaaatgc actgtattct tacagttaat gtttataact 2580
 atagtaaaaa attaatatat atcctattac ataaatgita tttcttaggt gtccattaa 2640
 gaagagcaat agaataatgc taaaaaataa tgcctataaa tcttcagagt ataaagacat 2700
 ccattcagaa acaaaaatta gcactaaatt ttttataaaa tagaccagat gacaaaattt 2760
 attttatttt taaacagtgg ttttgacaca aattatgita ttgaaaagca ttattaatgt 2820
 ttaatttatt taaaattttg gaatttgcca tttctcagac aatgatcagg ctttaggaaa 2880
 ttaatacagt agtagtaatc attttctagg ggaaaataaa agaataaatc actatactga 2940
 tatttigata taagcaagca ctacatgggt aatcactata tagatccaac ctgtggattt 3000
 tcttctatg tccatttaac tagaatatat tatttttaggt ataatttaca aatgtcacac 3060
 ctaataatct tttataatat accatatttc attaaagttt tgttagagaa gtatctacca 3120
 cagaggagtt tttgtcattg tgtacgttgt gtatttgaac ccacatgac agaaagtaaa 3180
 ttttaggaaa tagttatgag attaaggga aatctataaa aacaaggita gcataattctc 3240
 aacacagata ccaccattt ctttttccca ttatagacat ggtgaatcca cacagcatac 3300
 ttcatctctg agctttgttg tgattctca acacattacc ctaaccagcc agcagtaaca 3360
 gatttcagag taagataaag cagattctgt cticattgca aaaagttatt ctcaatggaa 3420
 gaatggcatc tgatctcata attactagtt tatattaata tagttttttt ctcccttttt 3480
 aataaaaataa ttacagtcat ccttcagtgt ctgtggggga ttggttccag ttaccectat 3540
 agataccaaa atctgcagat gctcaagtc ctgatataaa ctggcatagt agttgcataat 3600
 aatctatgca catctctctg tatacattaa gtcactctta gattatttat aacacttaat 3660
 acaatgtaaa tgctatgtag ttgittatacc atactggita gtgaataatt acatgaaaaa 3720
 aaagagtctg tacatcttca gagtttcagt cggcaatttc ttggccatgg atgtagaacc 3780
 tacagataag gtgagccaac tgcattagga aataactcta ataattctgt taattcttag 3840
 agaggaaaaac tttcaaaatc ttcctcaggt atttattaca actgccitta ccattttagt 3900
 igtaacacag tttaaattgt tatgataaca agtaaataag agcaaagaat ttatttctta 3960
 attcaaaact atacgtttga attcaataat gtataactta aagtggitata atacatacaa 4020
 tgcattgaatc ataatggatt cttttataag ttattaattt ttatggitta atcagtctaa 4080
 ttgttttgac tgttatagaa accaaatatt ttactgttct ttttaaggac taatattgtc 4140
 aaaaactgct gttattaact tcacttgagt tgtttaactt cttctgttt taagattgta 4200

10

20

30

40

attaaaaatt actatittgt tatatggaat ggtaatttt tacctaataa aaacatagat 4260
 gaaatacatt gtatittaga tattaaagtc tatcatttgc tctttaaaaa aaacaaaaaa 4320

<210> 16

<211> 1733

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7488348CB1

<400> 16

atcggatctg agcgaacgga acgggtgcgg gtgttcaaga tgtccattct caagatcattc 60
 catgcccgtg atatctttga atcccgtggg aatccaactg ttgaggtaga cctttacact 120
 aacaagggtg gtcgttcgg aagagctgct gticctagcg gtgcctcaac tggaaattat 180
 gaagctctgc tggaaattcg ggacaatgac aagacacgct acatgggggg gaagggtgtc 240
 tcaaaagctg ttgagcacat tatcaataaa acaattgcac ccgcactgat tagcaagaat 300
 gtcaatgtag tggagcaaga caagatcgat aacctgatgc tggacatgga tggatcagag 360
 aacaaatcta aatttgggtc caacgccatc ttgggtgtat ctctggctgt atgctcgaat 420
 gctgggtgcta ctgctgagaa ggggtgcccc ttgtaccgtc acattgctga ccttgcctggg 480
 aataaccag aagtcattct gccgtgtccc gcctttaacg tgatcaacgg tggctcccat 540
 gctggcaata agctggctat gcaggagttc atgateccct cctgtgggtg tgacaggttt 600
 aacgacgcaa tccgcattgg tgcagagggt taccacaacc tgaagaatgt catcaaggag 660
 aaatatggga aagatgccac caatgtgggg gatgaaggcg ggtttgctcc caacatcctg 720
 gagaataaag aaggcctgga gctgctgaag actgctattg ggaaagctgg ctacactgat 780
 aagggtgtca tcggcatgga cgtagcggcc tccgagttct tcaggctctgg gaagtatgac 840
 ctggacttca agtctccga tgacccagc aggtacatct cgctgacca gctggctgac 900
 ctgtacaagg gctttgtcct ggggcatgca gtgaagaact acccagtggg cgtctccatc 960
 gaggaccccc cattigacca ggaigactgg ggtagctgga agaagctgtt tactggcagc 1020
 ctggttggca tccaagtggg tggagacgat ctgaccgtga ctaagcccgga agcgcgtatt 1080
 gccaaaggctg tggaggaggt taaagcctgc aactgcctcc tctctctcaa ggtcaaccag 1140

10

20

30

40

attggatctg tgacggagtc cctacaagct tgcaagcttg cccagtcctaa cggctggggc 1200
 gtgatccccg tgagtcaccg cctctccgga gaaacagaag ataccctcat ggctgacctc 1260
 ggggtcgggc tctgcactgg tcagatcaaa actgggtccca ctgcccgatc tgagcgtcta 1320
 gccaaagtaca accagctgct gaggattgaa gaggctgagg ctggcagcaa ggcccgtttt 1380
 gctggaagaa acttcaggaa cccccgtatc aactaagctg cgtggatcga caccctttct 1440
 ggtaagttaa gcactagtca cctacttaga ctcacaatta ctgtattag aagatgaggg 1500
 gcaggctgaa gaaaagacca gtttgcaggt cctctccctt cctagatgac tctctttcac 1560
 ctacagtgtt tccaccagct ctgactgtt acttgtaacg atcgtgcttt gtagaacaat 1620
 cccagctttt gatgtttggg aggcgttttg acttgcagga cagcaagacg gtacctacaa 1680
 aacagctagt agtgitttta catgtgataa ataaaaagca tcaaacaaaa aaa 1733

<210> 17

<211> 2201

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5539166CB1

<400> 17

gacagggtta agcaattttt ttaaagcact acatatttat tgatagaata tctctccagg 60
 cagtattcac atgcittaaa gtgaagctcg tagcttggca gtgcagitta agcaaataaa 120
 ctccaggctgt tgaacacaat gattaacagg aagcagggga ggggacatag ttactagact 180
 ctggccggga caagaggctg aggttaaggt cttagagccg atgacttgca gaatggtcac 240
 ctccaccact tgcaatttca tggtaggaac aggtggacct cctggaatca gaacctctct 300
 gaggacatct gtttttgtgt agacacaggt tgcaggttag caggagaaca ggcaagccaa 360
 atgcaaagga gccacttcag aaatgtgtca cagaaaagtg aaaatgcaac ctagtggtaa 420
 gtgaagaggg gaagaagaaa gaaaaaggac cagaaccgtg aactgaaggg acggggaaca 480
 gccagacgag agcttcagcc atcacgagga tgatttcgga acctggagaa aatgtaagtt 540
 aaatatatct acactctgat cctatctcaa gagagagata ttttactcat ttcttggttg 600
 tgaatgatgg gctcttggaa gcactgtctt ttagcgcgt ctcttatctc tgccctgatt 660

10

20

30

40

ttgtatttg tttacaatac tgagttatgg gagaataaac gttttctgag ggcagctctg 720
 tccaatgctt cactgttagc agaagcctgt catcagattt ttgaggggaa agttttttac 780
 ccaacagaaa atgcattgaa aactaccctt gatgaagcta cctgctatga gtacatgggt 840
 cgaagccact atgtaacaga aacactctct gaagaagagg ctgggttccc tttagcttac 900
 acagtgacca tccacaaaga cttcggcact ttigagaggc tcttcagggc gatttatatg 960
 ccccaaaatg tctactgtgt gcacctggat cagaaggcga cggatgcctt taaagggtgca 1020
 gtgaaacagt tactcagctg ctcccaaatt gcttttctgg ctccaagaa ggagtcgggt 1080
 gtctatgggg ggatctccag gctccaggt gacctgaact gcctggaaga ccttgiggcc 1140
 tctgaagttc cctggaagta tgtcatcaac acctgcgggc aagactttcc cctgaaaacc 1200
 aacagggaaa tagttcagta tctgaaggga tttaaaggga aaaataacac ccccgagtg 1260
 ctgctctctg accacgctgt tggacggact aaatacgtcc accaagaact gttaaaccac 1320
 aaaaattcct acgtgattaa aacaacaaaa ttaaaaactc ctctctctca tgacatgggt 1380
 attaccttg gcacggccta cgtggctctc acaagggact ttgctaactt cgctctccaa 1440
 gaccagctcg cacttgactt actctcttgg tccaaggaca cctacagccc cgacgaacat 1500
 ttctgggtga cactcaacag gattcccgtt gtctctggct ctatgccaaa tgcactctgg 1560
 actggaaacc tcagagctat aaagtggagt gacatggaag acagacacgg aggcctgccac 1620
 ggccactatg tacatgggtat ttgtatctat ggaaacggag acttaaagtg gctgggtaat 1680
 tcaccaagcc tgtttgctaa caagtttgag cttaatacct acccccttac tgtggaatgc 1740
 ctagaactga ggcatcgga aagaacctc aatcagagt aaactgcgat acaaccagc 1800
 tggatTTTT gagctattca tgagctactc atgactgaag ggaaactgca gctgggaaga 1860
 ggagccgtgt ttgtgagag actttgcct tcgtaatgtt aaccgtttca ggaccaggt 1920
 tatagcttca ggacctggct acgttaattat acttaaaata tccactggac actgtgaaat 1980
 acactaacag gatggctggg tagagcaatc tgggcacitt ggccaatttt agtcttgctg 2040
 tttcttgatg ctcacctcta tattagttaa ttgttaggat catgataatt taatgacctc 2100
 agatcttggg cgatactctt ctccgggtt ttctaccgg gatgtgttct ctgctcgcat 2160
 acaatcgtcc tccgagggt tttaccttca gcccggggcc c 2201

<210> 18

<211> 457

<212> DNA

10

20

30

40

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500341CB1

<400> 18

```
gcttgcggct cgggtggctg agcgcgcggg gaaatggcca cggggacaga ccaggtaggtg 60
ggactcggcc tcgtcgccgt tagcctgac atcttcacct actacaccgc ctgggtgatt 120
ctcttgccat tcatcgacag tcagcatgtc atccacaagt atttctgcc ccgagcctat 180
gcgtcgcca tccactggc tgcaggcctc ctgctgtcc tgtttgtggg cagctttcct 240
tacctgaca cagccccaga cccacaaaag cttctggac ctggaaagcc tggggaagga 300
ctgacagacc ccaggaccag ccttggggct cagggcagcc accccgggcc gctgaccgac 360
tgacctctcc tcacggaggc ccagccccaa agccccaggg ctggcccggt tgggacagct 420
gaccaataaa cactgatggt gtgtttaaaa aaaaaaa 457
```

<210> 19

<211> 1513

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500846CB1

<400> 19

```
ggagctccga gtgtccacag gaagggaact atcagctcct ggcatctgta aggatgctgt 60
ccaigctgag gacaatgacc agactctgct tctgttatt cttctctgtg gccaccagt 120
ggtagcagtgc agcagcagcc tcttctcttg agatgtcttc gagggaaattc gaaacctgtg 180
ccttctcctt ttcttcctg cctagaagct gcaaagaaat caaggaacgc tgccatagt 240
caggtagatgg cctgtatctt ctcgcacca agaattggtgt tgtctaccag accttctgtg 300
acaigacttc tgggggtggc ggctggacc tggtagccag cgtgcacgag aatgacatgc 360
gtgggaagtg cacggtgggt gatcgtggt ccagtcagca gggcaacaaa gcagactacc 420
cagaggggga tggcaactgg gccaaactaca acaccttgg atctgcagag gcggccacga 480
```

gcgatgacta caagaacctt ggctactacg acatccaggc caaggacctg ggcatctggc 540
 atgtgcccac caagtcctcc atgcagcatt ggagaaacag cgccttgctg aggtaccgca 600
 ccaacactgg ctctctccag agactgggac ataatctgtt tggcatctac caggggaatt 660
 tgttgacgga ttcgttcagt tccgggtgtt taataacgag agagcagcca acgccccttg 720
 tgcctgggata aaagttactg gctgtaacac tgagcatcac tgcctcggtg gaggagggtt 780
 ctctccacag ggcaaacccc gtcagtgtgg ggacttctcc gcctttgact gggatggata 840
 tggaaactcac gtaagagca gctgcagtcg ggagataacg gaggcggctg tactcttgtt 900
 ctatagatga gacagagctc tgcgggtgtc gggcgagaac ccattctcca accccggcta 960
 ttgggagacg gaaaaactgg aattctaaca aggaggagag gagactaaat cacatcaatt 1020
 tgcaaaaaaa aaaaaagggg gggccgctct aaagaacccc ccaagggggc caaccitacc 1080
 ctgtcatgca attcataacc tctccccctc tgtaaaaagt gagaaggggc cggggccaga 1140
 acctctgggg tgccccacg ccccgggtgg tggccgatt gaccggttct ggaaggggta 1200
 accagtgggc ctittgtaaaa aggcctggag ttgccaagg tgtgctgatg gggtaggaaa 1260
 agacagttag gaggatatac ttcagagtgc ctgagaagcg cctataaaac atccctacga 1320
 agtggaaaag ccgggtgtgt gaataacacc ctgcccagc tccagaagct gcctctgtga 1380
 ccgctctaa aattcgcatg gaccgattgg aacctcccc aagggggaagt tgttgittagt 1440
 aggtgcaaca gccagggga atctgagatc attaccaac tgtatatccc actaccittg 1500
 ttgcgatgta ctc 1513

10

20

30

40

<210> 20

<211> 2084

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3230921CB1

<400> 20

ggccggaccg gcgggggtgt ggagactcga gcctggggtc ggccggagaca gctgggtgtc 60
 gaagccgctc gcgcccaggg tgacctgtt tgcagcacga tgtctgaaga agaggcggct 120
 cagatcccca gatccagtgt gtgggagcag gaccagcaga acgtgggtgca gcgtgtgggt 180

gctctgcccc tggtcagggc cacgtgcacc gcggtctgcg atgtttacag tgcagccaag 240
 gacaggcacc cgtctgtggg ctccgcctgc cgcctggctg agaactgcgt gtgcggcctg 300
 accacccgtg ccttggacca cggccagccg ctgctcgagc acctgcagcc ccagctggcc 360
 actatgaaca gcctcgccctg caggggcttg gacaagctgg aagagaagct tccctttctc 420
 cagcaacctt cggagacggt ggtgacctca gccaaaggac tggtagccag cagtgtcacg 480
 ggtgtgggtg acctggcccc gaggggcccg cgttggagcg tggagctgaa gcgctccgtg 540
 agccaigctg tggatgttgt actggaaaaa tcagaggagc tggtagatca ctccctgccc 600
 atgacggagg aagagctcgc ggcactggcg gctgaggctg aaggccctga agtgggttcg 660
 gtggaggatc agaggagaca gcagggtac ttigtgcgcc tcggctccct gtcagcacgg 720
 atccgccacc tggcctacga gcactctgtg gggaaactga ggcagagcaa acaccgtgcc 780
 caggacacc tggcccagct gcaggagacg ctggagctga tagaccacat gcagtgtggg 840
 gtgaccccca ccgccccggc ccgcccggg aaggtgcacg agctgtgggg ggaatggggc 900
 cagcgcctc cggagagccg ccgcccggagc caggcagagc tggagacgct ggtgtgttcc 960
 cgcagcctga cccaggagct gcagggcacg gtagaggctc tggagtccag cgtgcggggc 1020
 ctgcccggcg gcgcccagga gaagggtggc gaggtgcggc gcagtgtgga tgccctgcag 1080
 accgccttcg ctgatgcccg ctgcttcagg gacgtgccag cggccgcgct ggccgagggc 1140
 cggggtcgcg tggcccacgc gcacgcctgc gtggacgagc tgctggagct ggtgggtgcag 1200
 gccgtgccgc tgccctggct ggtgggaccc ttgcgcacca tccctgtgga gcgacccgag 1260
 cccctgcccg acctggcgga cctgggtggc gaggtcatcg ggggccccta ccccgctgg 1320
 gcgcacctgg actggccggc ccagcagaga gcctgggagg cagagcacag ggacgggagt 1380
 gggaaatggg atggggacag gatgggtgtt gccggggaca tctgcgagca ggaacccgag 1440
 acccccagct gcccggtaaa gcacacctg atgcccagc tggacttctg acccatgggc 1500
 cagtggaggc ggggaggaaa ggccacctgc acacccgat cctgtgtgcc cctgggtggc 1560
 cacacgtaag ctcgaggcct tggccttgac cttctttgg aatcaggccc aactccgat 1620
 ctctgaccac ctttttgta ttggactctc ccatttttc ctgaaacaca tggacaaaga 1680
 ggccccgggg agcagggcct cgaacctat tcaggccaac ttgagccaca agctgggttc 1740
 ttacctaag tctgtctccc tggctccaig aagcgaatcc aaatctttcc aagaggctgg 1800
 gcacagtggc tcacgcctgt aatcccagca ctttgggagt ctgaggcagg tggatcatct 1860
 gaagtcagga gttcgggac atcctggcca acatgatgaa accctgtctc tacttagaaa 1920

 gacagacaaa aaaaaaagg ggggggcccg cgattagga gctcgtggac ccggggatta 1980
 aatccgggac cggcccggcg ggggtaccag tcccctaata gggagcgtat aaagctgggg 2040
 aaccaaggca tacgggtcgt gtggaattgt tccgtcaat caaa 2084

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
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 02/097060 A2

(51) International Patent Classification: C12N

(21) International Application Number: PCT/US02/18354

(22) International Filing Date: 22 May 2002 (22.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/293,768 25 May 2001 (25.05.2001) US
60/309,548 1 August 2001 (01.08.2001) US
60/314,400 23 August 2001 (23.08.2001) US
60/343,706 19 October 2001 (19.10.2001) US
60/337,999 7 December 2001 (07.12.2001) 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): SWARNAKAR,
Anita [CA/US]; 8 Locksley Avenue #5D, San Francisco,
CA 94122 (US); GORVAD, Ann, E. [US/US]; 369 Marie
Common, Livermore, CA 94550 (US); HAFALIA, April,
J., A. [US/US]; 2227 Calle de Primavera, Santa Clara,
CA 95054 (US); DUGGAN, Brendan, M. [AU/US]; 243
Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US);
EMERLING, Brooke, M. [US/US]; 1735 Woodland
Avenue # 71, Palo Alto, CA 94303 (US); ISON, Craig, H.
[US/US]; 1242 Weathersfield Way, San Jose, CA 95118
(US); NGUYEN, Dannie, B. [US/US]; 1403 Ridgewood
Drive, San Jose, CA 95118 (US); LEE, Ernestine, A.
[US/US]; 20523 Crow Creek Road, Castro Valley, CA
94552 (US); YUE, Henry [US/US]; 826 Lois Avenue,
Sunnyvale, CA 94087 (US); YUE, Huijin [US/US];
1170 South Sielling Road, Cupertino, CA 95014 (US);
FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue,
Redwood City, CA 94061 (US); LI, Joana, X. [US/US];1264 Geneva Avenue, San Francisco, CA 94112 (US);
THANGAVELU, Kavitha [IN/US]; 1950 Montecito
Avenue #23, Mountain View, CA 94043 (US); WALIA,
Narinder, K. [US/US]; 890 Davis Street A # 205, San
Leandro, CA 94577 (US); BURFORD, Neil [GB/US];
105 Wilkewood Circle, Danbury, CT 06422 (US); MASON,
Patricia, M. [US/US]; 360 Clarke Lane, Morgan Hill, CA
95014 (US); LAL, Preeti, G. [US/US]; P.O. Box 5142,
Santa Clara, CA 95056 (US); LEE, Sally [US/US]; 3643
26th Street, San Francisco, CA 94110 (US); BECHA,
Shanya, D. [US/US]; 21062 Gary Drive # 117, Castro
Valley, CA 94546 (US); TANG, V. Tom [US/US]; 4230
Rainwick Court, San Jose, CA 95118 (US).(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(81) Designated States (national): AF, AG, AI, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PI, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PL, SE, TR), OAPI patent
(BJ, BJ, CI, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).Published:
without international search report and to be republished
upon receipt of that 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.

WO 02/097060 A2

(54) Title: CARBOHYDRATE-ASSOCIATED PROTEINS

(57) Abstract: The invention provides human carbohydrate-associated proteins (CHOP) and polynucleotides which identify and encode CHOP. 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 CHOP.

WO 02/097060

PCT/US02/18354

CARBOHYDRATE-ASSOCIATED PROTEINS**TECHNICAL FIELD**

This invention relates to nucleic acid and amino acid sequences of carbohydrate-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of carbohydrate-associated proteins.

BACKGROUND OF THE INVENTION

Carbohydrates, including sugars or saccharides, starch, and cellulose, are aldehyde or ketone compounds with multiple hydroxyl groups. Carbohydrates have three important roles in mammalian cells. Carbohydrates function as energy-storage molecules, as fuels, and as metabolic intermediates. Carbohydrates are broken down to release energy in glycolysis or may be stored as glycogen for later use. The importance of carbohydrate metabolism is demonstrated by the sensitive regulatory system in place for maintenance of blood glucose levels. Two pancreatic hormones, insulin and glucagon, promote increased glucose uptake and storage by cells, and increased glucose release from cells, respectively. The sugars deoxyribose and ribose form part of the structural support of DNA and RNA, respectively, providing a second example of carbohydrate function. Third, carbohydrates provide a means for post-translational modification of secreted and membrane proteins and lipids. Indeed, 2-10% of the content of eukaryotic cell membranes are contributed by oligosaccharides on membrane glycoproteins and glycolipids. Carbohydrate modifications on glycoproteins and glycolipids create great structural diversity, and since they are mainly located on the extracellular side of the plasma membrane, they play an important role in intercellular recognition (Stryer, L. (1988) *Biochemistry*, W.H. Freeman and Company, New York NY, pp. 298-299, 331-347).

Proteins are associated with carbohydrates in several ways. Carbohydrate-containing macromolecules, which include glycoproteins, glycolipids, glycosaminoglycans, and proteoglycans, are found on the cell surface and in the extracellular matrix. The extracellular matrix is composed of diverse glycoproteins, and carbohydrate-binding proteins which are secreted from the cell and assembled into an organized meshwork in close association with the cell surface. The interaction of the cell with the surrounding matrix profoundly influences cell shape, strength, flexibility, motility, and adhesion. These dynamic properties are intimately associated with signal transduction pathways controlling cell proliferation and differentiation, tissue construction, and embryonic development.

Glycoproteins have covalently attached carbohydrates which have been added to the proteins as they traverse the secretory pathway. Some proteins noncovalently associate with carbohydrate-

WO 02/097060

PCT/US02/18354

containing macromolecules for purposes of binding, modifying, or degrading the carbohydrates. Glycoproteins include cell adhesion molecules, receptors, blood group antigens, growth factors, and antibodies. These proteins are involved in cellular processes such as cell-cell recognition and signaling, recognition and/or destruction of neurotransmitters, transmission of neural impulses, and immune function.

Oligosaccharide modifications can provide great structural diversity. N- and O-linked oligosaccharides are transferred to proteins and modified in a series of enzymatic reactions that occur in the endoplasmic reticulum (ER) and Golgi. Oligosaccharides stabilize the protein during and after folding, orient the protein in the membrane, improve the protein's solubility, and act as a signal for lysosome targeting.

Heavily glycosylated glycoproteins are also referred to as proteoglycans. Proteoglycans in the extracellular matrix of connective tissues such as cartilage are essential for distributing the load in weight-bearing joints. Cell-surface-attached proteoglycans anchor cells to the extracellular matrix. Both extracellular and cell-surface proteoglycans bind growth factors, facilitating their binding to cell-surface receptors and subsequent triggering of signal transduction pathways (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY, pp. 1139-1142).

Carbohydrates also form glycosaminoglycans (GAGs), which are linear unbranched polysaccharides composed of repetitive disaccharide units. GAGs exist free or as part of proteoglycans, large molecules composed of a core protein attached to one or more GAGs. GAGs are found on the cell surface, inside cells, and in the extracellular matrix. The GAG hyaluronan (HA) is found in the extracellular matrix of many cells, especially in soft connective tissues, and is abundant in synovial fluid (Pitsillides, A.A. et al. (1993) *Int. J. Exp. Pathol.* 74:27-34). HA, which functions in water and plasma protein homeostasis, seems to play important roles in cell regulation, development, and differentiation.

Glycolipids, along with phospholipids and cholesterol, form the membranes of cells. Examples of glycolipids include blood group antigens on erythrocytes and gangliosides in the myelin sheath of neurons. Modifications to glycoproteins and glycolipids on the extracellular side of the plasma membrane are important for intercellular recognition (Stryer, *supra*, pp. 298-299, 331-347; Lodish, et al., *supra*, pp. 612-615).

Lectins are extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells (Drickamer, K. and Taylor, M.E. (1993) *Annu. Rev. Cell Biol.* 9:237-264). This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L.A. (1991) *J. Cell. Biochem.* 45:139-146; Palletta, E. et al. (1989) *J. Immunol.* 143:2850-2857).

WO 02/097060

PCT/US02/18354

Lectins are classified into subfamilies based on carbohydrate-binding specificity. The galectin subfamily, in particular, includes lectins that bind β -galactoside carbohydrate moieties in a thiol-dependent manner (Hadari, Y.R. et al. (1995) *J. Biol. Chem.* 270:3447-3453). Galectins are widely expressed and developmentally regulated. Because all galectins lack an N-terminal signal peptide, it is suggested that galectins are externalized through an atypical secretory mechanism. Two classes of galectins have been defined based on molecular weight and oligomerization properties. Galectins contain a characteristic carbohydrate recognition domain (CRD), also known as a galaptin domain, which is about 140 amino acids long and contains several conserved residues (See Prosite PDOC00279 Vertebrate galactoside-binding lectin signature).

Another example is intelectin, a Ca^{2+} dependent lectin that binds to galactofuranosyl residues and bacterial arabinogalactan. Intelectin may play a role in the recognition of bacterial carbohydrate and induction of the immune response to microorganisms.

Carbohydrate-modifying enzymes

The enzyme glutamine:fructose-6-phosphate amidotransferase (GFAT), also known as aminotransferase, catalyzes the reversible reaction of L-glutamine and D-fructose-6-phosphate to form L-glutamate and D-glucosamine-6-phosphate, which is the rate-limiting step in the hexosamine biosynthetic pathway (ExPASy ENZYME: EC 2.6.1.16). D-glucosamine-6-phosphate acts in the biosynthesis of UDP-N-acetyl-glucosamine (UDP-GlcNAc) and other hexosamines incorporated into glycoproteins and proteoglycans. GFAT regulates the availability of precursors for N- and O-linked glycosylation. Glucosamine enhances the production of transforming growth-factor (TGF)- β 1 (Kolm-Litty, V. et al. (1998) *J. Clin. Invest.* 101:160-169). GFAT activity plays a role in insulin resistance in Type II diabetes, and GFAT overexpression leads to insulin resistance. Hexosamine metabolism appears to regulate glycogen synthase, the rate-limiting enzyme in glycogen synthesis, as well as PPIG, a glycogen-bound protein phosphatase, pyruvate kinase, and the glucose transporter GLUT1 (McClain, D.A. and Crook, E.D. (1996) *Diabetes* 45:1003-1009).

The enzyme glucosamine-6-phosphate deaminase (GNPDA), also known as isomerase, catalyzes the reversible reaction of D-glucosamine-6-phosphate with water to form D-fructose-6-phosphate and ammonia (ExPASy ENZYME EC 5.3.1.10). This reaction links hexosamine systems with glycolytic pathways and may provide an energy source from the catabolism of hexosamines in glycoproteins, glycolipids, and sialic-acid-containing macromolecules. GNPDA is expressed in tissues with high energy requirements (Wolosker, H. et al. (1998) *FASEB J.* 12:91-99).

The enzyme UDP-glucose dehydrogenase (UDPGD) catalyzes the reversible reaction of UDP-glucose, 2 NAD^+ , and water to form UDP-glucuronate and 2 NADH (ExPASy ENZYME EC 1.1.1.22). UDP-glucuronate is needed for the biosynthesis of GAGs, which appear to play a role in signal transduction pathways (Binari, R.C. et al. (1997) *Development* 124:2623-2632).

WO 02/097060

PCT/US02/18354

Man₅-mannosidase is an α 1,2-mannosidase (glycosyl hydrolase) involved in the early processing of N-linked oligosaccharides. This enzyme catalyzes the specific cleavage of α 1,2-mannosidic linkages in Man₅-(GlcNAc)₂ and Man₅-(GlcNAc)₂. Multiple α 1,2-mannosidases have been identified in mammalian cells and may be needed for the processing of distinct classes of N-glycoproteins. Man₅-mannosidase is a Type II membrane protein with a short cytoplasmic tail, a single transmembrane domain, and a large luminal catalytic domain. The human kidney enzyme is localized to the Golgi (Bause, E. et al. (1993) Eur. J. Biochem. 217:535-540; Bieberich, E. and Bause, E. (1995) Eur. J. Biochem. 233:644-649).

DPM1 is an enzyme in the endoplasmic reticulum that catalyzes the production of dolichol phosphate-mannose (DPM) from GDP-mannose and dolichol phosphate. The activity of DPM1 is regulated by DPM2, which targets DPM1 to the endoplasmic reticulum (ER) and increases its affinity for dolichol phosphate. DMP2 resides in the (ER) membrane and contains two putative transmembrane domains and a putative ER-localization signal near its C-terminus.

Glycosylation

Glycosylation refers to the covalent attachment of any number of carbohydrate chains (oligosaccharides) to specific sites (glycosylation sites) on proteins. Glycosylation is a post-translational protein modification essential to the conformation, stability, transport, secretion, antigenicity, clearance and activity of glycosylated proteins (glycoproteins). The composition of the attached oligosaccharides is specific to a protein and may be simple (consisting primarily of mannose residues) or complex (with additional N-acetyl-glucosamine (GlcNAc), sialic acid, and galactose residues). Glycoproteins may have relatively few carbohydrate groups or may contain a larger percentage of carbohydrate than protein (based on molecular weight). These latter, heavily glycosylated glycoproteins are also referred to as proteoglycans to emphasize the predominant carbohydrate composition of the molecules. The type of saccharide bond (e.g., α , β , 1,2-, 1,4-) formed between any two constituent carbohydrate residues is also a critical molecular determinant for the structure and function of the glycoprotein.

Glycosylation confers increased hydrophilicity to proteins. Many glycoproteins, such as carrier proteins, antibodies, and lysosomal proteins, are found free in solution (e.g., plasma). Other glycoproteins are membrane-bound. In the case of membrane-associated glycoproteins, the carbohydrate side-chain serves to orient the glycoproteins in the membrane lipid bilayer. The glycosylated regions of the molecule interact with the aqueous environment on the inside or outside of the membrane while the more hydrophobic domains of the glycoprotein (typically consisting of non-polar amino acid residues that are not glycosylated) interact with the phospholipids in the membrane.

Addition of oligosaccharide side chains occurs at the -NH₂ group of asparagine (Asn)

WO 02/097060

PCT/US02/18354

residues (N-linked glycosylation) or at the -OH group of serine (Ser) residues (O-linked glycosylation). The process of N-linked glycosylation begins in the endoplasmic reticulum (ER) and is completed in the Golgi apparatus of eukaryotic cells. O-linked glycosylation occurs exclusively in the Golgi. Of all characterized glycoproteins, 90% are N-glycosylated, with or without additional O-glycosylation. Only 10% are exclusively O-glycosylated (Apweiler, R. et al. (1999) *Biochim. Biophys. Acta* 1473:4-8). Almost two-thirds of the approximately 75,000 SWISS-PROT protein sequences include putative N-glycosylation sites, underscoring the importance of this protein modification in nature (Apweiler et al., *supra*). Biochemical steps involved in N-linked glycosylation have been well characterized, and are reviewed below.

10 N-linked glycosylation

N-linkage of carbohydrates to proteins occurs via a nitrogen atom of asparagine (Asn) residue side-chains in the context of the primary amino acid sequence Asn-X-Ser or Asn-X-Thr (Ser = serine, Thr = threonine, and X = any amino acid residue except proline). While the composition of N-linked oligosaccharides is highly diverse, the pathways responsible for glycosylation have common first steps. A 14-residue core oligosaccharide, containing two N-acetylglucosamine (GlcNAc), nine mannose, and three glucose residues, is transferred as a unit from a dolichol phosphate donor molecule to the -NH₂ group of an acceptor Asn residue on the target protein. Typically, the three glucose residues of the core oligosaccharide are removed by glucosidases I and II resulting in "high mannose oligosaccharides" side chains. These partially processed N-linked glycoproteins are then sequentially transported from the ER through the *cis*-, *medial*-, and *trans*-cisternae of the Golgi (Bonay, P. et al. (1996) *J. Biol. Chem.* 271:3719-3726). Further modification to the oligosaccharide chains may occur to remove additional core mannose residues using the enzymes Golgi mannosidase I (*cis*-cisterna), N-acetyl-glucosaminyltransferase (GluNAcT; *medial*-cisterna), and Golgi mannosidase II (*trans*-cisterna). Following the removal of some of the mannose residues by Golgi mannosidase I, the addition of a single GlcNAc by GluNAcT is essential for the removal of the remaining mannose residues of the core oligosaccharide by Golgi mannosidase II.

Mannose-1-phosphate guanylttransferases are involved in early steps of protein glycosylation. They participate in sugar metabolism and their enzymatic products are channeled into glycoprotein synthesis. Mannose-1-phosphate guanylttransferase 1 (MPG1), also referred to as NDP-hexose pyrophosphorylase, catalyzes the conversion of GTP and α D-mannose 1-phosphate into diphosphate and CDP-ethanolamine in mannose metabolism. This enzyme is very similar to CDP-glucose pyrophosphorylase and may also be involved in the regulation of cell cycle progression. A cDNA coding for GTP: α D-mannose-1-phosphate guanylttransferase 1 (MPG1) was recently isolated from a cDNA library of a *Trichoderma reesei* strain (Kruszewska, J.S. et al. (1998) *Curr. Genet.* 33:445-50). The nucleotide sequence of the 1.6 kb cDNA revealed an ORF which encodes a protein of 364 amino

WO 02/097060

PCT/US02/18354

acids. Sequence comparisons demonstrate 70% identity with the yeast *Saccharomyces cerevisiae* glucanyltransferase gene 1 (MPG1) and 75% identity with the *Schizosaccharomyces pombe* homologue.

- Complex oligosaccharide side-chains result from the addition of N-acetyl-glucosamine, N-acetylneuraminic acid (sialic acid), and galactose, as well as other sugar moieties, to the remaining core sugar moieties on the partially-processed glycoprotein. These modifications occur in the *trans*-cisterna and *trans*-Golgi network (TGN), and involve a number of enzymes including N-acetylglucosaminyltransferase I (GlcNAcTsI), sialyltransferases (STs), and galactosyltransferases (GalTs). Multiple isoforms of many of these enzymes produce specific α or β , 1,2-, 1,3-, 1,4-, or 1,6-disaccharide bonds between constituent sugar residues of the oligosaccharide side-chain. The stereochemistry and type of bonds in a carbohydrate side-chain contribute to the overall structure and function of the resulting glycoprotein (Lehninger, A. L. et al. (1993) *Principles of Biochemistry*, Worth Publishers, New York NY, pp. 931; Lewin, B. (1997) *Genes VI*, Oxford University Press, New York NY, pp. 1030-1033).
- Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) *J. Biol. Chem.* 273:433-440; Amado, M. et al. (1999) *Biochim. Biophys. Acta* 1473:35-53). Galactosyltransferases are found in the Golgi, on the cell surface, and as soluble extracellular proteins, in addition to being present in the Golgi.
- β 1,3-galactosyltransferases form Type I carbohydrate chains with Gal (β 1-3)GlcNAc linkages. β 1,3-galactosyltransferases appear to have a short cytosolic domain, a single transmembrane domain, and a catalytic domain with eight conserved regions (Kolbinger et al., *supra*; Hennet, T. et al. (1998) *J. Biol. Chem.* 273:58-65). In mouse, UDP-galactose: β -N-acetylglucosamine β 1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is located at amino acid residues 264-275. A variant of a sequence found within mouse UDP-galactose: β -N-acetylglucosamine β 1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennet et al., *supra*). Recent work suggests that brainiac protein is a β 1,3-galactosyltransferase. (Yuan, Y. et al. (1997) *Cell* 88:9-11; Hennet et al., *supra*).
- UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) catalyzes the formation of Type II carbohydrate chains with Gal (β 1-4)GlcNAc linkages (Sato, T. et al. (1997) *EMBO J.* 16:1850-1857).
- A soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids

WO 02/097060

PCT/US02/18354

conserved among β 1,4-galactosyltransferases include two cysteines linked through a disulfide-bond and a putative UDP-galactose binding site in the catalytic domain (Yadav, S.P. and Brew, K. (1990) *J. Biol. Chem.* 265:14163-14169; Yadav, S.P. and Brew, K. (1991) *J. Biol. Chem.* 266:698-703; Shaper, N.L. et al. (1997) *J. Biol. Chem.* 272:31389-31399). β 1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. A β 1,4-galactosyltransferase functions as part of a heterodimer with α -lactalbumin in mammary lactose production. A β 1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface β 1,4-galactosyltransferases also function in cell adhesion, cell recognition, cell/basal lamina interaction, and normal and metastatic cell migration (Shur, B. (1993) *Curr. Opin. Cell Biol.* 5:854-863; Shaper, J. (1995) *Adv. Exp. Med. Biol.* 376:95-104; Masri, K.A. et al. (1988) *Biochem. Biophys. Res. Commun.* 157:657-663).

Synthetases are another class of carbohydrate-modifying enzymes that have critical roles in proper cell functioning. For example, production of sialylated glycoconjugates requires the synthesis of cytidine 5'-monophosphate N-acetylneuraminic acid (CMP-Neu5Ac or CMP-sialic acid), a reaction catalyzed by CMP-Neu5Ac synthetase (Munster, A.K. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:9140-9145). Sialic acids of cell surface glycoproteins and glycolipids contribute to proper structure and function in a variety of tissues. Sialyltransferases (STs) are a subset of glycosyltransferases that catalyze the transfer of sialic acid (from CMP-sialic acids) to the carbohydrate groups of glycoproteins and glycolipids. A variety of these Type II membrane proteins are present in the Golgi. Cloned members of this gene family share an N-terminal cytoplasmic tail region, a transmembrane region, and a large luminal region containing three sialyl motifs designated large (L), small (S), and very small (VS). The L-sialyl motif contributes to donor substrate binding and consists of eight invariant residues within a highly conserved stretch of 48-49 amino acids. The 23-amino acid S-sialyl motif contributes to the binding of both donor and acceptor substrates (Datta, A. et al. (1997) *Indian J. Biochem. Biophys.* 34:157-65). In the case of a representative sialyltransferase ST3GalII (~350 amino acids in length), the L, S, and VS, regions correspond to amino acids 138-182, 264-286, and 309-321, respectively. Other cloned members of the family include ST6GalNAcI and ST8SiaI. ST6GalNAcI is larger than the other known sialyltransferases, and is composed of more than 500 amino acid residues (Tsuiji, S. et al. (1996) *Glycobiology (letter)* 6:v-vii; Geremia, R. et al. (1997) *Glycobiology (letter)* 7:v-vii; Datta, A. et al. (1995) *J. Biol. Chem.* 270:1497-1500; Datta, A. et al. (1998) *J. Biol. Chem.* 273:9608-9618; Tsuiji, S. et al. (1998) *J. Biochem.* 120:1-13). Sialyltransferases are not abundant in cellular extracts, but several have been cloned and expressed. At least one inhibitor has been synthesized (Horenstein, B. et al. (1996) *J. Am. Chem. Soc.* 118:10371-10379).

A variety of other enzymes that are involved in sugar metabolism participate directly or

WO 02/097060

PCT/US02/18354

indirectly in glycosylation, upstream of events that occur in the ER and Golgi. Many of these enzymes were originally identified in bacteria and plants and are less well characterized in humans; however, human homologues may exist that perform similar functions. For example, ADP-glucose pyrophosphorylases catalyze a very important step in the biosynthesis of α 1,4-glucans (glycogen or starch) in bacteria and plants, namely the synthesis of the activated glucosyl donor, ADP-glucose, from glucose-1-phosphate and ATP. ADP-glucose pyrophosphorylases are tetrameric, allosterically-regulated enzymes. There are a number of conserved regions in the sequence of bacterial and plant ADP-glucose pyrophosphorylase subunits. Additionally, there are three regions which are considered signature patterns. The first two regions are N-terminal and have been proposed to be part of the allosteric and substrate-binding sites in the *Escherichia coli* enzyme. The third pattern corresponds to a conserved region in the central part of the enzymes.

Carbohydrate metabolism disorders

Carbohydrate metabolism is altered in several disorders. Diabetes mellitus is characterized by abnormally high blood glucose (hyperglycemia). Type I diabetes results from an autoimmune-related loss of pancreatic insulin-secreting cells. Type II diabetes results from insulin resistance and impaired insulin secretory response to glucose, and is associated with obesity. Hypoglycemia, or abnormally low blood glucose levels, has several causes including drug use, genetic deficiencies in carbohydrate metabolism enzymes, cancer, liver disease, and renal disease (Berkow, R. et al. (1992) The Merck Manual of Diagnosis and Therapy, Internet Edition, Section 8, Chapter 91, Diabetes Mellitus, Hypoglycemia).

Mutations in enzymes involved in protein glycosylation cause severe diseases. For example, alpha mannosidase mutations cause congenital dyserythropoietic anemia Type I and alpha B lysosomal mannosidosis (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York NY, pp. 2092-2093; and Online Mendelian Inheritance In Man, 224100).

Glucosidases represent another class of carbohydrate-modifying enzymes that catalyze the release of glucose from carbohydrates through hydrolysis of the glycosidic link in various glucosides. The inherited disorder type I Gaucher disease, which is characterized by hematologic abnormalities, can be detected in a heterozygous or homozygous individual through an assay of leukocyte beta-glucosidase levels (Raghavan, S.S. et al. (1980) *Am. J. Hum. Genet.* 32:158-173). Patients with all three types of Gaucher disease exhibit a deficiency of an enzyme called glucocerebrosidase that catalyzes the first step in the biodegradation of glucocerebroside. In the brain, glucocerebroside arises from the turnover of complex lipids during brain development and the formation of the myelin sheath of nerves. In other tissues, glucocerebroside arises mainly from the biodegradation of old red and white blood cells.

WO 02/097060

PCT/US02/18354

Galectins play a number of roles in diseases and conditions associated with cell-cell and cell-matrix interactions. For example, certain galectins associate with sites of inflammation and bind to cell surface immunoglobulin E molecules. In addition, galectins may play an important role in cancer metastasis. Galectin overexpression is correlated with the metastatic potential of cancers in humans and mice. Moreover, anti-galectin antibodies inhibit processes associated with cell transformation, such as cell aggregation and anchorage-independent growth.

Galectin-8, also known as prostate carcinoma tumor antigen 1 (PCTA-1), is a novel galectin implicated in cancer progression (Su, Z.-Z. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:7252-7257). Galectin-8 is expressed in invasive prostate carcinomas and early-stage prostate cancers, but not in normal prostate or benign prostatic hypertrophic tissue.

Defects in carbohydrate metabolism are also associated with cancer. Reduced GAG and proteoglycan expression is associated with human lung carcinomas (Nackaerts, K. et al. (1997) *Int. J. Cancer* 74:335-345). The carbohydrate determinants sialyl-LewisA and sialyl-LewisX are frequently expressed on human cancer cells. These determinants, ligands for the cell adhesion molecule E-selectin, are involved in the adhesion of cancer cells to vascular endothelium and contribute to hematogenous metastasis of cancer (Kannagi, R. (1997) *Glycoconj. J.* 14:577-584). Alterations of the N-linked carbohydrate core structure of cell surface glycoproteins are linked to colon and pancreatic cancers (Schwarz, R.E. et al. (1996) *Cancer Lett.* 107:285-291). Reduced expression of the Sda blood group carbohydrate structure in cell surface glycolipids and glycoproteins is observed in gastrointestinal cancer (Dohi, T. et al. (1996) *Int. J. Cancer* 67:626-631).

Changes in glycosaminoglycan levels are associated with several autoimmune diseases. Both increases and decreases in various GAGs occur in patients with autoimmune thyroid disease and autoimmune diabetes mellitus. Antibodies to GAGs were found in patients with systemic lupus erythematosus and autoimmune thyroid disease (Hansen, C. et al. (1996) *Clin. Exp. Rheum.* 14:S59-S67). The glycosaminoglycan hyaluronan (HA) induces tumor cell adhesion and migration, and its small fragments are angiogenic. Serum HA is diagnostic of liver disease and various inflammatory conditions, such as rheumatoid arthritis. Interstitial edema caused by accumulation of HA may cause dysfunction in various organs (Laurent, T.C. and Fraser, J.R. (1992) *FASEB J.* 6:2397-2404). Hyaluronidase is an enzyme that degrades HA to oligosaccharides by catalyzing the random hydrolysis of 1,4-linkages between N-acetyl- β -D-glucosamine and D-glucuronate residues. Hyaluronidases may function in cell adhesion, infection, angiogenesis, and signal transduction. Hyaluronidases are associated with reproduction, cancer, and inflammation. Hyaluronidase activity is significantly elevated in prostate tumor tissue compared to that in both normal prostate and benign prostate hyperplasia (Lokeshwar, V.B. et al. (1996) *Cancer Res.* 56:651-657).

PH-20, a protein expressed in the mammalian testis and present on the plasma membrane of

WO 02/097060

PCT/US02/18354

mouse and human sperm, has hyaluronidase activity (Lin, Y. et al. (1994) *J. Cell Biol.* 125:1157-63). PH-20 enables sperm to penetrate the mammalian egg's outer layer, which consists of about 3,000 cumulus cells embedded in an extracellular matrix rich in HA. Penetration of this layer is an essential step in the fertilization process. PH-20 is also expressed in some tumor cells. Non-testicular mammalian hyaluronidases include the HYAL1 hyaluronidase, expressed in human serum, and lysosomal hyaluronidase HYAL2, expressed in many cells (Lepperdinger, G. et al. (1998) *J. Biol. Chem.* 273:22466-22470). HYAL2 may have a role in producing distinct HA fragments that can induce angiogenesis and the expression of enzymes involved in signal transduction pathways, such as nitric oxide synthase. A lysosomal-type hyaluronidase may degrade HA in lung fibroblasts in a cytokine-regulated process (Sampson, P.M. et al. (1992) *J. Clin. Invest.* 90:1492-1503). The venom of numerous animals including various snakes, bees, hornets, stone fish, platypus, scorpions, and lizards contain hyaluronidase. Venom hyaluronidase is thought to act as an aid in the diffusion of toxins.

A number of human diseases are linked to genetic or acquired deficiencies in protein glycosylation. Carbohydrate-deficient glycoprotein syndromes (CDGSs) include a host of alterations in glycosylation in a number of disorders and diseases. CDGSs are a group of hereditary multisystem disorders (Matthijs, G. et al. (1997) *Nat. Genet.* 16:88-92) causing severe psychomotor and mental retardation, as well as blood coagulation abnormalities seen in thrombosis, bleeding, or stroke-like episodes. The characteristic biochemical abnormality of CDGSs is the hypoglycosylation (N-linked) of glycoproteins (Freeze, H. and Aebi, M. (1999) *Biochim. Biophys. Acta* 1455:167-78). Depending on the type of CDGS, the carbohydrate side chains of glycoproteins are either truncated or completely missing from the protein core. Several different types of CDGS have been classified. The most common form, CDGS type 1A, is caused by phosphomannomutase (PMM1) deficiency (Matthijs, G. (1998) *Am. J. Hum. Genet.* 62:542-50). PMM1 functions upstream of MPG1 (see above) and catalyzes the conversion of D-mannose-6-phosphate to D-mannose-1-phosphate, which is required for the initial steps of protein glycosylation.

A second form of CDGS, designated CDGS type 1B, has also been described (Niehues, R. et al. (1998) *J. Clin. Invest.* 101:1414-1420). Psychomotor dysfunction and mental retardation are not present in this disease; instead, CDGS type 1B is a gastrointestinal disorder characterized by protein-losing enteropathy, severe hypoglycemia, vomiting, diarrhea, and congenital hepatic fibrosis. Nonetheless, some patients who are affected with CDGS type 1B suffer from thrombosis and life-threatening bleeding. A deficiency of phosphomannose isomerase (PMT) was identified as the most likely cause of this syndrome. Most symptoms can be controlled with dietary mannose supplements (Niehues et al., *supra*; Freeze and Aebi, *supra*). This form of CDGS is the first inherited disorder in human metabolism that shows a decrease in available mannose.

WO 02/097060

PCT/US02/18354

Defects in glucosyltransferase function also play an important role in some human diseases. Galactosyltransferases may be involved in autoimmune/inflammatory disorders as many humans with autoimmune thyroid disorders have high levels of circulating antibodies directed against the enzymatic product of α 1,3-galactosyltransferase (Etienne-Decert, J. et al. (1987) *Acta Endocrinol.* 115:67-74). An aberrantly-cleaved, soluble β 1,4-galactosyltransferase is secreted by a human ovarian cancer cell line (Uejima, T. et al. (1992) *Cancer Res.* 52:6158-6163). β 1,4-GalT-deficient transgenic mice exhibited growth retardation in one experiment (Asano, M. et al. (1997) *EMBO J.* 16:1850-1857), while targeted inactivation of the mouse α 1,4-GalT in another study was usually lethal (Furukawa, K. et al. (1999) *Biochim. Biophys. Acta.* 1473:54-66). In a third study, the constitutive overexpression of an α 1,3-galactosyltransferase in transgenic mice led to the increased secretion of proteins in the urine, low body weight, partial damage to hair growth, and early death (Ikematsu, S. et al. (1999) *Glycoconj. J.* 1999 16:73-76). Galactosyltransferases have also been implicated in the regulation of cellular growth, development, and differentiation and may play an important role in embryogenesis as well as tumor development. Secreted galactosyltransferases, derived in some cases from proteolytic cleavage of membrane-bound forms, may trigger cell surface receptors by binding their bound carbohydrates or may modify carbohydrates on cell surface molecules in a regulated fashion. Extracellular carbohydrate moieties are developmentally regulated and are likely involved in the regulation of cell migration (Shur, B. et al. (1984) *Mol. Cell. Biochem.* 61:143-158; Paulson, J. and Colley, K. (1989) *J. Biol. Chem.* 264:17615-17618). The expression of β 1,6-GlcNAc-bearing N-linked glycoproteins has been used as a marker of tumor progression in human breast and colon cancer, and astrocytes from human glioma specimens were found to contain increased levels of these type of glycoproteins compared to astrocytes from normal individuals (Yamamoto, H. et al. (2000) *Cancer Res.* 2000 60:134-142). These observations suggest that the dysfunction of another isoform of a glucosyltransferase, a β 1,6-GlcNAcT, may also play a role in tumor formation or invasivity.

Sialyltransferases have also been implicated in human disease. Elevated levels of 2,6-sialyltransferase (but not 2,3-sialyltransferase) are detected in human choriocarcinoma tissues, apparently the result of upregulation at the transcriptional level (Fukushima, K. (1998) *Cancer Res.* 58:4301-4306). Transient transfection of 2,6-sialyltransferase into human, tumorigenic, glioma cell line, reduces the invasivity of the cells (Yamamoto, H. (1997) *J. Neurochem.* 68:2566-2576). Chronic alcohol (ethanol) consumption causes a decrease in Gal- β 1,4-GlcNAc- α 2,6-sialyltransferase (α 2,6-ST) activity in the livers of rats who obtained at least one-third of their calories from alcohol for a period of one month or longer. Liver α 2,6-ST activity returned to normal after a week of abstinence from alcohol consumption. Based on the results of nuclear run-on assays and mRNA stability assays, the reduction in α 2,6-ST activity was the result of a 50% decrease in the half-life of

WO 02/097060

PCT/US02/18354

α 2,6-ST mRNA (Rao, M. (1999) *Metabolism* 48:797-803). A significant decrease in the plasma α 2,6-sialyltransferase activity was also observed in a group of individuals suffering from clinical depression. This particular form of depression was attributed to a change in glucocorticoid receptor (GR) functionality. These findings suggested that α 2,6-ST enzyme or activity levels may be a contributing factor in clinical depression or at least a useful biochemical marker of cortisol receptor dysfunction (Maguire, T. et al. (1997) *Biological Psychiatry* 41:1131-1136).

Additional human diseases that involve defects in glycosylation, and the enzyme deficiencies that cause them, include (i) aspartylglycosaminuria, an aspartylglycosaminidase deficiency that causes mental retardation, (ii) GM₁ and GM₂ gangliosidosis, β -galactosidase and β -N-acetylhexosaminidase deficiencies, respectively, that cause glycolipid storage diseases, (iii) α -mannosidosis and β -mannosidosis, caused by a deficiency of α -mannosidase or β -mannosidase, respectively, that cause neurological dysfunction, and (iv) sialidosis, caused by a neuraminidase deficiency, characterized by hepatosplenomegaly as well as impaired neural development.

Expression Profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder.

The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in tissues from subjects with colon cancer may be compared with the levels and sequences expressed in normal tissue.

Colon cancer is causally related to both genes and the environment. Several molecular pathways have been linked to the development of colon cancer, and the expression of key genes in any of these pathways may be lost by inherited or acquired mutation or by hypermethylation. There is a particular need to identify genes for which changes in expression may provide an early indicator of colon cancer or a predisposition for the development of colon cancer.

For example, it is well known that abnormal patterns of DNA methylation occur consistently in human tumors and include, simultaneously, widespread genomic hypomethylation and localized areas of increased methylation. In colon cancer in particular, it has been found that these changes occur early in tumor progression such as in premalignant polyps that precede colon cancer. Indeed, DNA methyltransferase, the enzyme that performs DNA methylation, is significantly increased in histologically normal mucosa from patients with colon cancer or the benign polyps that precede

WO 02/097060

PCT/US02/18354

cancer, and this increase continues during the progression of colonic neoplasms (Wafik, S. et al. (1991) Proc. Natl. Acad. Sci. USA 88:3470-3474). Increased DNA methylation occurs in G+C rich areas of genomic DNA termed "CpG islands" that are important for maintenance of an "open" transcriptional conformation around genes, and hypermethylation of these regions results in a "closed" conformation that silences gene transcription. It has been suggested that the silencing or downregulation of differentiation genes by such abnormal methylation of CpG islands may prevent differentiation in immortalized cells (Antequera, F. et al. (1990) Cell 62:503-514).

Familial Adenomatous Polyposis (FAP) is a rare autosomal dominant syndrome that precedes colon cancer and is caused by an inherited mutation in the adenomatous polyposis coli (APC) gene.

FAP is characterized by the early development of multiple colorectal adenomas that progress to cancer at a mean age of 44 years. The APC gene is a part of the APC- β -catenin-Tcf (T-cell factor) pathway. Impairment of this pathway results in the loss of orderly replication, adhesion, and migration of colonic epithelial cells that results in the growth of polyps. A series of other genetic changes follow activation of the APC- β -catenin-Tcf pathway and accompanies the transition from normal colonic mucosa to metastatic carcinoma. These changes include mutation of the K-ras proto-oncogene, changes in methylation patterns, and mutation or loss of the tumor suppressor genes p53 and Smad4/DPC4. While the inheritance of a mutated APC gene is a rare event, the loss or mutation of APC and the consequent effects on the APC- β -catenin-Tcf pathway is believed to be central to the majority of colon cancers in the general population.

Hereditary nonpolyposis colorectal cancer (HNPCC) is another inherited autosomal dominant syndrome with a less well defined phenotype than FAP. HNPCC, which accounts for about 2% of colorectal cancer cases, is distinguished by the tendency to early onset of cancer and the development of other cancers, particularly those involving the endometrium, urinary tract, stomach and biliary system. HNPCC results from the mutation of one or more genes in the DNA mismatch-repair (MMR) pathway. Mutations in two human MMR genes, MSH2 and MLH1, are found in a large majority of HNPCC families identified to date. The DNA MMR pathway identifies and repairs errors that result from the activity of DNA polymerase during replication. Furthermore, loss of MMR activity contributes to cancer progression through accumulation of other gene mutations and deletions, such as loss of the BAX gene which controls apoptosis, and the TGF- β receptor II gene which controls cell growth. Because of the potential for irreparable damage to DNA in an individual with a DNA MMR defect, progression to carcinoma is more rapid than usual.

Although ulcerative colitis is a minor contributor to colon cancer, affected individuals have about a 20-fold increase in risk for developing cancer. Progression is characterized by loss of the p53 gene which may occur early, appearing even in histologically normal tissue. The progression of the disease from ulcerative colitis to dysplasia/carcinoma without an intermediate polyp state suggests a

WO 02/097060

PCT/US02/18354

high degree of mutagenic activity resulting from the exposure of proliferating cells in the colonic mucosa to the colonic contents.

Almost all colon cancers arise from cells in which the estrogen receptor (ER) gene has been silenced. The silencing of ER gene transcription is age related and linked to hypermethylation of the ER gene (Issa, J.P. et al. (1994) Nat. Genet. 7:536-540). Introduction of an exogenous ER gene into cultured colon carcinoma cells results in marked growth suppression. The connection between loss of the ER protein in colonic epithelial cells and the consequent development of cancer has not been established.

Clearly there are a number of genetic alterations associated with colon cancer and with the development and progression of the disease, particularly the downregulation or deletion of genes, that potentially provide early indicators of cancer development, and which may also be used to monitor disease progression or provide possible therapeutic targets. The specific genes affected in a given case of colon cancer depend on the molecular progression of the disease. Identification of additional genes associated with colon cancer and the precancerous state would provide more reliable diagnostic patterns associated with the development and progression of the disease.

The discovery of new carbohydrate-associated proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of carbohydrate-associated proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, carbohydrate-associated proteins, referred to collectively as "CHOP" and individually as "CHOP-1," "CHOP-2," "CHOP-3," "CHOP-4," "CHOP-5," "CHOP-6," "CHOP-7," "CHOP-8," "CHOP-9," and "CHOP-10." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-10.

The invention further provides an isolated polynucleotide encoding a polypeptide selected

WO 02/097060

PCT/US02/18354

- from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-10. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:11-20.
- 10 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:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. 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.
- 20 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:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.
- 30 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:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an
- 35

WO 02/097060

PCT/US02/18354

immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

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:11-20, 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:11-20, 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.

10 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:11-20, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of
15 SEQ ID NO:11-20, 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
20 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:11-20, 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:11-20, 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
30 comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a
35 polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid

WO 02/097060

PCT/US02/18354

sequence selected from the group consisting of SEQ ID NO:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CHOP, 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:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. 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 CHOP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. 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 CHOP, comprising

WO 02/097060

PCT/US02/18354

administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. 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.

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:1-10, 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:1-10, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide

WO 02/097060

PCT/US02/18354

comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20, 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:11-20, 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:11-20, 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:11-20, 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, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are 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.

Table 4 lists the SEQ ID NO., Incyte ID, and length of the assembled polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

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

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

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

WO 02/097060

PCT/US02/18354

parameters.

DESCRIPTION OF THE INVENTION

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

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

Unless defined otherwise, all technical and scientific terms used herein have the same
 15 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
 20 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

"CHOP" refers to the amino acid sequences of substantially purified CHOP obtained from
 any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and
 25 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
 CHOP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other
 compound or composition which modulates the activity of CHOP either by directly interacting with
 CHOP or by acting on components of the biological pathway in which CHOP participates.

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

WO 02/097060

PCT/US02/18354

in a given sequence.

"Altered" nucleic acid sequences encoding CHOP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CHOP or a polypeptide with at least one functional characteristic of CHOP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CHOP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CHOP. 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 CHOP. 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 CHOP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

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

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

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

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

WO 02/097060

PCT/US02/18354

translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLEH). The coupled peptide is then used to immunize the animal.

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

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries.

15 Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules,
20 e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and Gold, L. (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in
25 the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

30 The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having
35 modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense

WO 02/097060

PCT/US02/18354

molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

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

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

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CHOP or fragments of CHOP 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.

35	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys

WO 02/097060

PCT/US02/18354

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

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

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

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

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

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

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

A "fragment" is a unique portion of CHOP or the polynucleotide encoding CHOP which is

WO 02/097060

PCT/US02/18354

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

A fragment of SEQ ID NO:11-20 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:11-20, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:11-20 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:11-20 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:11-20 and the region of SEQ ID NO:11-20 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:1-10 is encoded by a fragment of SEQ ID NO:11-20. A fragment of SEQ ID NO:1-10 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-10. For example, a fragment of SEQ ID NO:1-10 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-10. The precise length of a fragment of SEQ ID NO:1-10 and the region of SEQ ID NO:1-10 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.

WO 02/097060

PCT/US02/18354

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
 5 Higgins and Sharp, 1989 (CABIOS 5:151-153) and in Higgins 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.

10 Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence
 15 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
 20 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

25 *Penalty for mismatch: -2*

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

30 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous
 35 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length

WO 02/097060

PCT/US02/18354

supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

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

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

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: 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

WO 02/097060

PCT/US02/18354

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.

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

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

20 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 et al., 1989 (Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, 30 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 35 circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions

WO 02/097060

PCT/US02/18354

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.

5 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., C₃t or R₃t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

10 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.

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

20 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.

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

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.

30 "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

WO 02/097060

PCT/US02/18354

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.

5 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 CHOP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will
10 vary by cell type depending on the enzymatic milieu of CHOP.

"Probe" refers to nucleic acid sequences encoding CHOP, 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.

15 "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

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

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

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to
35 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer

WO 02/097060

PCT/US02/18354

selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection 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,

WO 02/097060

PCT/US02/18354

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.

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

10 The term "sample" is used in its broadest sense. A sample suspected of containing CHOP, nucleic acids encoding CHOP, 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.

15 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.

20 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.

25 "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.

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

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or

WO 02/097060

PCT/US02/18354

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.

5 A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with
10 a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). 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
15 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
20 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
25 sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the
30 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
35 presence of SNPs may be indicative of, for example, a certain population, a disease state, or a

WO 02/097060

PCT/US02/18354

propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

10 THE INVENTION

The invention is based on the discovery of new human carbohydrate-associated proteins (CHOP), the polynucleotides encoding CHOP, and the use of these compositions for the diagnosis, treatment, or prevention of carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer.

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

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

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the

WO 02/097060

PCT/US02/18354

MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

5 Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are carbohydrate-associated proteins. For example, SEQ ID NO:1 is 54% identical, from residue Q170 to residue W909, to a murine α -glucosidase II, alpha subunit (GenBank ID g2104689) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.4\text{e-}258$, which indicates the probability
10 of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a glycosyl hydrolase family 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 MOTIFS analyses provide further corroborative evidence that SEQ ID NO:1 is a glycosyl hydrolase. In an alternative example, SEQ ID NO:4 is 55% identical,
15 from residue E20 to residue T276, to bull α 1,3-galactosyltransferase (GenBank ID g163124) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $5.8\text{e-}77$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from further BLAST analyses provide corroborative evidence that SEQ ID NO:4 is an α 1,3-galactosyltransferase. In another example, SEQ ID NO:7 is 77%
20 identical, from residue M2 to residue F402, to mouse β 1,6-N-acetylglucosaminyltransferase B (GenBank ID g9650954) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $2.1\text{e-}177$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a core-2/I-branching enzyme domain as determined by searching for statistically significant matches in the
25 hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from further BLAST analyses provide corroborative evidence that SEQ ID NO:7 is a β 1,6-N-acetylglucosaminyltransferase B. In an alternative example, SEQ ID NO:9 is 81% identical, from residue M8 to residue Q200, to human intelectin (GenBank ID g8096221) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.1e-
30 83 , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 is an intelectin, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:9 also contains a fibrinogen domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analysis of the PRODOM
35 database provide further corroborative evidence that SEQ ID NO:9 is a lectin. In another example,

WO 02/097060

PCT/US02/18354

SEQ ID NO:10 is 36% identical, from residue T35 to residue E387, to human cargo selection protein (mannose-6-phosphate receptor binding protein) (GenBank ID g14043157) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.1e-51, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also has homology to proteins that are localized to the endosomal vesicles, which mediate the transport of mannose-6-phosphate receptors from endosomes to the Golgi apparatus, and are endosome/endosomal vesicle cytoplasmic tail-interacting protein of 47 kDa and perilipin proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:10 also contains a perilipin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analysis against the BLAST_PRODOM database provides further corroborative evidence that SEQ ID NO:10 is a homolog of perilipin. SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:6, and SEQ ID NO:8 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-10 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:11-20 or that distinguish between SEQ ID NO:11-20 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (i.e., those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (i.e., those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (i.e., those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in

WO 02/097060

PCT/US02/18354

column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N_{1,2,3,...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAA_gBBBBB_1_N is 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
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).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which

WO 02/097060

PCT/US02/18354

were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses CHOP variants. A preferred CHOP 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 CHOP amino acid sequence, and which contains at least one functional or structural characteristic of CHOP.

The invention also encompasses polynucleotides which encode CHOP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:11-20, which encodes CHOP. The polynucleotide sequences of SEQ ID NO:11-20, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CHOP. 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 CHOP. 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:11-20 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:11-20. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CHOP.

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

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CHOP, some bearing minimal

WO 02/097060

PCT/US02/18354

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 CHOP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CHOP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CHOP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CHOP 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 CHOP 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 CHOP and CHOP 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 CHOP 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:11-20 and fragments thereof under various conditions of stringency (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 Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied

WO 02/097060

PCT/US02/18354

Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (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 CHOP 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 (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 (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 (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 (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 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

WO 02/097060

PCT/US02/18354

software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

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

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

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat.
20 Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CHOP, 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
25 properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of
30 homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CHOP may be synthesized, in whole or in part, using chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, CHOP

WO 02/097060

PCT/US02/18354

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 (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) *Science* 269:202-204). Automated synthesis may be achieved using the
 5 ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CHOP, 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
 10 chromatography (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 (Creighton, *supra*, pp. 28-53).

In order to express a biologically active CHOP, the nucleotide sequences encoding CHOP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains
 15 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 CHOP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CHOP. Such signals
 20 include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CHOP 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
 25 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 (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
 30 vectors containing sequences encoding CHOP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook et al., *supra*, 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).

35 A variety of expression vector/host systems may be utilized to contain and express sequences

WO 02/097060

PCT/US02/18354

- encoding CHOP. 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 (Sambrook et al., *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 (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; 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 CHOP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CHOP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of sequences encoding CHOP 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 (Van Heeke and Schuster, *supra*). When large quantities of CHOP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CHOP 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 CHOP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation (Ausubel et al., 1995, *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994)

WO 02/097060

PCT/US02/18354

Bio/Technology 12:181-184).

Plant systems may also be used for expression of CHOP. Transcription of sequences encoding CHOP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; 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 (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

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

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

For long term production of recombinant proteins in mammalian systems, stable expression of CHOP in cell lines is preferred. For example, sequences encoding CHOP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *aprt*⁻ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or

WO 02/097060

PCT/US02/18354

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 (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 (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β -glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (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 CHOP is inserted within a marker gene sequence, transformed cells containing sequences encoding CHOP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CHOP 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 CHOP and that express CHOP 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 CHOP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CHOP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (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

WO 02/097060

PCT/US02/18354

hybridization or PCR probes for detecting sequences related to polynucleotides encoding CHOP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CHOP, 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 Biosciences, 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 CHOP 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 CHOP may be designed to contain signal sequences which direct secretion of CHOP through a prokaryotic or eukaryotic cell membrane.

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

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CHOP 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 CHOP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CHOP 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

WO 02/097060

PCT/US02/18354

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 CHOP encoding sequence and the heterologous protein sequence, so that CHOP may be cleaved away from the heterologous moiety following purification.

5 Methods for fusion protein expression and purification are discussed in Ausubel et al., 1995 (*supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

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

CHOP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CHOP. At least one and up to a plurality of test compounds may be screened for specific binding to CHOP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules. In one embodiment, the compound thus identified is closely related to the natural ligand of CHOP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified is a natural ligand of a receptor CHOP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci. 22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, the compound can be closely related to the natural receptor to which CHOP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for CHOP which is capable of propagating a signal, or a decoy receptor for CHOP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Immunex Corp., Seattle WA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG₁ (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

35 In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit CHOP involves producing appropriate cells which express CHOP, either as a secreted protein or on

WO 02/097060

PCT/US02/18354

the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CHOP or cell membrane fractions which contain CHOP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CHOP or the compound is analyzed.

5 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 CHOP, either in solution or affixed to a solid support, and detecting the binding of CHOP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a
10 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.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include
15 radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to
20 improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

CHOP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CHOP. Such compounds may include agonists, antagonists, or partial or
25 inverse agonists. In one embodiment, an assay is performed under conditions permissive for CHOP activity, wherein CHOP is combined with at least one test compound, and the activity of CHOP in the presence of a test compound is compared with the activity of CHOP in the absence of the test compound. A change in the activity of CHOP in the presence of the test compound is indicative of a compound that modulates the activity of CHOP. Alternatively, a test compound is combined with an
30 *in vitro* or cell-free system comprising CHOP under conditions suitable for CHOP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CHOP 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 CHOP or their mammalian homologs may
35 be "knocked out" in an animal model system using homologous recombination in embryonic stem

WO 02/097060

PCT/US02/18354

(ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 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 therapeutic or toxic agents.

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

Polynucleotides encoding CHOP 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 CHOP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease.

Alternatively, a mammal inbred to overexpress CHOP, e.g., by secreting CHOP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CHOP and carbohydrate-associated proteins. In addition, examples of tissues expressing CHOP can be found in Table 6 and can also be found in Example XI. Therefore, CHOP appears to play a role in carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer. In the treatment of disorders associated with increased CHOP expression or activity, it is desirable to decrease the expression or activity of CHOP. In the treatment of disorders associated with decreased CHOP expression or activity, it is desirable to increase the expression or activity of CHOP.

WO 02/097060

PCT/US02/18354

Therefore, in one embodiment, CHOP 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 CHOP. Examples of such disorders include, but are not limited to, a carbohydrate metabolic disorder, such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, fructose-1,6-diphosphatase deficiency, galactosemia, glucagonoma, hereditary fructose intolerance, hypoglycemia, mannosidosis, neuraminidase deficiency, obesity, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, carbohydrate-deficient glycoprotein syndromes (CDGS types 1A and 1B), autoimmune thyroid disorders, aspartylglycosaminuria, GM₁ gangliosidosis, GM₂ gangliosidosis, β -galactosidase β -N-acetylhexosaminidase deficiency, glycolipid storage diseases, neurological dysfunction, sialidosis, hepatosplenomegaly, and inherited abnormalities of pyruvate metabolism; 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; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, 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; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency,

WO 02/097060

PCT/US02/18354

delayed puberty, retrograde ejaculation and anejaculation, haemangioblastomas,
 cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac
 tumours; a genetic disorder, such as adrenoleukodystrophy, Alport's syndrome, choroideremia,
 Duchenne and Becker muscular dystrophy, Down syndrome, cystic fibrosis, chronic granulomatous
 5 disease, Gaucher's disease, Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic
 dystrophy, pycnodysostosis, Refsum's syndrome, retinoblastoma, sickle cell anemia, thalassemia,
 Werner syndrome, von Willebrand's disease, Wilms' tumor, Zellweger syndrome, peroxisomal acyl-
 CoA oxidase deficiency, peroxisomal thiolase deficiency, peroxisomal bifunctional protein
 deficiency, mitochondrial carnitine palmitoyl transferase and carnitine deficiency, mitochondrial
 10 very-long-chain acyl-CoA dehydrogenase deficiency, mitochondrial medium-chain acyl-CoA
 dehydrogenase deficiency, mitochondrial short-chain acyl-CoA dehydrogenase deficiency,
 mitochondrial electron transport flavoprotein and electron transport flavoprotein:ubiquinone
 oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-
 chain 3-hydroxyacyl-CoA dehydrogenase deficiency; a transport disorder, such as akinesia,
 15 amyotrophic lateral sclerosis, ataxia telangiectasia, Bell's palsy, Charcot-Marie Tooth disease,
 diabetes insipidus, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's
 disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, dystonias, peripheral
 neuropathy; cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia,
 hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear
 20 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,
 25 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, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria,
 iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as ischemic
 30 cerebrovascular disease, stroke, cerebral neoplasms, Pick's disease, Huntington'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
 35 system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-

WO 02/097060

PCT/US02/18354

Strassler-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 seasonal affective disorder (SAD), akathisia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, postherpetic neuralgia, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; 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 CHOP 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 CHOP including, but not limited to, those described above.

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

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

In a further embodiment, an antagonist of CHOP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CHOP. Examples of such disorders include, but are not limited to, those carbohydrate metabolism, cell proliferative, autoimmune/inflammatory, reproductive, genetic, transport, and neurological disorders and cancer, described above. In one aspect, an antibody which specifically binds CHOP 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 CHOP.

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

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

WO 02/097060

PCT/US02/18354

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 CHOP may be produced using methods which are generally known in the art. In particular, purified CHOP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CHOP. Antibodies to CHOP 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. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with CHOP 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 CHOP 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 CHOP 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 CHOP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (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.

WO 02/097060

PCT/US02/18354

et al. (1984) Mol. Cell. Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad.

- 5 Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CHOP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. 10 (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

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

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

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

- Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CHOP. Affinity is expressed as an 30 association constant, K_a, which is defined as the molar concentration of CHOP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CHOP epitopes, represents the average affinity, or avidity, of the antibodies for CHOP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a 35 particular CHOP epitope, represents a true measure of affinity. High-affinity antibody preparations

WO 02/097060

PCT/US02/18354

with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CHOP-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 CHOP, 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 CHOP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*, and Coligan et al., *supra*).

In another embodiment of the invention, the polynucleotides encoding CHOP, 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 CHOP. 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 CHOP (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 (Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; 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 (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 (Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736).

In another embodiment of the invention, polynucleotides encoding CHOP 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-

WO 02/097060

PCT/US02/18354

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 CHOP expression or regulation causes disease, the expression of CHOP 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 CHOP are treated by constructing mammalian expression vectors encoding CHOP and introducing these vectors by mechanical means into CHOP-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. R  c  pon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of CHOP include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CHOP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter

WO 02/097060

PCT/US02/18354

(Rossi and Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CHOP from a normal individual.

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

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CHOP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CHOP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

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

WO 02/097060

PCT/US02/18354

(Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinuzzi et al., 1999 (Annu. Rev. Nutr. 19:511-544) and Verma and Somia, 1997 (Nature 389:239-242), both incorporated by
 5 reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CHOP to target cells which have one or more genetic abnormalities with respect to the expression of CHOP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CHOP to cells of the central nervous system, for which HSV has a
 10 tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby
 15 incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins et al., 1999 (J. Virol. 73:519-532) and Xu et al., 1994 (Dev. Biol.
 20 163:152-161), 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
 25 deliver polynucleotides encoding CHOP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA,
 30 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 CHOP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CHOP-coding RNAs and the synthesis of high levels of CHOP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a
 35 persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN)

WO 02/097060

PCT/US02/18354

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 CHOP 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
 5 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,
 10 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 (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
 15 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,
 20 engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CHOP.

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,
 25 corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared
 30 by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding CHOP. 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
 35 that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines,

WO 02/097060

PCT/US02/18354

cells, or tissues.

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

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

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CHOP 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 CHOP 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 CHOP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without

WO 02/097060

PCT/US02/18354

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 (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 CHOP, antibodies to CHOP, and mimetics, agonists, antagonists, or inhibitors of CHOP.

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

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the

WO 02/097060

PCT/US02/18354

lung have enabled the practical delivery of drugs such as insulin to blood circulation (Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

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

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

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

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

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and

WO 02/097060

PCT/US02/18354

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

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

DIAGNOSTICS

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

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

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

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CHOP or closely related molecules may be used to identify nucleic acid sequences which encode CHOP. 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

WO 02/097060

PCT/US02/18354

conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CHOP, 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 CHOP 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:11-20 or from genomic sequences including promoters, enhancers, and introns of the CHOP gene.

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

Polynucleotide sequences encoding CHOP may be used for the diagnosis of disorders associated with expression of CHOP. Examples of such disorders include, but are not limited to, a carbohydrate metabolic disorder, such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, fructose-1,6-diphosphatase deficiency, galactosemia, glucagonoma, hereditary fructose intolerance, hypoglycemia, mannosidosis, neuraminidase deficiency, obesity, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, carbohydrate-deficient glycoprotein syndromes (CDGS types 1A and 1B), autoimmune thyroid disorders, aspartylglycosaminuria, GM1, gangliosidosis, GM2, gangliosidosis, β -galactosidase β -N-acetylhexosaminidase deficiency, glycolipid storage diseases, neurological dysfunction, sialidosis, hepatosplenomegaly, and inherited abnormalities of pyruvate metabolism; 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; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, 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,

WO 02/097060

PCT/US02/18354

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

5 extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of

10 the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and anejaculation, haemangioblastomas,

15 cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumours; a genetic disorder, such as adrenoleukodystrophy, Alport's syndrome, choroideremia, Duchenne and Becker muscular dystrophy, Down syndrome, cystic fibrosis, chronic granulomatous disease, Gaucher's disease, Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic dystrophy, pycnodysostosis, Refsum's syndrome, retinoblastoma, sickle cell anemia, thalassemia,

20 Werner syndrome, von Willebrand's disease, Wilms' tumor, Zellweger syndrome, peroxisomal acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, peroxisomal bifunctional protein deficiency, mitochondrial carnitine palmitoyl transferase and carnitine deficiency, mitochondrial very-long-chain acyl-CoA dehydrogenase deficiency, mitochondrial medium-chain acyl-CoA dehydrogenase deficiency, mitochondrial short-chain acyl-CoA dehydrogenase deficiency,

25 mitochondrial electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; a transport disorder, such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, Bell's palsy, Charcot-Marie Tooth disease, diabetes insipidus, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's

30 disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, dystonias, peripheral neuropathy; 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

35 associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression,

WO 02/097060

PCT/US02/18354

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,

5 Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as ischemic cerebrovascular disease, stroke, cerebral neoplasms, Pick's disease, Huntington's disease, and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive

10 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

15 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,

20 inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and seasonal affective disorder (SAD), akathisia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, postherpetic neuralgia, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and cancers including

adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

25 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 CHOP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-

30 like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CHOP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CHOP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CHOP may be labeled by standard methods and added to a fluid or tissue sample

35 from a patient under conditions suitable for the formation of hybridization complexes. After a

WO 02/097060

PCT/US02/18354

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 CHOP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CHOP, 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 CHOP, 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 CHOP 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 CHOP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CHOP, 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 CHOP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

WO 02/097060

PCT/US02/18354

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 CHOP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

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

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

WO 02/097060

PCT/US02/18354

presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

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

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

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of 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

WO 02/097060

PCT/US02/18354

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 (Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

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

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical

WO 02/097060

PCT/US02/18354

density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

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

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the 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

WO 02/097060

PCT/US02/18354

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 (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 CHOP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington et al., *supra*; 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) (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 (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 CHOP 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,

WO 02/097060

PCT/US02/18354

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 (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, CHOP, 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 CHOP 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 (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CHOP, or fragments thereof, and washed. Bound CHOP is then detected by methods well known in the art. Purified CHOP 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 CHOP specifically compete with a test compound for binding CHOP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CHOP.

In additional embodiments, the nucleotide sequences which encode CHOP 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. 60/293,768, U.S. Ser. No. 60/309,548, U.S. Ser. No. 60/314,400, U.S. Ser. No. 60/343,706, and U.S. Ser. No. 60/337,999, are expressly incorporated by reference herein.

WO 02/097060

PCT/US02/18354

EXAMPLES

I. Construction of cDNA Libraries

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

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

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPRT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel, *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 Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Invitrogen.

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,

WO 02/097060

PCT/US02/18354

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).

10 III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences 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 (Amersham Biosciences); 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, *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; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002)

WO 02/097060

PCT/US02/18354

Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

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

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:11-20. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative carbohydrate-associated proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpr1 and gbhtg).

WO 02/097060

PCT/US02/18354

Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (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

5 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 carbohydrate-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for carbohydrate-associated proteins. Potential carbohydrate-associated proteins were also identified by homology to Incyte cDNA sequences that

10 had been annotated as carbohydrate-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbprl 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

15 sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or

20 unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example

25 III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on

30 more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear

35 along their parent sequences to generate the longest possible sequence, as well as sequence variants.

WO 02/097060

PCT/US02/18354

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 gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

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

20 VI. Chromosomal Mapping of CHOP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:11-20 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:11-20 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

30 Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation

WO 02/097060

PCT/US02/18354

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.

5 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 (Sambrook et al., *supra*, ch. 7; Ausubel et al., 1995, *supra*, ch. 4 and 16).

10 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:

15

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

The product score takes into account both the degree of similarity between two sequences and the
 20 length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by
 25 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
 30 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 CHOP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA
 35 sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is

WO 02/097060

PCT/US02/18354

classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding CHOP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of CHOP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,

WO 02/097060

PCT/US02/18354

Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). 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 Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) 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 Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

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

IX. Identification of Single Nucleotide Polymorphisms in CHOP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:11-20 using the LIFSEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice

WO 02/097060

PCT/US02/18354

variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:11-20 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] adenosine triphosphate (Amersham Biosciences), 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 Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

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

WO 02/097060

PCT/US02/18354

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the
 5 aforementioned technologies should be uniform and solid with a non-porous surface (Schna, *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to
 10 those of ordinary skill in the art and may contain any appropriate number of elements (Schna, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be
 15 selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser
 20 desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

25 Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (2Imet), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse
 30 transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified
 35 using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.

WO 02/097060

PCT/US02/18354

(CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

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

10 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

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,

15 and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic

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

25 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered

30 with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X

35 SSC), and dried.

WO 02/097060

PCT/US02/18354

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). Array elements that exhibited at least about a two-fold change in expression, a signal-to-background

WO 02/097060

PCT/US02/18354

ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

Expression

For example, SEQ ID NO:19 showed differential expression in colon tissue from patients with colon cancer compared to matched microscopically normal tissue from the same donors as determined by microarray analysis. The expression of SEQ ID NO:19 was decreased at least two-fold in cancerous colon tissue. Therefore, SEQ ID NO:19 is useful in disease staging and diagnostic assays for cell proliferative disorders, including colon cancer.

XII. Complementary Polynucleotides

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

XIII. Expression of CHOP

Expression and purification of CHOP is achieved using bacterial or virus-based expression systems. For expression of CHOP 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 CHOP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CHOP 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 CHOP 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 (Engelhard et al., *supra*; Sandig et al., *supra*).

WO 02/097060

PCT/US02/18354

In most expression systems, CHOP 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 Biosciences). Following purification, the GST moiety can be proteolytically cleaved from CHOP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al., 1995 (*supra*, ch. 10 and 16). Purified CHOP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

CHOP function is assessed by expressing the sequences encoding CHOP 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 plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, 1994 (*Flow Cytometry*, Oxford, New York NY).

The influence of CHOP on gene expression can be assessed using highly purified populations

WO 02/097060

PCT/US02/18354

of cells transfected with sequences encoding CHOP 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 CHOP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of CHOP Specific Antibodies

CHOP substantially purified using polyacrylamide gel electrophoresis (PAGE; Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the CHOP 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 (Ausubel et al., 1995, *supra*, ch. 11).

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

XVI. Purification of Naturally Occurring CHOP Using Specific Antibodies

Naturally occurring or recombinant CHOP is substantially purified by immunoaffinity chromatography using antibodies specific for CHOP. An immunoaffinity column is constructed by covalently coupling anti-CHOP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CHOP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CHOP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CHOP 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 CHOP is collected.

XVII. Identification of Molecules Which Interact with CHOP

WO 02/097060

PCT/US02/18354

CHOP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (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 CHOP, washed, and any wells with labeled CHOP complex are assayed. Data obtained using different concentrations of CHOP are used to calculate values for the number, affinity, and association of CHOP with the candidate molecules.

Alternatively, molecules interacting with CHOP are analyzed using the yeast two-hybrid system as described in Fields and Song, 1989 (*Nature* 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

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

XVIII. Demonstration of CHOP Activity

Galactosyltransferase Activity Assay #1

β 1,3-galactosyltransferase and β 1,4-galactosyltransferase activity of CHOP is determined by measuring the transfer of galactose from UDP-galactose to a GlcNAc-terminated oligosaccharide chain in a radioactive assay (Kolbinger, F. et al. (1998) *J. Biol. Chem.* 273:58-65; Hennet, T. et al. (1998) *J. Biol. Chem.* 273:58-65). An aliquot of CHOP is incubated with 14 μl of assay stock solution (180 mM sodium cacodylate, pH 6.5, 1 mg/ml bovine serum albumin, 0.26 mM UDP-galactose, 2 μl of UDP-[^3H]galactose, 1 μl of MnCl_2 (500 mM), and 2.5 μl of $\text{GlcNAc}\beta\text{O}-(\text{CH}_2)_6\text{-CO}_2\text{Me}$ (37 mg/ml in dimethyl sulfoxide)) for approximately 1 hr at 37°C. The reaction is quenched by the addition of 1 ml of water and loaded on a C18 Sep-Pak cartridge (Waters). The column is washed twice with 5 ml of water to remove unreacted UDP-[^3H]galactose. The reaction product, [^3H]galactosylated $\text{GlcNAc}\beta\text{O}-(\text{CH}_2)_6\text{-CO}_2\text{Me}$, remains bound to the column during the water washes and is eluted with 5 ml of methanol. The level of radioactivity in the eluted material is measured by liquid scintillation counting and is proportional to CHOP galactosyltransferase activity in the sample.

Galactosyltransferase Activity Assay #2

In the alternative, β 1,4-galactosyltransferase activity of CHOP is determined by quantitating the transfer of [^{14}C]galactose from UDP-[^{14}C]Gal to ovalbumin. The approximately 50 μl reaction contains 50 mM HEPES (pH 7.35), 10 mM MnCl_2 , 1.5 mg of ovalbumin, 50 mM NaCl, 5 μl UDP-[^{14}C]galactose (25 nCi), and 5 μl of CHOP. The assay is incubated at 60 °C for 30 minutes and terminated by the addition of ice-cold 2.5% phosphotungstic acid (w/v) in 1 M HCl. Unincorporated UDP-[^{14}C]Gal is separated by filtration through Whatman GF/C glass fiber filters. The filters are washed once with 2.5% phosphotungstic acid (w/v) and then rinsed with ice-cold ethanol. The filters

WO 02/097060

PCT/US02/18354

are dried, and radioactivity is determined using a scintillation counter. The amount of radioactivity is proportional to the activity of CHOP (Verdon, B. and Berger, E. (1983) in Galactosyltransferase. Methods of Enzymatic Analysis, Bergmeyer, H. et al. (eds.), Vol. III, 3rd Ed., Verlag Chemie, Weinheim, Deerfield Beach, Basel, pp. 374-381; Tilo, S. et al. (1996) *J. Biol. Chem.* 271:3398-3405).

5 Sialyltransferase Activity Assay #1

In the alternative, sialyltransferase activity of CHOP is assayed as follows. The sialyltransferase acceptors used in the assay are derived from aminophenylglycosides reacted with 6(5-fluorescein-carboxamido)-hexanoic acid succinidyl ester (FCHASE). Briefly, 10 mg *p*-aminophenylglycoside is dissolved in 0.5 ml of 0.2 M triethylamine acetate buffer, pH 8.2.

- 10 FCHASE was dissolved in 0.5 ml of methanol and added to the aminophenylglycoside solution. The mixture is stirred in the dark for about 3 h at room temperature, lyophilized, resuspended in approximately 200 μ l of 50% acetonitrile, and spotted on a silica thin-layer chromatography (TLC) plate which is developed with an ethyl acetate/methanol/water/acetic acid solvent system. Following air drying in a fume hood, the yellow product is scraped, eluted with distilled water, concentrated, 15 desalted, and bound to a Sep-Pak C18 reverse phase cartridge. After washing the cartridge with several volumes of water, the product is eluted in 50% acetonitrile and quantitated by well-known spectrophotometric methods. Following preparation of the sialyltransferase acceptors, enzyme reactions are performed at 37 °C in 20 μ l volumes in a buffer consisting of 50 mM MES (pH 6.0), 10 mM MnCl₂, with 0.2 or 1.0 mM labeled acceptor, 0.2 mM CMP-Neu5Ac donor, and various amounts 20 of CHOP. The reaction is terminated by diluting the reaction with 10 mM NaOH prior to analysis by capillary electrophoresis. Capillary electrophoresis (CE) is performed using an Argon-ion laser-induced fluorescence detector (excitation = 488 nm; emission = 520 nm). The product peak, consisting of FCHASE-2,3-sialyl-N-acetylactosamine, is identified and quantitated and is proportional to the sialyltransferase activity in the sample (Wakarchuk, W. et al. (1996) *J. Biol. Chem.* 271:19166-19173; Gilbert, M. (1996) *J. Biol. Chem.* 271:28271-28276).

25 Sialyltransferase Activity Assay #2

Alternatively, sialyltransferase activity of CHOP is assayed as follows. Sialyltransferase assays are performed in a reaction mixture containing 10 mM MgCl₂, 0.3% Triton CF-54, 100 mM sodium cacodylate buffer (pH 6.0), 0.66 mM unlabeled CMP-Neu5Ac donor, 4,400 dpm/ μ l 30 CMP-[¹⁴C]Neu5Ac (tracer), the CHOP solution, and substrates in total volume of 20 to 50 μ l. The reaction mixture is incubated at 37 °C for 2 h then terminated by addition of 500 μ l of water. The products are isolated by C18 Sep-Pak cartridge and analyzed by thin layer chromatography (TLC) using the solvent system ethanol/pyridine/n-butanol/acetate/water (100:10:10:3:30) (Okajima, T. et al. (1999) *J. Biol. Chem.* 274:11479-11486). The radiolabeled products are visualized by standard 35 autoradiography techniques familiar to persons skilled in the art and sialyltransferase activity of

WO 02/097060

PCT/US02/18354

CHOP is proportional to the signal on the autoradiogram.

Sialyltransferase Activity Assay #3

In the alternative, sialyltransferase activity of CHOP is assayed as follows. Human embryonic kidney cells (293) are stably transfected with a plasmid encoding the CHOP. The cells are grown to confluence in 225 cm² tissue culture flasks and harvested by scraping cells into phosphate-buffered saline (PBS). Cells are pelleted and resuspended in approximately 1 ml of 1% Triton X-100, 50 mM NaCl, 5 mM MnCl₂, and 25 mM MES (pH 6.0). The cell pellet is solubilized by repeated pipetting and vortexing. This crude homogenate is cleared by centrifugation at 1000 x g for 10 min and used directly as the enzyme source. The assay mixture consists of 50 μM CMP-Neu5Ac with 250,000 cpm of CMP-[¹⁴C]Neu5Ac added as a tracer, 0.1% Triton CF-54, 20 mM cacodylate (pH 6.0) and 10 μl of CHOP-containing extract in a 30-μl reaction volume. Glycoprotein and glycolipid products are separated from CMP-Neu5Ac by gel filtration and the amount of label in the eluted fractions are quantitated to determine the relative amount of CHOP activity in the sample (Sjoberg, E. et al. (1996) J. Biol. Chem. 271:7450-7459).

O-glucosyltransferase Transferase Activity Assay

O-glucosyltransferase activity of CHOP is assayed as follows. CHOP preparations are diluted with cold desalt buffer (20 mM Tris pH 8.0, 20% glycerol, 0.02% NaN₃) immediately prior to use. Peptide substrates (Kreppel, L. et al. (1999) J. Biol. Chem. 274:32015-32022) are used as acceptors at a concentration of approximately 3 mM. The peptides in the reaction are separated from the reactants using a SP-Sephadex column. The modified and unmodified peptides are loaded onto a Sep-pak C18 cartridge. Unmodified peptides are eluted with 50 mM formic acid, 10 ml of 50 mM formic acid containing 0.5 M NaCl, and 10 ml of distilled H₂O. Modified peptides are eluted from the cartridge directly into scintillation vials using methanol. Enzyme activity is expressed in terms of micromoles of GlcNAc transferred per minute, which is proportional to the level of CHOP activity in the sample (Kreppel et al., *supra*).

Mannosidase Activity Assay

Mannosidase activity in CHOP is measured by its ability to release mannose from Man₆(GlcNAc)₂ oligosaccharide (Schweden, J. et al. (1986) Eur. J. Biochem. 157:563-570). CHOP, in 200 mM phosphate buffer, pH 6.5 and 1% Triton X-100, is mixed with [¹⁴C](Man₆)(GlcNAc)₂ (2-3 x 10³ cpm) in a final volume of 30 μl at 37°C for 60 minutes. The reaction is terminated by the addition of 30 μl glacial acetic acid. The amount of liberated [¹⁴C]mannose, analyzed by paper chromatography in 2-propanol/acetic acid/water (29/4/9, by volume), is proportional to the activity of CHOP in the starting sample.

Carbohydrate Binding Assay

CHOP activity is also demonstrated by the ability of CHOP to bind to β-galactoside sugars.

WO 02/097060

PCT/US02/18354

CHOP is applied to a lactosyl-Sepharose column, and the column is eluted with 0.1 M lactose. The presence of CHOP in the eluate is detected by sodium dodecyl sulfate polyacrylamide gel electrophoresis and indicates the ability of CHOP to bind β -galactoside sugars.

Hyaluronan Hydrolysis Assay

- 5 CHOP activity is also measurable by its ability to hydrolyze hyaluronan (HA) (Leppendinger, *supra*). Radioactively labeled HA is immobilized on microtiter plates with the aid of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and N-hydroxy-sulfosuccinimide. The radioactivity solubilized after incubation with CHOP is measured using a liquid scintillation counter and is proportional to the activity of CHOP in the starting sample.

10 Cellular Transformation Assay

- Alternatively, CHOP activity is measured by its ability to regulate transformation of NIH3T3 mouse fibroblast cells. A cDNA encoding CHOP is subcloned into an appropriate eukaryotic expression vector. This construct is transfected into NIH3T3 cells using methods known in the art. Transfected cells are compared with non-transfected cells for the following quantifiable properties characteristic of oncogenically-transformed cells: high-density growth in culture associated with loss of contact inhibition; growth in suspension or in soft agar; reduced serum requirements; and ability to induce tumors when injected into immunodeficient mice. The activity of CHOP is proportional to the extent of transformation of NIH3T3 cells transfected with CHOP, compared to non-transfected cells.

XIX. Expression and purification of CHOP-associated α 1,3-Galactosyltransferase activity

- 20 Soluble, catalytically-active CHOP with α 1,3-galactosyltransferase (α 1,3-GalT) activity is purified from culture media of nearly confluent human embryonic kidney (293) cells transfected with a plasmid expressing CHOP. The culture media is dialyzed overnight at 4 °C against 20 mM HEPES (pH 7.0) and 5 mM $MnCl_2$, with several changes of buffer and then applied to an affinity column of UDP-hexanolamine-Sepharose (4 μ mol/ml resin). The column is washed with dialysis buffer and dialysis buffer containing 0.75 M NaCl until the A_{280} of the eluate reaches background levels. The bound enzyme is eluted with 5 mM UDP in dialysis buffer and concentrated using a Centricon-10 (Millipore). The homogeneity of the purified enzyme is determined by separation on a 10% SDS-PAGE gel (Kim, S. et al. (1997) J. Biol. Chem. 272:13622-13628).

- Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

WO 02/097060

PCT/US02/18354

Table 1

Inocyte Project ID	Polypeptide SEQ ID NO:	Inocyte Polypeptide ID	Polynucleotide SEQ ID NO:	Inocyte Polynucleotide ID
7247177	1	7247177CD1	11	7247177CB1
5596327	2	5596327CD1	12	5596327CB1
4179651	3	4179651CD1	13	4179651CB1
7482109	4	7482109CD1	14	7482109CB1
2867618	5	2867618CD1	15	2867618CB1
7488348	6	7488348CD1	16	7488348CB1
5539166	7	5539166CD1	17	5539166CB1
7500341	8	7500341CD1	18	7500341CB1
7500846	9	7500846CD1	19	7500846CB1
3230921	10	3230921CD1	20	3230921CB1

WO 02/097060

PCT/US02/18354

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	Annotation
1	724717CD1	g2104689	1.4E-258	[Mus musculus] alpha-glucosidase II, alpha subunit [Arendt, C.W. and Ostergaard, H.L. (1997) J. Biol. Chem. 272:1317-1325.
2	5596327CD1	g2621120	1.9E-17	[Melioidi thermobacter thermotrophicus] O-linked GlcNAc transferase [Smith, D.R. et al. (1997) J. Bacteriol. 179:7135-7155.
3	4179651CD1	g14150450	0.0	[Rattus norvegicus] UDP-GalNAc:polypeptide N-acetyl(galactosaminyl)transferase T9 [Ten Hagen, K.G. et al. (2001) J. Biol. Chem. 276:17395-17404.
4	7482109CD1	g163124	5.8E-77	[Bos taurus] alpha-1,3-galactosyltransferase [Jozaissie, D.H. et al. (1989) J. Biol. Chem. 264:14290-14297.
5	2867618CD1	g2642187	4.4E-182	[Rattus norvegicus] endo-alpha-D-mannosidase [Spero, M.J. et al. (1997) J. Biol. Chem. 272:29356-29363.
6	7488348CD1	g211179	4.1E-243	[Homo sapiens] endoglucanase [Homo sapiens] beta-1,6-N-acetylglucosaminyltransferase B [Chen, G.Y. et al. (2000) Glycobiology 10:1001-1011.
7	5539166CD1	g9650954	2.1E-177	[Homo sapiens] DPM2 [Maeda, Y. et al. (1998) EMBO J. 17:4920-4929.
8	7500341CD1	g5790363	5.4E-30	[Lemon, G. et al. (1996) Genomics 33:151-152.
		339872DPM2	4.8E-31	[Homo sapiens] [Regulatory subunit; Transferase] [Endoplasmic reticulum; Cytoplasmic] Dolichol phosphate mannose synthase regulatory subunit, regulates synthesis of dolichol phosphate mannose, which is a donor of mannose residues for N-linked oligosaccharide precursors and for the core of glycosylphosphatidylinositol anchors
9	7500846CD1	g806221	2.1E-83	[Homo sapiens] intelectin [Tsujii, S. et al. (2001) J. Biol. Chem. 276:23456-23463.
		5851071Idn	4.2E-80	[Mus musculus] Intelectin, expressed in small intestinal Paneth cells and may have a role in the defense against microorganisms
10	3230921CD1	g14043157	1.1E-51	[Homo sapiens] cargo selection protein (mannose-6-phosphate receptor binding protein) [Diaz, E. and Pfeffer, S.R. (1998) Cell 93:433-443.

WO 02/097060

PCT/US02/18354

Table 2

Polypeptide SEQ ID NO:	Inverte Polypeptide ID	GenBank ID NO:	Probability Score	Annotation
10 cont.		342866 TIP47	9.2E-53	[Homo sapiens]Endosome/Endosomal vesicles; Cytoplasmic; Tail-interacting protein of 47 kDa, protein that binds to the cytoplasmic domain of mannose-6-phosphate receptors and mediates their transport from endosomes to the Golgi apparatus; serum levels are elevated in cervical carcinoma and normal pregnancy Oresl, J. G., et al. (2000) Proc. Natl. Acad. Sci. USA 97:9047-9051.
		339060 ADFP	9.8E-33	[Homo sapiens]Endoplasmic reticulum; Cytoplasmic; Lipid particles; Adipose differentiation-related protein, a membrane-associated protein involved in fatty acid transport, a marker of lipid accumulation
		583559 Adfp	2.7E-26	[Mus musculus]Small molecule-binding protein[Cytoplasmic; Unspecified membrane] Adipose differentiation-related protein, a membrane-associated protein involved in fatty acid and cholesterol binding and long-chain fatty acid transport, a marker of lipid accumulation
		337032 PIN	2.8E-16	Trusey, J. T., et al. Proc. Natl. Acad. Sci. U.S.A. 98:6494-9. (2001) [Homo sapiens] Perilipin, may be multiply phosphorylated by protein kinase A in adipocytes
		329340 Plin	3.5E-16	[Rattus norvegicus] Perilipin, regulates lipolysis and is multiply phosphorylated by protein kinase A in adipocytes

WO 02/097060

PCT/US02/18354

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7247177CD1	912	S10 S38 S56 S145 S427 S477 S577 S804 S809 S835 S844 T42 T51 T52 T103 T108 T180 T184 T290 T340 T650 T732 T759 T907 Y246 Y449	N286 N843	Glycosyl hydrolases family: P138-T875	HMMER_PPAM
					Transmembrane domain: W592-F620 N-terminus is cytosolic	TMAP
					Glycosyl hydrolases family BL00129: G353-D398, G464-Y490, R565-W592, L603-A646, R667-F702, T783-S820	BLMP5_BLOCKS
					HYDROLASE GLYCOSIDASE GLYCOPROTEIN ALPHA-GLUCOSIDASE MALTASE PRECURSOR SIGNAL PROTEIN GLUCOSIDASE SUCRASE ISOMALTASE	BLAST_PRODOM
					PD001543: W195-F518, Q523-T874 HYDROLASE GLYCOSIDASE ALPHA- GLUCOSIDASE GLYCOPROTEIN MALTASE SUCRASE ISOMALTASE PRECURSOR SIGNAL ACID INTESTINAL	BLAST_PRODOM
					PD001716: G329-W472, H224-Y302 SUCRASE/ISOMALTASE DM01359[P1025]P283-930: H537-L598 DM01359[P2379]P278-931: Q523-A906 DM01359[P14410]P268-925: N534-I905 DM01359[P14024]P99-864: E522-Q822	BLAST_DOMO

Table 3

SRQ ID	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1				Glycosyl hydrolase family 31 signature: P507-E514	MOTIFS
2	5596327CD1	691	S80 S137 S198 S291 S327 S510 S615 S658 S664 S674 S684 Y99 T541 T561 T671 Y270		ATP synthase Alpha chain, C-terminal: R128-L142 TPK Domain: I149-H182, V183-N216, P115-F148, P48-A81, R82-Q114 Transmembrane domain: E575-Y591, P595-K617 N-terminus is non-cytosolic MITOCHONDRIAL OUTER MEMBRANE PROTEIN, Y9K Glycosyl transferase: S31-G236	HMIMER_PPAM
3	4179651CD1	506	S21 S29 S40 S75 S90 S108 S141 S253 S319 S337 S443 S465 S470 T2 T114 T137 T409 T422 T437 T498 Y196 Y238 Y459	N27 N49 N496	BLAST_DOMO	HMIMER_PPAM

WO 02/097060

PCT/US02/18354

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	cont.				N-ACETYL GALACTOSAMINYLTRANSFERASE TRANSFERASE POLYPEPTIDE ACETYL GALACTOSAMINYLTRANSFERASE UDP GALNAc:POLYPEPTIDE GLYCOSYLTRANSFERASE PROTEIN UDP PROTEIN UDP N PD0031621202-A362	BLAST_PRODOM
					do ACETYL GALACTOSAMINYLTRANSFERASE: POLYPEPTIDE DM03891Q075712-558: D3-F503 DM03891P3467837-600: D8-E893 DM03891H3740521-571: Y11-F492 signal_cleavage: M1-V19	BLAST_DOMO
4	7482109CD1	276	S3 S153 S260 T47 T84 T131 T139 T213 Y70	N74 N259	Signal Peptide: M1-R21, M1-Y22, M1-R24 TRANSFERASE ALPHA 1.3- GALACTOSYLTRANSFERASE N- ACETYL GALACTOSAMINIDE GALACTOSYLTRANSFERASE UDP- GALACTOSE: BETA D-GALACTOSYL 1 GROUP 4-N-ACETYL D-GLUCOSAMINIDE GLYCOSYLTRANSFERASE PD010022-K41-T276 SIGNAL-ANCHOR TRANSMEM (GALACTOSYLTRANSFERASE FUCOSYLGlycoprotein ALPHA-1 ALPHA) DM07533P147696-367: E20-T276 DM07533P1644216-353: R42-T276	SPSCAN HMMER BLAST_PRODOM
						BLAST_DOMO

WO 02/097060

PCT/US02/18354

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	cont.				dc GALACTOSYLTRANSFERASE, ALPHA-1; ACETYLLACTOSAMINIDE; DM08008 A56480 1-376; E20-T276 DM08003 I49698 1-371; R24-T276	BLAST_DOMO
5	286761RCD1	378	S80 S85 S206 S256 S280 S317 S336 S359 T125 T190 T325 T355 Y369		ENDO-ALPHA-D-MANNOSIDASE PD141586; P3-S378	BLAST_PRODOM
6	748834RCD1	438	S263 S272 S306 S370 S392 S424 T27 T246 T341 T421 Y61 Y279	N76 N108	Enolase: S2-R456	HMMER_Pfam
					Enolase proteins BL00164; F403-A441, R35-D57, I153-K202, N229-K271, P202-D316, I331-N366	BLIMPS_BLOCKS
					Enolase signature: Q314-L362	PROFILESKAN
					Enolase signature PR00148; Y38-E52, G113-G129, A173-R186, G335-R346, L361-L373, S392-V409	BLIMPS_PRINTS
					ENOLASELYASE GLYCOLYSIS MAGNESIUM HYDROLYASE 2-PHOSPHOGLYCERATE DEHYDRATASE 2-PHOSPHO-D-GLYCERATE PHOSPHO-D-GLYCERATE PD000902; M1-Y293	BLAST_PRODOM

WO 02/097060

PCT/US02/18354

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	cont.				LYASE MAGNESIUM ENOLASE GLYCOLYSIS HYDROL YASE DEHYDRATASE 2- PHOSPHOGLYCERATE 2-PHOSPHO-D- GLYCERATE 2-PHOSPHO-D-GLYCERATE PD003216: D275-E440	BLAST_PRODOM
					ENOLASE LYASE GLYCOLYSIS MAGNESIUM HYDROL YASE 2-PHOSPHOGLYCERATE DEHYDRATASE 2-PHOSPHO-D-GLYCERATE 2- PHOSPHO-D-GLYCERATE PD000948: V408-R456	BLAST_PRODOM
					ENOLASE DM00487[S2207]P: 455: S2-R456 DM00487[S52838]P: 431: M1-P455 DM00487[P1039]P: 431: M1-P455 DM00487[P15429]P: 430: I7-R456 Endase signature: L361-S374	BLAST_DOMO
7	5539166CD1	402	S4 S89 S154 S158 S295 S329 T167 T68 T73 T87 T199 T247 Y206	N41 N316 N390	signal_cleavage: M1-N25	MOTIFS SPSCAN
					Core-2/1-Branching enzyme: T73-L382 Cytosolic Domain: M1-K6 Transmembrane Domain: H7-W29 Non-cytosolic Domain: E30-F402	HMMER PFAM TMHMMER

WO 02/097060

PCT/US02/18354

Table 3

SEQ ID NO.	Protein ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	cont.				PROTEIN BETA 1-TRANSFERASE GLYCOSYLTRANSFERASE ENZYME CORE 6-N- ACETYLGALACTOSAMINYLTRANSFERASE 3- GALACTOSYLOGLYCOSYL GLYCOPROTEIN 6- N-ACETYLGALACTOSAMINYLTRANSFERASE PD065410: L171-H344 PD065338: T73-D170	BLAST_PRODOM
					BETA 1-TRANSFERASE GLYCOSYLTRANSFERASE ENZYME CORE 6-N- ACETYLGALACTOSAMINYLTRANSFERASE 3- GALACTOSYLOGLYCOSYL GLYCOPROTEIN 6- N-ACETYLGALACTOSAMINYLTRANSFERASE BRANCHING PD011484: G345-F402 LUMENAL DOMAIN DM07544/Q064308-399: L9-F402 DM07544/Q0742129-427: C74-R385	BLAST_PRODOM
8	7500341CD1	109	T3 T73		signal cleavage: M1-S67 Cytosolic domains: M1-V8, P69-D109 Transmembrane domains: V91-L31, F46-F68 Non-cytosolic domain: P32-L45	SPSCAN TMHMMER
9	7500846CD1	225	S30 S52 S62 S96 T80 T107 T142		signal cleavage: M1-A26 Signal Peptide: M1-A27, M1-A28, M4-A26, M4-A28, M8-C24, M8-A26, M8-A28, M8-A29 Fibrinogen beta and gamma chains, C-terminus: L49-V94	SPSCAN HMMER HMMER_PPAM

WO 02/097060

PCT/US02/18354

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	cont.				Cytosolic domain: M1-C12 Transmembrane domain: F13-L35 Non-cytosolic domain: S36-G225 LECTIN INTELECTIN CORTICAL GRANULE PRECURSOR GLYCOPROTEIN PD040823; R51-Q200	TMHMMER BLAST_PRODOM
10	3230921CD1	463	S2 S13 S117 S140 S187 S205 S273 S277 S304 T68 T169		signal cleavage: M1-A38 Perilipin family: P10-Y369 PROTEIN PERILIPIN ADIPOSE DIFFERENTIATION RELATED ADRP MEMBRANE CARGO SELECTION TIP47 A/B PD018256; S148-E384, Q19-G128	SPSCAN HMMER_PTAM BLAST_PRODOM

WO 02/097060

PCT/US02/18354

Table 4

Polymer Seq ID NO./ Accession ID/ Sequence Length	Sequence Fragments
11/724717/CB1/ 3150	1-388, 1-394, 1-730, 1-737, 5-338, 5-423, 35-603, 220-394, 228-423, 239-978, 239-980, 285-389, 295-485, 295-905, 297-544, 311-418, 311-709, 454-719, 498-1352, 508-1352, 546-1352, 549-669, 549-751, 549-812, 550-1352, 568-1168, 580-1352, 585-1352, 589-1088, 589-1236, 619-1352, 681-1352, 790-1650, 790-1661, 821-1714, 872-1147, 876-1584, 885-1584, 956-1608, 971-1615, 1002-1621, 1063-1942, 1070-1961, 1075-1277, 1098-1704, 1099-1781, 099-1864, 1184-1635, 1193-1893, 1224-1901, 1235-1955, 1236-1961, 1250-1955, 1315-1574, 1324-1955, 1345-1955, 1398-1967, 1457-1954, 1496-1721, 1498-1949, 1505-1952, 1505-1955, 1512-1955, 1516-1864, 1651-1664, 1786-2256, 1817-2103, 1827-2567, 1827-2584, 1827-2598, 1827-2614, 1827-2637, 1831-2622, 1865-2227, 1892-2474, 1894-2196, 1918-2032, 1923-2092, 1946-2324, 1975-2840, 1988-2386, 2188-2897, 2200-2897, 2205-2897, 2206-2897, 2208-2897, 2413-2526, 2417-2885, 2447-3062, 2489-3150, 2605-2951, 2641-2804, 2789-2932, 2802-2951, 2869-2951, 3113-3138
12/596327/CB1/ 3290	1-788, 303-1028, 315-1031, 591-798, 662-962, 683-960, 683-1394, 812-1327, 840-1045, 887-1605, 890-1054, 894-1126, 1010-1623, 1031-1502, 1052-1144, 1090-1723, 1140-2052, 1349-1637, 1372-1618, 1524-1761, 1594-2220, 1636-1965, 1659-1956, 1662-1946, 1748-2036, 1919-2156, 1919-2371, 1919-2397, 2062-2282, 2214-2912, 2256-2538, 2256-2569, 2285-2466, 2285-2593, 2397-2603, 2397-2605, 2424-2682, 2424-2698, 2427-2854, 2453-3065, 2462-2765, 2462-3197, 2495-2631, 2543-3211, 2784-3290, 2801-3060, 2924-3185, 2941-3188, 3072-3154, 3072-3221
13/417961/CB1/ 4279	1-695, 2-126, 7-281, 265-4279, 267-440, 301-523, 380-573, 389-1062, 351-1058, 560-1043, 560-1108, 857-1252, 938-1061, 994-1560, 1081-1379, 1169-1391, 1265-1556, 1328-1369, 1373-1993, 1393-1433, 1554-1786, 1554-2093, 1754-2218, 1756-2060, 1793-2052, 1825-2259, 1874-2127, 1875-2460, 1922-2189, 1932-2180, 1967-2451, 1969-2254, 1981-2640, 2042-2534, 2307-2541, 2357-2783, 2389-2670, 2451-2923, 2455-3049, 2464-2724, 2510-3099, 2562-3072, 2619-2856, 2662-3273, 2663-2858, 2666-3246, 2719-2948, 2743-2995, 2747-3269, 2778-3037, 2986-3292, 3139-3416, 3162-3390, 3162-3656, 3186-3469, 3187-3859, 3199-3501, 3210-3493, 3232-3800, 3280-3480, 3296-3818, 3333-3593, 3335-3617, 3340-3605, 3362-3609, 3362-3835, 3362-3843, 3406-3764, 3432-3860, 3525-3751, 3556-3781, 3556-3782, 3641-3905, 3695-3878, 3698-3866, 3698-3870, 3709-3953, 3709-4209, 3933-4202, 3953-4214, 3968-4107, 4033-4279

WO 02/097060

PCT/US02/18354

Table 4

Polymer/oligonucleotide SEQ ID NO./ Nucleotide ID/ Sequence Length	Sequence Fragments
1477482109CB1/ 1042	1-119, 70-257, 72-831, 677-1042, 678-1042, 680-1041, 680-1042
1512867618CB1/ 4320	1-641, 1-633, 565-1278, 877-1420, 1055-1618, 1079-1693, 1106-1729, 1142-1922, 1188-1774, 1338-2100, 1358-1974, 1394-2074, 1478-1760, 1481-1878, 1760-2479, 1820-2098, 1841-2531, 1885-2045, 1977-2517, 1986-2165, 2001-2591, 2035-2507, 2057-2531, 2193-2532, 2140-2524, 2145-2747, 2151-2532, 2172-2532, 2221-2547, 2225-2493, 2330-3026, 2652-2874, 2706-3044, 2710-3396, 2856-3130, 2876-3159, 2892-3127, 2892-3451, 2918-3173, 3031-3596, 3070-3504, 3086-3538, 3108-3586, 3157-3493, 3174-3453, 3246-3549, 3255-3463, 3255-3707, 3308-3579, 3319-4013, 3321-4010, 3342-3695, 3526-4170, 3570-4148, 3575-4261, 3615-3869, 3617-3904, 3675-3942, 3701-3871, 3742-4032, 3742-4035, 3758-4296, 3776-4247, 3789-4264, 3814-4265, 3818-4268, 3820-4083, 3824-4102, 3834-4273, 3844-4016, 3844-4103, 3844-4234, 3846-4100, 3846-4320, 3847-4273, 3849-4267, 3858-4263, 3910-4277, 3938-4221, 3991-4222, 3991-4255, 3991-4266, 4002-4292, 4004-4260, 4037-4273, 4064-4272, 4075-4261, 4080-4268, 4085-4273, 4113-4264
167748548CB1/ 1733	1-1733, 33-1404
175539166CB1/ 2201	1-430, 209-939, 228-476, 239-1055, 262-859, 350-991, 453-1120, 457-1013, 457-1019, 611-1443, 682-869, 756-963, 756-1218, 756-1230, 820-1338, 1079-1530, 1411-1529, 1469-1998, 1469-2107, 1470-1758, 1470-1908, 1470-1943, 1470-2011, 1470-2065, 1470-2066, 1470-2201
187500341CB1/ 487	1-126, 1-218, 1-230, 1-457, 2-230, 12-126, 15-164, 16-230, 17-223, 19-230, 20-230, 22-168, 23-230, 28-175, 28-230, 29-230, 31-209, 36-445, 36-451, 37-202, 38-230, 42-202, 50-452, 228-428, 232-445, 310-446
197500846CB1/ 1513	1-268, 1-276, 1-495, 1-593, 1-1020, 4-266, 8-578, 13-221, 16-283, 112-643, 214-499, 246-678, 301-1011, 332-998, 384-648, 442-861, 654-803, 660-1023, 669-948, 669-972, 669-1013, 669-1023, 669-1032, 693-1023, 775-1020, 854-1513, 871-1111

Table 4

Polynucleotide SEQ ID NO./ Ineye ID/ Sequence Length	Sequence Fragments
20/3230921CBI/ 2084	1-240, 1-261, 1-266, 28-291, 28-650, 33-619, 36-332, 46-717, 47-896, 53-583, 115-354, 115-660, 160-297, 160-599, 160-818, 231-355, 231-369, 231-378, 231-422, 231-438, 238-438, 250-438, 254-438, 343-593, 439-568, 439-576, 439-582, 439-604, 439-613, 439-614, 439-619, 443-619, 444-572, 445-619, 449-619, 453-619, 456-619, 458-619, 466-619, 468-689, 468-795, 468-1004, 472-619, 483-1054, 483-1124, 497-616, 620-889, 620-1000, 620-1058, 620-1118, 641-1123, 646-1110, 653-597, 664-1126, 700-771, 716-1118, 725-1118, 750-1127, 751-1118, 759-1118, 931-1027, 957-1127, 1048-1118, 1068-1411, 1119-1915, 1120-1696, 1128-1677, 1144-1325, 1144-1446, 1144-1483, 1144-1605, 1144-1624, 1164-1408, 1164-1635, 1167-1652, 1200-1583, 1217-1333, 1223-1770, 1250-1893, 1261-1612, 1300-1691, 1304-1569, 1304-1750, 1319-1916, 1325-1920, 1401-1698, 1419-1722, 1419-1921, 1467-1727, 1482-1794, 1485-2084

WO 02/097060

PCT/US02/18354

Table 5

Polynucleotide SEQ ID No.	Protein Project ID	Representative Library
11	724177CB1	MIXDTXB01
12	5596327CB1	OVARNOT09
13	4179651CB1	BRAGNON02
15	2867618CB1	HPQNON01
17	5539166CB1	KIDNEFC01
18	7500341CB1	LINODNOT03
19	7500846CB1	PROSNOT06
20	3230921CB1	PANCNOT23

WO 02/097060

PCT/US02/18354

Table 6

Library	Vector	Library Description
BRAGNON02	pNCY	This normalized substantia nigra tissue library was constructed from 4.2×10^7 independent clones from a substantia nigra tissue library. Starting RNA was made from RNA isolated from substantia nigra tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis.
		Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolosclerosis, nodular colloid goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease. The library was normalized in two rounds using conditions adapted from Soares et al., 1994 (Proc. Natl. Acad. Sci. USA 91:9228-9232) and Bonaldi et al., 1996 (Genome Res. 6:791).
HIPONON01	PSPORT1	This normalized hippocampus library was constructed from 1.13M independent clones from a hippocampus tissue library. RNA was isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis. The normalization and hybridization conditions were adapted from Soares et al., 1994 (Proc. Natl. Acad. Sci. USA 91:9228-9232).
KIDNREC01	PBLUESCRIPT	Library was constructed using RNA isolated from kidney tissue removed from a pool of twelve Caucasian male and female fetuses that were spontaneously aborted at 19-23 weeks' gestation.
LNODNOT03	pNCY	Library was constructed using RNA isolated from lymph node tissue obtained from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. On microscopic exam, this tissue was found to be extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, congestive heart failure, atherosclerotic coronary artery disease.

WO 02/097060

PCT/US02/18354

Table 6

Library	Vector	Library Description
MIXDTEX01	PBK-CMV	This 5' biased random primed library was constructed using pooled cDNA from nine donors. cDNA was generated using mRNA isolated from Jurkat cell line derived from the T cells of a male (donor A), THP-1 cell line derived from the peripheral blood of a 1-year-old male (donor B), Daudi cell line derived from B-lymphoblasts from a 16-year-old black male (donor C), RPMI-1666 cell line derived from lymphoma tissue from a 29-year-old Caucasian male (donor D), spleen from a 1-year-old Caucasian male (donor E), thymus removed from a 21-year-old Caucasian male (donor F) during a thymectomy, lymph node from a 42-year-old Caucasian female (donor G), thymus tumor from a 56-year-old Caucasian female (donor H) during a total thymectomy and PBMC's from a pool of donors (donor I). The patients presented with anemia and persistent hyperplastic thymus (H). Patient history included acute T-cell leukemia (A); acute monocytic leukemia (B); Burkitt's lymphoma (C); Hodgkin's disease (D); Bronchitis (E); hydrocele, regional enteritis or the small intestine, atopic dermatitis and benign neoplasm of the parathyroid (F); heart murmur and cardiac arrest (G); and cardiac dysrhythmia and left bundle branch block (H). Previous surgeries included an appendectomy and parathyroid surgery (F); unspecified heart surgery (G); and a normal delivery (H). Family history included benign hypertension in the grandparent(s) and coronary artery disease in the father of donor F.
OVARNOT09	pINCY	Library was constructed using RNA isolated from ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Pathology indicated multiple follicular cysts ranging in size from 0.4 to 1.5 cm in the right and left ovaries, chronic cervicitis and squamous metaplasia of the cervix, and endometrium in weakly proliferative phase. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.
PANCN0123	pINCY	Library was constructed using RNA isolated from diseased pancreatic tissue removed from a 43-year-old Caucasian female who died from a gunshot wound to the head. Patient history included type I diabetes and serology positive CMV antibody.
PROSNOT06	PSFORT1	Library was constructed using RNA isolated from the diseased prostate tissue of a 57-year-old Caucasian male during radical prostatectomy, removal of both testes and excision of regional lymph nodes. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma (Gleason grade 3+3). Patient history included a benign neoplasm of the large bowel and type I diabetes. Family history included a malignant neoplasm of the prostate and type I diabetes.

WO 02/097060

PCT/US02/18354

Table 7

Program	Description	Reference	Parameter Threshold
ABI/FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABUPARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value=1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfastn, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity=95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCI, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nussliel, Cambridge Univ. Press pp. 1-350.	PFAM, INCI, SMART, or TIGRFAM hits: Probability value=1.0E-3 or less Signal peptide hits: Score=0 or greater

WO 02/097060

PCT/US02/18354

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores>GCC-specified "HGR" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A 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 Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nelson, 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.	Bairoch, 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 02/097060

PCT/US02/18354

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10,
 - 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:1-3, SEQ ID NO:5, and SEQ ID NO:7-10,
 - c) a polypeptide comprising a naturally occurring amino acid sequence at least 93% identical to the amino acid sequence of SEQ ID NO:4,
 - d) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:6,
 - e) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, and
 - f) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:11-20.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:

WO 02/097060

PCT/US02/18354

- 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
- 5 b) recovering the polypeptide so expressed.
- 10 10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.
11. An isolated antibody which specifically binds to a polypeptide of claim 1.
12. 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:11-20,
- 15 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:11-20,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- 20 e) an RNA equivalent of a)-d).
13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.
- 25 14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under
- 30 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.
- 35 15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

WO 02/097060

PCT/US02/18354

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
 - 5 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.
- 10 18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.
19. A method for treating a disease or condition associated with decreased expression of functional CHOP, comprising administering to a patient in need of such treatment the composition of claim 17.
- 15 20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
 - 20 b) detecting agonist activity in the sample.
21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.
- 25 22. A method for treating a disease or condition associated with decreased expression of functional CHOP, comprising administering to a patient in need of such treatment a composition of claim 21.
- 30 23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
 - b) detecting antagonist activity in the sample.
- 35 24. A composition comprising an antagonist compound identified by a method of claim 23

WO 02/097060

PCT/US02/18354

and a pharmaceutically acceptable excipient.

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

5

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

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- 10 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.

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

- 15 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 20 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.

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

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- 30 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.

29. A method of assessing toxicity of a test compound, the method comprising:

- 35 a) treating a biological sample containing nucleic acids with the test compound,

WO 02/097060

PCT/US02/18354

- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- 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.

30. A diagnostic test for a condition or disease associated with the expression of CHOP in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

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

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

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

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of CHOP

WO 02/097060

PCT/US02/18354

in a subject, comprising administering to said subject an effective amount of the composition of claim 34,

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, 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 specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

37. A polyclonal antibody produced by a method of claim 36.

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

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-10, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10.

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

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

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab

WO 02/097060

PCT/US02/18354

expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

5

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 10 b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-10 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID
- 20 NO:1-10.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

25 47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a
- 30 hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first

35 oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous

WO 02/097060

PCT/US02/18354

nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is
5 completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is
completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

10 51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is
completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

15 53. An array of claim 48, further comprising said target polynucleotide hybridized to a
nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to
said solid substrate.

20 55. An array of claim 48, wherein each distinct physical location on the substrate contains
multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical
location have the same sequence, and each distinct physical location on the substrate contains
nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at
25 another distinct physical location on the substrate.

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

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

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

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

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

WO 02/097060

PCT/US02/18354

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.
- 5 63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.
64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
- 10 66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.
67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.
- 15 68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.
69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.
- 20 70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.
- 25 71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.
72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.
- 30 73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.
74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
- 35

WO 02/097060

PCT/US02/18354

NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:20.

5

WO 02/097060

PCT/US02/18354

```

<110> INCYTE GENOMICS, INC.
      SWARNAKAR, Anita
      GORVAD, Ann E.
      HAFALIA, April J.A.
      DUGGAN, Brendan M.
      EMERLING, Brooke M.
      ISON, Craig H.
      NGUYEN, Dannie B.
      LEE, Ernestine A.
      YUE, Henry
      YUE, Huibin
      FORSYTHE, Ian J.
      LI, Joana X.
      THANGAVELU, Kavitha
      WALIA, Narinder K.
      BURFORD, Neil
      MARSON, Patricia M.
      LAL, Preeti G.
      LEE, Sally
      BECHA, Shanya D.
      TANG, Y. Tom

<120> Carbohydrate-Associated Proteins

<130> PF-0982 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/293,768; 60/309,548; 60/314,400; 60/343,706; 60/337,999
<151> 2001-05-25; 2001-08-01; 2001-08-23; 2001-10-19; 2001-12-07

<160> 20

<170> PERL Program

<210> 1
<211> 912
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7247177CD1

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

```

WO 02/097060

PCT/US02/18354

Ala Asn Pro Phe	125	130	135
Lys Val Asp Leu Val	140	Ser Glu Glu Glu Val	145
Ile Ser Ile Asn	145	Ser Glu Glu His Leu	150
Ser Leu Gly Gln Leu	155	Tyr Phe Glu His Leu	155
Ile Leu His Lys	160	Gln Arg Ala Ala Lys	165
Gln Arg Ala Ala Lys	170	Glu Asn Glu Glu Thr	175
Ser Val Asp Thr	175	Glu Asp Leu Gly Leu Trp	180
Ser Gln Glu Asn Gln	185	Glu Asp Leu Gly Leu Trp	185
Glu Glu Lys Phe	190	Lys Ala Asn Gly Pro	195
Gly Lys Phe Val Asp	200	Ile Lys Ala Asn Gly Pro	205
Ser Ser Ile Gly	205	His Gly Phe Glu His Leu	210
Leu Asp Phe Ser Leu	215	His Gly Phe Glu His Leu	215
Tyr Gly Ile Pro	220	His Gln Leu Lys Asn Thr	225
Gln His Ala Glu Ser	230	His Gln Leu Lys Asn Thr	235
Gly Asp Gly Asp	235	Asn Leu Asp Val Tyr Gly	240
Ala Tyr Arg Leu Tyr	245	Asn Leu Asp Val Tyr Gly	245
Tyr Gln Ile Tyr	250	Tyr Gly Ser Val Pro Tyr	255
Asp Lys Met Gly Ile	260	Tyr Gly Ser Val Pro Tyr	265
Leu Leu Ala His	265	Ile Gly Ile Phe Trp Leu	270
Lys Leu Gly Arg Thr	275	Ile Gly Ile Phe Trp Leu	275
Asn Ala Ser Glu Thr	280	Asn Thr Glu Pro Ala Val	285
Leu Val Glu Ile	290	Asn Thr Glu Pro Ala Val	295
Glu Tyr Thr Leu Thr	295	Val Ala Ala Lys Gln Lys	300
Gln Met Gly Pro	305	Val Ala Ala Lys Gln Lys	305
Val Arg Ser Arg	310	Met Ser Glu Ser Gly Ile	315
Thr His Val His Trp	320	Met Ser Glu Ser Gly Ile	325
Ile Asp Val Phe	325	Pro Ser Asp Val Phe	330
Leu Thr Gly Pro	335	Pro Ser Asp Val Phe	335
Lys Gln Tyr Ser	340	Gln Ala Met Pro Pro Leu	345
His Leu Thr Gly Thr	350	Gln Ala Met Pro Pro Leu	355
Phe Ser Leu Gly Tyr	355	Asn Tyr Glu Asp Glu	360
His Gln Cys Arg Trp	365	Asn Tyr Glu Asp Glu	365
Gln Asp Val Lys	370	Phe Asp Glu His Asp Ile	375
Ala Val Asp Ala Gly	380	Phe Asp Glu His Asp Ile	385
Pro Tyr Asp Ala Met	385	Glu His Thr Glu Gly Lys	390
Trp Leu Asp Ile	395	Glu His Thr Glu Gly Lys	395
Arg Tyr Phe Thr	400	Pro Asn Pro Lys Arg	405
Trp Asp Lys Asn Arg	410	Pro Asn Pro Lys Arg	415
Met Gln Glu Leu	415	Arg Lys Leu Val Val Ile	420
Leu Arg Ser Lys Lys	425	Arg Lys Leu Val Val Ile	425
Ser Asp Pro His	430	Tyr Ser Val Tyr Val	435
Ile Lys Ile Asp Pro	440	Tyr Ser Val Tyr Val	445
Lys Ala Lys Asp	445	Lys Asn Gln Glu Gly Glu	450
Gln Gly Phe Phe Val	455	Lys Asn Gln Glu Gly Glu	455
Asp Phe Glu Gly Val	460	Leu Ser Ser Tyr Leu Asp	465
Cys Trp Pro Gly	470	Leu Ser Ser Tyr Leu Asp	475
Phe Thr Asn Pro	475	Tyr Ser Ser Leu Phe Ala	480
Lys Val Arg Glu Trp	485	Tyr Ser Ser Leu Phe Ala	485
Phe Pro Val Tyr	490	Ile Leu Phe Leu Trp Asn	495
Gln Gly Ser Thr Asp	500	Ile Leu Phe Leu Trp Asn	505
Asp Met Asn Glu Pro	505	Pro Glu Gln Thr Met	510
Ser Val Phe Arg Gly	515	Pro Glu Gln Thr Met	515
Gln Lys Asn Ala	520	Tyr Glu His Arg Glu Leu	525
Ile His His Gly Asn	530	Tyr Glu His Arg Glu Leu	535
His Asn Ile Tyr	535	Met Ala Thr Ala Glu Gly	540
Gly Phe Tyr His Gln	545	Met Ala Thr Ala Glu Gly	545
Leu Ile Lys Arg	550	Pro Phe Val Leu Thr	555
Ser Lys Gly Lys Glu	560	Pro Phe Val Leu Thr	565
Arg Ser Phe Phe	565	Tyr Gly Ala Val Trp Thr	570
Ala Gly Ser Gln Lys	575	Tyr Gly Ala Val Trp Thr	575
Gly Asp Asn Thr	580	Leu Lys Ile Ser Ile Pro	585
Ala Glu Trp Ser Asn	590	Leu Lys Ile Ser Ile Pro	595
	595		600

WO 02/097060

PCT/US02/18354

```

Met Leu Leu Thr Leu Ser Ile Thr Gly Ile Ser Phe Cys Gly Ala
605 610 615
Asp Ile Gly Gly Phe Ile Gly Asn Pro Glu Thr Glu Leu Leu Val
620 625 630
Arg Trp Tyr Gln Ala Gly Ala Tyr Gln Pro Phe Phe Arg Gly His
635 640 645
Ala Thr Met Asn Thr Lys Arg Arg Ala Pro Trp Leu Phe Gly Glu
650 655 660
Glu His Thr Arg Leu Ile Arg Glu Ala Ile Arg Glu Arg Tyr Gly
665 670 675
Leu Leu Pro Tyr Trp Tyr Ser Leu Phe Tyr His Ala His Val Ala
680 685 690
Ser Gln Pro Val Met Arg Pro Leu Trp Val Glu Phe Pro Asp Glu
695 700 705
Leu Lys Thr Phe Asp Met Glu Asp Glu Tyr Met Leu Gly Ser Ala
710 715 720
Leu Leu Val His Pro Val Thr Glu Pro Lys Ala Thr Thr Val Asp
725 730 735
Val Phe Leu Pro Gly Ser Asn Glu Val Trp Tyr Asp Tyr Lys Thr
740 745 750
Phe Ala His Trp Glu Gly Gly Cys Thr Val Lys Ile Pro Val Ala
755 760 765
Leu Asp Thr Ile Pro Val Phe Gln Arg Gly Gly Ser Val Ile Pro
770 775 780
Ile Lys Thr Thr Val Gly Lys Ser Thr Gly Trp Met Thr Glu Ser
785 790 795
Ser Tyr Gly Leu Arg Val Ala Leu Ser Thr Lys Gly Ser Ser Val
800 805 810
Gly Glu Leu Tyr Leu Asp Asp Gly His Ser Phe Gln Tyr Leu His
815 820 825
Gln Lys Gln Phe Leu His Arg Lys Phe Ser Phe Cys Ser Ser Val
830 835 840
Leu Ile Asn Ser Ser Ala Asp Gln Arg Gly His Tyr Pro Ser Lys
845 850 855
Cys Val Val Glu Lys Ile Leu Val Leu Gly Phe Arg Lys Glu Pro
860 865 870
Ser Ser Val Thr Thr His Ser Ser Asp Gly Lys Asp Gln Pro Val
875 880 885
Ala Phe Thr Tyr Cys Ala Lys Thr Ser Ile Leu Ser Leu Glu Lys
890 895 900
Leu Ser Leu Asn Ile Ala Thr Asp Trp Glu Gly Ser
905 910

```

<210> 2
 <211> 691
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5596327CD1

```

<400> 2
Met Leu Gln Glu Glu Pro Asp Leu Val Ser Ala Ile Tyr Gly Arg
1 10 15
Gly Ile Ala Tyr Gly Lys Lys Gly Leu His Asp Ile Lys Asn Ala
20 25 30
Glu Leu Ala Leu Phe Glu Leu Ser Arg Val Ile Thr Leu Glu Pro
35 40 45
Asp Arg Pro Glu Val Phe Glu Gln Arg Ala Glu Ile Leu Ser Pro
50 55 60
Leu Gly Arg Ile Asn Glu Ala Val Asn Asp Leu Thr Lys Ala Ile
65 70 75

```

WO 02/097060

PCT/US02/18354

Gln Leu Gln Pro Ser Ala Arg Leu Tyr Arg His Arg Gly Thr Leu 80 85 90
 Tyr Phe Ile Ser Glu Asp Tyr Ala Thr Ala His Glu Asp Phe Gln 95 100 105
 Gln Ser Leu Glu Leu Asn Lys Asn Gln Pro Ile Ala Met Leu Tyr 110 115 120
 Lys Gly Leu Thr Phe Phe His Arg Gly Leu Leu Lys Glu Ala Ile 125 130 135
 Glu Ser Phe Lys Glu Ala Leu Lys Gln Lys Val Asp Phe Ile Asp 140 145 150
 Ala Tyr Lys Ser Leu Gly Gln Ala Tyr Arg Glu Leu Gly Asn Phe 155 160 165
 Glu Ala Ala Thr Glu Ser Phe Gln Lys Ala Leu Leu Leu Asn Gln 170 175 180
 Asn His Val Gln Thr Leu Gln Leu Arg Gly Met Met Leu Tyr His 185 190 195
 His Gly Ser Leu Gln Glu Ala Leu Lys Asn Phe Lys Arg Cys Leu 200 205 210
 Gln Leu Glu Pro Tyr Asn Glu Val Cys Gln Tyr Met Lys Gly Leu 215 220 225
 Ser His Val Ala Met Gly Gln Phe Tyr Glu Gly Ile Lys Ala Gln 230 235 240
 Thr Lys Val Met Leu Asn Asp Pro Leu Pro Gly Gln Lys Ala Ser 245 250 255
 Pro Glu Tyr Leu Lys Val Lys Tyr Leu Arg Glu Tyr Ser Arg Tyr 260 265 270
 Leu His Ala His Leu Asp Thr Pro Leu Thr Glu Tyr Asn Ile Asp 275 280 285
 Val Asp Leu Pro Gly Ser Phe Lys Asp His Trp Ala Lys Asn Leu 290 295 300
 Pro Phe Leu Ile Glu Asp Tyr Glu Glu Gln Pro Gly Leu Gln Pro 305 310 315
 His Ile Lys Asp Val Leu His Gln Asn Phe Glu Ser Tyr Lys Pro 320 325 330
 Glu Val Gln Glu Leu Ile Cys Val Ala Asp Arg Leu Gly Ser Leu 335 340 345
 Met Gln Tyr Glu Thr Pro Gly Phe Leu Pro Asn Lys Arg Ile His 350 355 360
 Arg Ala Met Gly Leu Ala Ala Leu Glu Val Met Gln Ala Val Gln 365 370 375
 Arg Thr Trp Thr Asn Ser Lys Val Arg Met Asn Gly Lys Thr Arg 380 385 390
 Leu Met Gln Trp Arg Asp Met Phe Asp Ile Ala Val Lys Trp Arg 395 400 405
 Arg Ile Ala Asp Pro Asp Gln Pro Val Leu Trp Leu Asp Gln Met 410 415 420
 Pro Ala Arg Ser Leu Ser Arg Gly Phe Asn Asn His Ile Asn Leu 425 430 435
 Ile Arg Gly Gln Val Ile Asn Met Arg Tyr Leu Glu Tyr Phe Glu 440 445 450
 Lys Ile Leu His Phe Ile Lys Asp Arg Ile Leu Val Tyr His Gly 455 460 465
 Ala Asn Asn Pro Lys Gly Leu Leu Glu Val Arg Glu Ala Leu Glu 470 475 480
 Lys Val His Lys Val Glu Asp Leu Leu Pro Ile Met Lys Phe Asn 485 490 495
 Thr Lys Thr Lys Asp Gly Phe Thr Val Asn Thr Lys Val Pro Ser 500 505 510
 Leu Lys Asp Gln Gly Lys Glu Tyr Asp Gly Phe Thr Ile Thr Ile 515 520 525
 Thr Gly Asp Lys Val Gly Asn Ile Leu Phe Ser Val Glu Thr Gln 530 535 540
 Thr Thr Glu Glu Arg Thr Gln Leu Tyr His Ala Glu Ile Asp Ala

WO 02/097060

PCT/US02/18354

```

545      550      555
Leu Tyr Lys Asp Leu Thr Ala Lys Gly Lys Val Leu Ile Leu Ser
560      565
Ser Glu Phe Gly Glu Ala Asp Ala Val Cys Asn Leu Ile Leu Ser
575      580
Leu Val Tyr Tyr Phe Tyr Asn Leu Met Pro Leu Ser Arg Gly Ser
590      595
Ser Val Ile Ala Tyr Ser Val Ile Val Gly Ala Leu Met Ala Ser
605      610
Gly Lys Glu Val Ala Gly Lys Ile Pro Lys Gly Lys Leu Val Asp
620      625
Phe Glu Ala Met Thr Ala Pro Gly Ser Glu Ala Phe Ser Lys Val
635      640
Ala Lys Ser Trp Met Asn Leu Lys Ser Ile Ser Pro Ser Tyr Lys
650      655
Thr Leu Pro Ser Val Ser Glu Thr Phe Pro Thr Leu Arg Ser Met
665      670
Ile Glu Val Leu Asn Thr Asp Ser Ser Pro Arg Cys Leu Lys Lys
680      685
Leu

```

```

<210> 3
<211> 506
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Inocyte ID No: 4179551CD1

```

```

<400> 3
Met Thr Asp Ala Glu Arg Val Asp Gln Ala Tyr Arg Glu Asn Gly
1      5      10      15
Phe Asn Ile Tyr Val Ser Asp Lys Ile Ser Leu Asn Arg Ser Leu
20      25      30
Pro Asp Ile Arg His Pro Asn Cys Asn Ser Lys Arg Tyr Leu Glu
35      40      45
Thr Leu Pro Asn Thr Ser Tie Ile Ile Pro Phe His Asn Glu Gly
50      55      60
Trp Ser Ser Leu Leu Arg Thr Val His Ser Val Leu Asn Arg Ser
65      70      75
Pro Pro Glu Leu Val Ala Glu Ile Val Leu Val Asp Asp Phe Ser
80      85      90
Asp Arg Glu His Leu Lys Lys Pro Leu Glu Asp Tyr Met Ala Leu
95      100      105
Phe Pro Ser Val Arg Ile Leu Arg Thr Lys Lys Arg Glu Gly Leu
110      115      120
Ile Arg Thr Arg Met Leu Gly Ala Ser Val Ala Thr Gly Asp Val
125      130      135
Ile Thr Phe Leu Asp Ser His Cys Glu Ala Asn Val Asn Trp Leu
140      145      150
Pro Pro Leu Leu Asp Arg Ile Ala Arg Asn Arg Lys Thr Ile Val
155      160      165
Cys Pro Met Ile Asp Val Ile Asp His Asp Asp Phe Arg Tyr Glu
170      175      180
Thr Gln Ala Gly Asp Ala Met Arg Gly Ala Phe Asp Trp Glu Met
185      190      195
Tyr Tyr Lys Arg Ile Pro Ile Pro Pro Glu Leu Gln Lys Ala Asp
200      205      210
Pro Ser Asp Pro Phe Glu Ser Pro Val Met Ala Gly Gly Leu Phe
215      220      225
Ala Val Asp Arg Lys Trp Phe Trp Glu Leu Gly Gly Tyr Asp Pro

```

WO 02/097060

PCT/US02/18354

```

230      235      240
Gly Leu Glu Ile Trp Gly Gly Glu Gln Tyr Glu Ile Ser Phe Lys
245      250      255
Val Trp Met Cys Gly Gly Arg Met Glu Asp Ile Pro Cys Ser Arg
260      265      270
Val Gly His Ile Tyr Arg Lys Tyr Val Pro Tyr Lys Val Pro Ala
275      280      285
Gly Val Ser Leu Ala Arg Asn Leu Lys Arg Val Ala Glu Val Trp
290      295      300
Met Asp Glu Tyr Ala Glu Tyr Ile Tyr Gln Arg Arg Pro Glu Tyr
305      310      315
Arg His Leu Ser Ala Gly Asp Val Ala Val Gln Lys Lys Leu Arg
320      325      330
Ser Ser Leu Asn Cys Lys Ser Phe Lys Trp Phe Met Thr Lys Ile
335      340      345
Ala Trp Asp Leu Pro Lys Phe Tyr Pro Pro Val Glu Pro Pro Ala
350      355      360
Ala Ala Trp Gly Glu Ile Arg Asn Val Gly Thr Gly Leu Cys Ala
365      370      375
Asp Thr Lys His Gly Ala Leu Gly Ser Pro Leu Arg Leu Glu Gly
380      385      390
Cys Val Arg Gly Arg Gly Glu Ala Ala Trp Asn Asn Met Gln Val
395      400      405
Phe Thr Phe Thr Trp Arg Glu Asp Ile Arg Pro Gly Asp Pro Gln
410      415      420
His Thr Lys Lys Phe Cys Phe Asp Ala Ile Ser His Thr Ser Pro
425      430      435
Val Thr Leu Tyr Asp Cys His Ser Met Lys Gly Asn Gln Leu Trp
440      445      450
Lys Tyr Arg Lys Asp Lys Thr Leu Tyr His Pro Val Ser Gly Ser
455      460      465
Cys Met Asp Cys Ser Glu Ser Asp His Arg Ile Phe Met Asn Thr
470      475      480
Cys Asn Pro Ser Ser Leu Thr Gln Gln Trp Leu Phe Glu His Thr
485      490      495
Asn Ser Thr Val Leu Glu Lys Phe Asn Arg Asn
500      505

```

<210> 4

<211> 276

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7482109CD1

<400> 4

```

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

```

WO 02/097060

PCT/US02/18354

```

110      115      120
Leu Pro Asp Ile Glu Pro Ser Pro Leu Arg Thr Phe Lys Ala Phe
125      130      135
Lys Val Gly Thr Glu Arg Trp Trp Leu Asp Gly Pro Leu Val His
140      145      150
Val Lys Ser Leu Gly Glu His Ile Ala Ser His Ile Gln Asp Glu
155      160      165
Val Asp Phe Leu Phe Ser Met Ala Ala Asn Gln Val Phe Gln Asn
170      175      180
Glu Phe Gly Val Glu Thr Leu Gly Pro Leu Val Ala Gln Leu His
185      190      195
Ala Trp Trp Tyr Phe Arg Asn Thr Lys Asn Phe Pro Tyr Glu Arg
200      205      210
Arg Pro Thr Ser Ala Ala Cys Ile Pro Phe Gly Gln Gly Asp Phe
215      220      225
Tyr Tyr Gly Asn Leu Met Val Gly Gly Thr Pro His Asn Ile Leu
230      235      240
Asp Phe Ile Lys Glu Tyr Leu Asn Gly Val Ile His Asp Ile Lys
245      250      255
Asn Gly Leu Asn Ser Thr Tyr Glu Lys His Leu Asn Lys Tyr Phe
260      265      270
Tyr Leu Asn Lys Pro Thr
275

```

```

<210> 5
<211> 378
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2867618CD1

```

```

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

```

WO 02/097060

PCT/US02/18354

```

215          220          225
Lys Tyr Asp Ile Leu Gln Arg Val Asp Gly Ile Tyr Thr Tyr Phe
230          235          240
Ala Thr Asn Gly Phe Thr Tyr Gly Ser Ser His Gln Asn Gly Ala
245          250          255
Ser Leu Lys Leu Leu Cys Asp Lys Tyr Asn Leu Ile Phe Ile Pro
260          265          270
Ser Val Gly Pro Gly Tyr Ile Asp Thr Ser Ile Arg Pro Trp Asn
275          280          285
Thr Gln Asn Thr Arg Asn Arg Ile Asn Gly Lys Tyr Tyr Glu Ile
290          295          300
Gly Leu Ser Ala Ala Leu Gln Thr Arg Pro Ser Leu Ile Ser Ile
305          310          315
Thr Ser Phe Asn Glu Trp His Glu Gly Thr Gln Ile Glu Lys Ala
320          325          330
Val Pro Lys Arg Thr Ser Asn Thr Val Tyr Leu Asp Tyr Arg Pro
335          340          345
His Lys Pro Gly Leu Tyr Leu Glu Leu Thr Arg Lys Trp Ser Glu
350          355          360
Lys Tyr Ser Lys Glu Arg Ala Thr Tyr Ala Leu Asp Arg Gln Leu
365          370          375
Pro Val Ser

```

```

<210> 6
<211> 458
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7488348CD1

```

```

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

```

WO 02/097060

PCT/US02/18354

```

215      220      225
Ile Leu Glu Asn Lys Glu Gly Leu Glu Leu Leu Lys Thr Ala Ile
230      235      240
Gly Lys Ala Gly Tyr Thr Asp Lys Val Val Ile Gly Met Asp Val
245      250      255
Ala Ala Ser Glu Phe Phe Arg Ser Gly Lys Tyr Asp Leu Asp Phe
260      265      270
Lys Ser Pro Asp Asp Pro Ser Arg Tyr Ile Ser Pro Asp Gln Leu
275      280      285
Ala Asp Leu Tyr Lys Gly Phe Val Leu Gly His Ala Val Lys Asn
290      295      300
Tyr Pro Val Gly Val Ser Ile Glu Asp Pro Pro Phe Asp Gln Asp
305      310      315
Asp Trp Gly Ala Trp Lys Lys Leu Phe Thr Gly Ser Leu Val Gly
320      325      330
Ile Gln Val Val Gly Asp Asp Leu Thr Val Thr Lys Pro Glu Ala
335      340      345
Arg Ile Ala Lys Ala Val Glu Glu Val Lys Ala Cys Asn Cys Leu
350      355      360
Leu Leu Leu Lys Val Asn Gln Ile Gly Ser Val Thr Glu Ser Leu
365      370      375
Gln Ala Cys Lys Leu Ala Gln Ser Asn Gly Trp Gly Val Met Pro
380      385      390
Val Ser His Arg Leu Ser Gly Glu Thr Glu Asp Thr Phe Met Ala
395      400      405
Asp Leu Val Val Gly Leu Cys Thr Gly Gln Ile Lys Thr Gly Pro
410      415      420
Thr Cys Arg Ser Glu Arg Leu Ala Lys Tyr Asn Gln Leu Leu Arg
425      430      435
Ile Glu Glu Ala Glu Ala Gly Ser Lys Ala Arg Phe Ala Gly Arg
440      445      450
Asn Phe Arg Asn Pro Arg Ile Asn
455

```

```

<210> 7
<211> 402
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 5539166CD1

```

```

<400> 7
Met Met Gly Ser Trp Lys His Cys Leu Phe Ser Ala Ser Leu Ile
1      5      10      15
Ser Ala Leu Ile Phe Val Phe Val Tyr Asn Thr Glu Leu Trp Glu
20     25     30
Asn Lys Arg Phe Leu Arg Ala Ala Leu Ser Asn Ala Ser Leu Leu
35     40     45
Ala Glu Ala Cys His Gln Ile Phe Glu Gly Lys Val Phe Tyr Pro
50     55     60
Thr Glu Asn Ala Leu Lys Thr Thr Leu Asp Glu Ala Thr Cys Tyr
65     70     75
Glu Tyr Met Val Arg Ser His Tyr Val Thr Glu Thr Leu Ser Glu
80     85     90
Glu Glu Ala Gly Phe Pro Leu Ala Tyr Thr Val Thr Ile His Lys
95     100    105
Asp Phe Gly Thr Phe Glu Arg Leu Phe Arg Ala Ile Tyr Met Pro
110    115    120
Gln Asn Val Tyr Cys Val His Leu Asp Gln Lys Ala Thr Asp Ala
125    130    135
Phe Lys Gly Ala Val Lys Gln Leu Leu Ser Cys Phe Pro Asn Ala

```

WO 02/097060

PCT/US02/18354

```

140      145      150
Phe Leu Ala Ser Lys Lys Glu Ser Val Val Tyr Gly Gly Ile Ser
155      160      165
Arg Leu Gln Ala Asp Leu Asn Cys Leu Glu Asp Leu Val Ala Ser
170      175      180
Glu Val Pro Trp Lys Tyr Val Ile Asn Thr Cys Gly Gln Asp Phe
185      190      195
Pro Leu Lys Thr Asn Arg Glu Ile Val Gln Tyr Leu Lys Gly Phe
200      205      210
Lys Gly Lys Asn Ile Thr Pro Gly Val Leu Pro Pro Asp His Ala
215      220      225
Val Gly Arg Thr Lys Tyr Val His Gln Glu Leu Leu Asn His Lys
230      235      240
Asn Ser Tyr Val Ile Lys Thr Thr Lys Leu Lys Thr Pro Pro Pro
245      250      255
His Asp Met Val Ile Tyr Leu Gly Thr Ala Tyr Val Ala Leu Thr
260      265      270
Arg Asp Phe Ala Asn Phe Val Leu Gln Asp Gln Leu Ala Leu Asp
275      280      285
Leu Leu Ser Trp Ser Lys Asp Thr Tyr Ser Pro Asp Glu His Phe
290      295      300
Trp Val Thr Leu Asn Arg Ile Pro Gly Val Pro Gly Ser Met Pro
305      310      315
Asn Ala Ser Trp Thr Gly Asn Leu Arg Ala Ile Lys Trp Ser Asp
320      325      330
Met Glu Asp Arg His Gly Gly Cys His Gly His Tyr Val His Gly
335      340      345
Ile Cys Ile Tyr Gly Asn Gly Asp Leu Lys Trp Leu Val Asn Ser
350      355      360
Pro Ser Leu Phe Ala Asn Lys Phe Glu Leu Asn Thr Tyr Pro Leu
365      370      375
Thr Val Glu Cys Leu Glu Leu Arg His Arg Glu Arg Thr Leu Asn
380      385      390
Gln Ser Glu Thr Ala Ile Gln Pro Ser Trp Tyr Phe
395      400

```

<210> 8

<211> 109

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500341CD1

<400> 8

```

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

```

WO 02/097060

PCT/US02/18354

<210> 9
 <211> 225
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7500846CD1

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

<210> 10
 <211> 463
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3230921CD1

<400> 10
 Met Ser Glu Glu Glu Ala Ala Gln Ile Pro Arg Ser Ser Val Trp
 1 5 10 15
 Glu Gln Asp Gln Gln Asn Val Val Gln Arg Val Val Ala Leu Pro
 20 25 30
 Leu Val Arg Ala Thr Cys Thr Ala Val Cys Asp Val Tyr Ser Ala
 35 40 45
 Ala Lys Asp Arg His Pro Leu Leu Gly Ser Ala Cys Arg Leu Ala
 50 55 60
 Glu Asn Cys Val Cys Gly Leu Thr Thr Arg Ala Leu Asp His Ala
 65 70 75
 Gln Pro Leu Leu Glu His Leu Gln Pro Gln Leu Ala Thr Met Asn

WO 02/097060

PCT/US02/18354

```

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

```

<210> 11

<211> 3150

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7247177CB1

<400> 11

aaataaatca ttgtagaaac gctagtttgg gcttgaaaaa ttccaggagc aagagtcaag 60

WO 02/097060

PCT/US02/18354

```

atgttgcaact ccatgagaat ctggaggagg ccccttccc agaaacttga cgatgaagta 120
ctgggttgtaa ttttagaaag acacocaaac ggctttttta aaagatoycc caggggccctt 180
gtctctgagag ctggggagctg gtccggagtg cagagaagcc atgggaagcag cagtgaaga 240
ggaaataaag gttagagatg aagctgtaga taaaacatt ttccaggact gtacacagat 300
cgcatlittac aggcgtcaga aacagtggct ttccaagaag tccacclato ggcatlitt 360
ggattcagtc acaacagatg aagacagcac caggttccaa atcatcaatg aagcaagtaa 420
ggttccclctc ctggctgaaa ttatcggtat agaaggaaac atttccaggc ttaaaattaa 480
cgaagagact cctctaaaaa ccagatttga agttccggat gtccctcaca gcaagccaag 540
cactglaagg ctgatttcat gctctgggga cacaggcagt ctgatttgg cagattgaaa 600
aggagacctg aagtgcctaa tcacagcaaa ccattcaag gtgactttgg tgtctgaaga 660
agaggtttgt attagcataa attcctctgg caattatata tttagacatc tacagattct 720
tcacaaacaa agagctgcta aagaaaaatg ggaggagaca tcagtggaca cctctcagga 780
aaatcaagaa gatctgggccc gtggggaaga gaaatttggg aaatttggg atatcaaaag 840
taatggccct tcttctattg gtttggattt cctcttgcat ggaatttgag atctttatgg 900
gatccacaaa catgcagaat cacaocaaact taaaaatact ggtgatggag atgottacog 960
tcttataaac ctggatgtct atggatacca aatatatgat aaaaatgggca tttatggttc 1020
aglaccltal cctctggccc acaaactggg cagaactala ggtatttctt ggtgaaatgc 1080
ctcggaacaa ctggtgaga tcaatacaga gctctcagta gactacacac tgacccagat 1140
ggggccagtt gctgctaaa aaaaagtcag atctcgact catgtgact ggagtcaaga 1200
gaatggcatt attgatgttt ttctgctgac aggcctaca cctctgatg tcttcaaaac 1260
gtactcacac ctacacaggc cacaagccat gccccctctt ttctctttg gataccacca 1320
gtccogctgg aactatgaag atgagcagga tgtaaaagca gtggatgcag ggtttgatga 1380
gcattgacatt ccttatgatg ccatgtggct ggacatagag cactactagg gcaagaggtg 1440
ctcacctgg gacaaaaaca gattcccaaa cccaagagg atgcaagagc tgctcaggag 1500
caaaaagcgt aagcttgttg tcatccagtga tcccacatc aagattgac ctgactactc 1560
agtatatgtg aaggccaaag atcagggctt ctttgtgaag aatcaggaag gggaaagctt 1620
tgaaggggtg tgttgccag gtctctctctc ttacttgat ttcaccaatc ccaaggtcag 1680
agaggtgtat caagctctt ttgctttccc ttttlatcag ggaatcaggg acatcccttt 1740
cctttggaat gacatgaatg agccttctgt ctttagaggg ccagagcaaa ccatgcagaa 1800
gaatgcatt catcatggca attgggagca caggagagct caacaaactc acggttttta 1860
tcatacaatg gctactgag aaggactgat aaaaagatct aaaggggaag agagacccct 1920
tgttttata cgtcttctt ttctcggtat acaaaagat ggtgcctgt ggaacggcga 1980
caacacagca gaatggagca acttgaaaat tctatocca atgttactca ctctcagcat 2040
tactgggata tctttttgca gactgacat aggcgggttc attggaatc cagagacaga 2100
gctgtagatg cgttggtacc aggtcgagc ctaccagccc ttcttcogtg gccatgccac 2160
catgaacacc aagcgacgag cgccttggtc ctttggggag gaacacaccc gactcatcag 2220
agaagccatc agagagcgct atggcctcct gccatttgg tattctctgt tctaccatgc 2280
acaagtggtc tccaaacctg tcatgagggc tctgtgggta gacttccctg atgaactaaa 2340
gacttttgat atgggaagat aatacctgct ggggagtgca ttattgggtc atccagtcac 2400
agaacaaaaa gccacacag ttgatgtgtt tottccagga tcaaatgagg tctggtatga 2460
ctataagaca ttgtctcatt gggaaaggag gtgtactgta aagatccag tagccttgga 2520
cactatccca gtgtttcagc gaggcggag tttgatacca ataaagacaa ctglagaaa 2580
atccacagc tggatgactg aatctctca tggactccgg gtgtgtcaa gcaataagg 2640
ttctcagtg ggtgagttat attttgatga tggccattca tcccaatacc tccacacaga 2700
gcaatttttg cacaggaagt tttcatctgt ttccagtgt ctgatcaata gttctgtga 2760
ccagaggggt cattatccca gcaagtgtgt ggtggagag atcttggtct taggctcag 2820
gaaggagcaa tcttctgtga ctaccactc atctgatggt aaagatcagc ctgtggcttt 2880
tactattgt gccaaaacat ccatcctgag cctggagag ctctcactca acattgccac 2940
tgactggggg ggttccctaaa aaagctctaa aatgaatgca tagtgtgtct ttcttcaag 3000
atgtaaatgt caaggaatga gttggcaaat cagtttaaaa acacaaacta cccaacagaa 3060
aaatcagaaa gttcccatc cagttctgtc ggaagagacc tccggaataa tgcgcttgcc 3120
agcacactgc ggcgtaacg tgagggggcc 3150

```

<210> 12

<211> 3290

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5596327CB1

<400> 12

WO 02/097060

PCT/US02/18354

```

ggctcagggt agcgagcggc ttgtctgctg ctgctgcttc tggggcgccg ctgtgcccg 60
cggcgccgcc cgggttgctct gctgctgctg ctgctggggg tctgtccgcc gggctgccc 120
aggcgccctg gccaccgagc actactcgcc gtctccctgc tcaagcagga gctgcagacc 180
ggcagagagg aggcgccggg gggggggggt gagcccgagt cggggggact ggggggacca 240
gtactctgcc gaggcgccgg agtcactcct ttggaacttc catgactcag actgcagacc 300
caagggatca tcaacctgtg actccttgct ttccctcaac actgagagga tcttgagcca 360
ggcgaagctc attgcagaac agagagagat cccgtttgcc actgataag acagcaaaaa 420
tgaagagtta gctattgctt atgtcttgat tggcagtggt ctgtatgatg aagcaatcac 480
gcatctttca acaatgcttc agcagagacc tgatctggtt agtgcacttt atggccgagg 540
gatagcctat ggaaegaagg cactacatga cattaaagt gctgagcttg cctgtgttga 600
actgagccga gtaattacct tggaaaccaga tcgtccagag gtatttgagc agcgagcaga 660
aattctgtcc cctctgggac gaattaatga agcagtgaaat gacctacta aagctatcca 720
actgcagccc tcagcacggc tgtacagaca tcgggggaacc ctgtacttca tatcagagga 780
catgcaaca gcccatgaag acttcoagca gtcccttagaa ctgaacaaaa accagcctat 840
agctatgcta tacaagggtt taactttcct tcacagagga cttctgaggg aagctattga 900
atccttcaaa gaagctttga agcagaaagt tgactttatt gatgcataa aaagtctagg 960
gcaggcatat agagaactgg gcaattttga agcagccact gagagcttcc aaaaggcact 1020
gtgtctcaac caaaatcatg tgcacaacct ccagctccgg ggaatgatgc tctaccacca 1080
cggcagctta caggagggcc tlaagaacct caaggtgtgt ctgcagctag agccataaaa 1140
tgaagtgtgc cagiatatga aagggtcag ccatgttgcc atgggaagct ttatggaagg 1200
gtataaagca caaaacaagg ttatgctgaa tgatcctctc ccaggccaga aggtcagccc 1260
tgatgatctt aaagtaaatg atctccgaga gtattctcga tatctctatg cacaccttga 1320
taccocctct accgaatata acattgatgt ggatctgcct ggaagcttta aggaccactg 1380
ggctaaaaat ttgcctttcc tcatagaaga ctacgaagag cagccagggt tgcaccccca 1440
cataaaagat gtgttacctc agaattttga gagttataag cctgaagtac aggagctgat 1500
tgtgtggctc gatcgttttg gatccctgat gcaatatgaa acacctggtt tctgccaata 1560
caagagaata cacagagcta tgggtttggc cgcattggag gtcatgcaag ccgtgcagcg 1620
tactatggac aactcgaaag ttcgaatgaa tgggaagaca cggttgatgc agtgagagga 1680
catgtttgac attgcagtta aatggagaa gattgctgac ccagaccagc ccgtgctgtg 1740
gttagatcaa atgccagcac gaagtcttag cagaggtttt aacaaccaca ttaatttaat 1800
cagggggcac gtgatcaaca tgaataacct agaatatttt gagaasatac ttaattttat 1860
taaaagcaga attctgtttt atcatggagc taataacct aaaggattgc tggaaagtgc 1920
ggaaagcctg gaaaaggtac caaaagtga agacctctt ccgattatga agtttaatac 1980
taaaacgaag gatgggttca cgtgaacac aaaagttccc agccttaagc accaaggga 2040
gaaalatgat ggattcacaa tcaagattac aggagacaaa gtggcaata tattattttc 2100
tgtggaactc caaacacggc aagaaggac acaattatct catgctgaaa tagatgcact 2160
ttataaagat ttgacagcaa aggaaaagt attgattctt tcatcagaat ttggggaggg 2220
tgatgctgtc tgaacttaa tottatcctt agtttatlac ttttataatt taatgccact 2280
ctctcgagga tccagttaa ttgcttactc ggtcatcggt ggagcactga tggcaagtg 2340
aaaagaagta gcaggaaaaa ttcccaagg gaagttagtc gactttgaag ctatgacagc 2400
cctgtgttca gaggccttta gcaagctgc caaaagctgg atgaacttga aaagtatttc 2460
accttcttat aagactcttc catcagtttc agaaaogttt ccaacgttaa gatcgatgat 2520
tgagggtgcta aacacagact ctctccacgg ttgtcttaag aaactctagt tctgtcgtg 2580
tatttataga agtaaggggc cggacctctt gctctttaa gtatttttta aaacatgga 2640
ttataaagaa attagttcct ctattacttt tgataccaat ttttatagga attgaatact 2700
tgaatacatt ttctttactt ttctgaacca taatgaanaa catgaaagag gctttcgggg 2760
ggaaaaagct acttgacttg gagggaaagc gtagttatta gcgcgagccg tgttggttcc 2820
cacttataaa gcatataaac tcttctctta cagtaaaaaa aattttcttc acaactaaa 2880
atatataact gaagatgaac aaaaaaataa ccaaatagga gtgtttgtat tcttttagag 2940
atctggatat ttttcaaga ctgctttttt acgcacattg tgatagagca tcttttggtg 3000
actcacgtaa cctccatgtg tgggtccttt gttttattag aacocatgga gaaggttctc 3060
gggccatcag tatgtgttgc agagctcaca gccagggaaca cttattttga acaatgcaga 3120
cagtaattgc cagagtcoca aatgctgaat tattgaacca tgtttttttt tctctttgtt 3180
gaaaaaataa aaccaggtga ttctataaaa tcaccagttt ttggaaaaaa aaaaaaaag 3240
ggggcgccgg aaaaaggagc tctggaaccc gggaataaat ccgggacccg 3290

```

<210> 13
 <211> 4279
 <212> DNA
 <213> Homo sapiens

 <220>
 <221> misc_feature

WO 02/097060

PCT/US02/18354

<223> Incyte ID No: 4179651CB1

<400> 13

```

ggaagagagg ctoctgaagg cggctggcgc ggtgtgtggg gccctgggtc tectgcccac 60
cgtgggggtt tgggcgtgtt acgcgagcg gcagcccgac ggcacccctg tgtgggatcg 120
ggggcgtgtg tglgtgtgtg cgtgctgtgt gtgacatggg ctacacatgt tgcgcaaaat 180
gaaaacgitt tcttctgtga gatgggcaga agctgaagtg actggcatga caatggaggc 240
catccgcatg ggcagctcag cggctaggaa atggagaaca aggaagacct taccocatga 300
ccgatgtcga gagagtggat caggcatacc gagaaaatgg atttaacatc tacgtcagtg 360
ataaaatctc cttgaatcgc tctctccccc atatccggca cccaaactgc aacagcaagc 420
gtacactcga gacactcccc aacacaagca tcatcatccc cttcccaaac gagggtcgtg 480
cctccctcct ccgcacccgc cacagtgtgc tcaatcgtcc gctccagagc ctggctgcgc 540
agattgtact ggtcgacgac ttcatgtatc gagagacct gaagagacct cttgaagact 600
acatggccct ttccccagt gtgaggatcc ttgaaaccaa gaaacgggaa gggctgataa 660
ggacccgaat gctgggggccc tcaagtgcaa ctggggatgt catcacatcc ttggattcac 720
actgtgaagc caatgtcaac tgggttcccc ccttgcctga ccgcatctgc cggaaactga 780
agacacttgt gtgcccgatg atctgatgaa ttgacccatg cgaactttcg tacgagacac 840
aggcagggga tgccalgcgg ggaacttttg actgggagat gtactacaag cggatccgga 900
tccctccaga actgcagaaa gctgacccca gcgacccatc tgaatctccc gtgatggccg 960
gtggactggt agccgtggat cgggaagtgt tctgggaact cggcgggtat gaccaggctc 1020
tgagatctgt gggagggggag cagtatgaaa tctccttcaa ggtgtggatg tgtggggccc 1080
gcattggagg catccccctg tccagggtgg gccatatcta caggagatgt gtgccctaca 1140
aggctccggc cggagtcagc ctggcccgga acctaaagcg ggtggccgaa gtgtggatgg 1200
atgagtacgc agagtacatt taccagcgcc ggootgaata ccgccaacct tccgctgggg 1260
atgtcgcagt ccagaaaaag ctccgcagct cctttaactg caagagtttc aagtgttcta 1320
tgacgaagat agcctgggac ctgcccacat tctaccaccc cgtggagccc ccggtcgacg 1380
cttgggggga gatccgaat gtgggcacag ggtgtgtgct agacacaaag cccggggcct 1440
tgggtcccc actaaggcta gagggtctgg tccagggcgg tgggggggct gccctggaac 1500
acatgcaggc atccaccttc agctggagag aggaactccg gcttgagac ccccgagaca 1560
ccagaaagt ctgctttgat gccatttccc aacccagccc tgtcagctg tacgactgcc 1620
acagactgaa gggcaaccag ctgttgaaat accgcaaaag caagaccctg taccacctg 1680
tcaagtggag ctgcattggc tgcattgaaa gtgacccatg gatcttcatg aacacctgca 1740
accatccctc tctaccccag cagtggctgt ttgaaacac caactcaaca gtcttggaaa 1800
aatccaatag gaactgagcc ctcatgtccc ctggccagcg cccccagggt ctgggacctc 1860
ctgcagactt cctctttcaa gggaggcagg gccctgtgg gcactagggt taaaaggtyc 1920
tgcccaaatg gttcagggtg aagagggtct ttgattcagg ggtgggggtc tgcctggctc 1980
ttgagccctc gatttgtggg ggtagggtga agagcatatc ccacaagagg cccacagggy 2040
agcagagact gctttaatcc ctgctgacat caoggaagag caacagagcc ttttcaactt 2100
tgtcaactat tccccctgaa cattatgtgg gagaaaccca aggtagoccta ggccaaccaa 2160
aagtgaatcc tggaggttg cccagccctc agatggctct cctacatgat ggtgtcttag 2220
aaacaaaggc aaaaattggc tgtttggggc agcttttagt atcgatgcca ctcatctgca 2280
gcagaaagga aagaagtccc ctggggggtt tttagttctt gccgtcctgg ggggaacatt 2340
gcagttaactg cacagcttct gttctctgtc acaacccagc gtgattttgt ccggtcaaa 2400
gccataactg gggccctaag agtgttcatg attgaatgct gatcagctgc cagggtgagga 2460
gtcagaaggg gggcccccct tagacatttc tttagcgcta tggacatgct ggtatctccc 2520
ccctgctctc tgggtatttg aaatgtcaat tttagacatc tccaggcaaa aggcagccc 2580
agcaccagct ttacagggca gtgtttcaga tggcccttag cccagggaaa aggcagagta 2640
gacctccaaa ctagnaatgc tggctgattt gccctgctcc atgcttccat ttccctgtct 2700
ctcttcccca ggcaattact ggcctcaaaa gaggaacaga ggtgctgcga ggtgctcacc 2760
tcaagagtc tggaggcctc caggatcaac tgtgggcaaa gtgctgctct ctgacctcat 2820
catggttcta gttctcatcc agaactccag aatttttaaa gaactctata attggattgc 2880
aaactaggat gctacatagg attctggtat tccacatcca atatggallt ctgagaatgt 2940
gtgatgaag gagccagcca ggtgtaatac agtcaaggca gcccccagcc tagagacaa 3000
ctgtgaatc caaagtggg ggtgttggga aagcaaggga acatgtctcc ctcaagctag 3060
cagaggctgt ggtacaacat ggtccttggt gaagacctgc acccctggaa cctccacaca 3120
tcataccaac tglagtctca ttgcagtg agaaaagac ccgacgtccc acagccagat 3180
atacccccag ctccatgcca gccctctcat tttacottttt gotttgttaa ttacatgca 3240
gactccataga ggcctctcag actaatagga agcatcttct taaccaacct gccaccact 3300
gattcagaaa tggaaatcac attccacaat ctatggcttc caccagctag cccaggaaat 3360
acttgaatc agcaattcaa ttatgttga gtctcttgat tgtgtcatct accaataaa 3420
taactgagac ctaagtctgg gaacagagcc acgaatctgc ctttgagatg ctggcagatc 3480
tcaagggcat caattattgg gggaggggag gacaaacatc cccaatcacc caccagtag 3540
actgaatgtg tagctggoga ggaattactt ccaactctgg cccagcaaaa gccctgcttt 3600

```

WO 02/097060

PCT/US02/18354

```

ggccacctgt ctgcaagaga ggccggccct gtgcttgcaa cgttacgtg ttgatccag 3560
tgtccctttc caaatgagtg ctgtagcttt agaagtggcc ctctatagaa agaagtcaaa 3720
agatgaggcc ccttctagaa tctaggataa caagagtgtt gacagtttga ggagtcgaat 3780
tggagttcat catcaagag caatgcagcg tcgttaaaat aaaaactgtg ccttttataa 3840
agaaaatatc aatatagag caatcccta aacttgacc atttctagt gcttgctag 3900
acaacataaa agggcgagg ttgtggggg gggaacgtt ttgttggtgt gtcgagttcc 3960
tactggatga gtgtgtgtt ctaaatgtct catccaagc agagatgtt gggtctggag 4020
caagatctga gactgagatg tcccccaag gtgacaggtc agatttattt cagtcagtgc 4080
aagtccacct ccaactaaa ggctgggag tgcgtttagt ctgtggccta gaagacctca 4140
tcagcccccg agaggaggcc ttgcattacc tcaactggag ctgtttgggg gccatccctg 4200
cggctccacc tctgaacccc cagcatgttt gtctgttttc tcacoattgg gtttaaggggt 4260
gcctgagcact agctagact 4279

```

```

<210> 14
<211> 1042
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7482109CB1

```

```

<400> 14
atgaattcta aaagaatgct gttattggtt ttatttgctt tttaactgat gttgggtgag 60
cgttatttca ggaatcacca agtagaagaa cttcggctct cagactgggt tcactctaga 120
aaacgccttg atgttataac gaaaacagac tggctcgctc ctgtcctatg ggaagggact 180
ttcgacaggg ggttcttgga aaaacattac agaagcgga atatcactgt gggcctggcc 240
gtcttttgta ctggcaggtt tgcagaggag taccctgagg cgttcttaca ctccgcaaat 300
aagcacttca tgaagaggta ccgagtgatc ttctacatca tggtaggagc cttcttcaag 360
ctgcctgaca tagagcccgag tctctcttga acgttcaaa cattlaaagt gggcaccgag 420
aggttggtgg tcatggccc cctgggtgat gtgaagagcc tgggtgaaca catcgccagt 480
cacatccagg acgaggtgga cttctcttct agcatggctg ccacccaggt cttccagaat 540
gagttcgggg tggagacccc ggcccccgtg gtgcccagc tccacgctg gtggtatttc 600
agaaacacca agaacttccc ttatgagagg aggcgcacct cagcagcttg catccogttt 660
ggcacaggag atttctatta tggcaacttg atggttggtg gcacacccca taatatltta 720
gacttcatca aagaatatct gaacggagtt attcatgaca tcaaaaatgg actcaatagc 780
acttatgaaa agcaccttaa caaatatttt tacttcaata aacccacctg gctgttatca 840
ccagcataca gctgggatct tgcattttct cctctccacc agatccaata cgtcaaggtc 900
gcacatgatt cccagaggaa attatgaatt acgcatagag gtttatgag aataaatgaa 960
tcacggcaat ttcttataaa tgaggggaag cttaagacct tccatttttt agagacatac 1020
aaatttccat cctcccaaaa aa 1042

```

```

<210> 15
<211> 4320
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2867618CB1

```

```

<400> 15
attcatttgg ggaataattt tgatttccaa agagtgcag aatcaacagt gaaacgaata 60
ccaagaattt aaaaagtgtt gaaatcacta tgaacacctc caaggcctct gaacttaact 120
tggatgaact accacctctg aacatttata tactgtatt ttattacagt tggatagaaa 180
atccacaatt tgatggttaa tatacattt ggatccatcc agtgttagag ccttgggacc 240
ctagaatagc caagaattat ccacaaggga gacacaaccc tccagatgac attgggtcca 300
gcttttatcc tgaattggga agttacagt ctcgggatcc ttctgtcata gaaactcaca 360
tgagacaaat gcgctcagct tcaattggtg tactagccct ctcttggtac caactgatg 420
taaatgatga aatggagaa octactgata aottgggtacc cactattttg gataaagctc 480
ataaatataa cctaaggtt acttttcaca tagaaccata tagcatcgag atgatcaaac 540
atgtacaaaa tgcctagtat attatagaca aatatggaaa tcatccggcc ttttacaggt 600
acaagacgaa gactggcaat gctcttccta tgttttatgt ctatgatccc tatgtgacca 660

```

WO 02/097060

PCT/US02/18354

```

agcctgaaaa aaggggccaat ctgttaacca cctcagggtc tcggagtagt cgcaattctc 720
cttatgggtgg actgttttatt gccctctctgg tagaagaaga acataagtat gatattcttc 780
aaagggttga tgggaatttac acatatcttg ccacaaatgg cttaacttat ggcctcatoc 840
atcagaatgg ggotagoccta aaattattgt gtgataaata caacttaata ttatcccaa 900
gtcggggccc aggatcacata gatcacagga tccgtccatg gacactcgca aacactcgga 960
acggaatcaa tgggaagtat tatgaaatg gtctgagtc cgcacttcag aacggcccca 1020
gcttaatttc tctcaactct tttaatgagt ggcattgagg aactcagatt gaaaaagctg 1080
ttcccaaaag aaccagtaab acagtgtacc tagattacgg tctctataaa ccaggtcttt 1140
acctagaac cgaactcgcaag tggctctgaaa aatacagtaa ggaagagca actatgcat 1200
tagatcgcca gctgctctgt tcttaatgca ttgattaaag cttaactgtt atcaaaatca 1260
cctaattttt aaaaaatgct ttgtttttga gtctgggaaa gaaaactgtc aaaaatcagta 1320
tatactatta gttatattta aaaaattttt tttaaattct ttacagataa tattataact 1380
gttacccttc acaataccac atgagaaaat atctgagaca aaatgtatac aaatatattc 1440
cttatggcat aatttattgc atttctgact gaaatcaaaa tcttgatttg atggcaattg 1500
aattttcatt ttacaataga taaatgcttg tgttaacctaa agcaacttagc acacagttaa 1560
attatattta oactctagac ccaataaat aggattgtgt gtatatttgg gatattctat 1620
gaagaaaaaa agaaaacccc ttaaagataa tgcactgct tcatgtcatg tctttaaaat 1680
aatttaacaa actttattgt cttagtattt agactctgga taactctaca ataatgagga 1740
aattcttaag aatacaaaa tctctgtacc tctctctcna ttttgctgtg aacctgaat 1800
ggctttaat taatactctt atttttatt taatttaatt acataaatta aacctacca 1860
tgaccaaat gtgttaggaa ggctgtctat ctacagcaca gtgtgtcatt tgcagatttg 1920
tggttaccta taccagcta ggtgttttga catgtttagt gttctgtctt tacagtgtg 1980
aattccatct tttagaagct atgaaagctc ttttatgaaa aagttactga ttgctctca 2040
gttattagga aaacagttgt ttcaaatata ttatgtatgt atgatgacca aatatcattt 2100
ttagtatact ttgtctatgt ttaagtttgt actattgtgt tattcatgct tttaaatcag 2160
ataccaaata ttttttagga aagaaaaatg ttattactgt cattaggttg tcttttaata 2220
ctttaagtta ttttgacgaa aagtaataga gaaaatttac ttgacttttt agattctaga 2280
gacatggaaa tgaaaattat tttatgtcta gagtaggtcc tgaagtttgg ctttacatta 2340
agtttagcac tgtatcagaa tgaggaactt aatattttac ataaaaacta atactttcaa 2400
tttttatat agtaatatcc ccattttgta aagtttagac ttttatcata cctgtaagt 2460
aaaatactgt ttataataa ctgtctcagt tgggaacaa ttgtcacttg tgtcggaaa 2520
ataaatctcc ttaaaaacta aataaagtgc actgtattct taagttaat gtttataact 2580
atagtaaaa attaatatct atctattac ataaatgtta tttcttaggt gttcattaa 2640
gaagagcaat agaataatgc taaaaataa tgcotataaa ttttcagagt ataaagacat 2700
ccattcagaa acaaaaatta gcactaaatt ttttataaaa tagaccagat gacaaaaatt 2760
attttatttt taaacagttg ttttgacaca aattatgtta ttgaaaagca ttattaatgt 2820
ttaatttatt taaaattttg gaatttgoca tttctcagac aatgatoagg ccttaggaaa 2880
ttaatacagt agtagtaact attttctagg ggaataaata agataaact actatactga 2940
tattttgata taagcaagca ctacactggt aatcaactata tagatccaac ctgtggattt 3000
tcttctatgt cccatttaac tagaatatat tattttaggt ataatttaca aatgtccac 3060
ctaataatct ttataatct accatatttc attaaagttt tgttagagaa gtatctacca 3120
cagaggagtt ttgtcatttg tgtacgttgt gtatttgac ccacatgac agaaagttaa 3180
tttagggaaa tagttatgag attaaggga aatctataaa aacaaagttg gcatattctc 3240
aacacagata ccacactttt ctttttcca ttatagacat ggtgaatcca cacagcatc 3300
ttcatctctg agcttttgtg tgattctctca acaacttacc ctacccagcc agcagtaaca 3360
gatctcagag taagataaag cagattctgt ctcaattgca aaaagttatt ctcaatggaa 3420
gaatggcatc tgaatctcat attactagtt tatattanta tagttttttt ctcccttttt 3480
aataaaataa ttacagtcat cctcagttgt ctgtggggga ttggttccag ttaccctat 3540
agataccaaa atctgcagat gctcaagctc ctgatataaa ctggcactag agttgcatat 3600
aatctatgca catctcctg tatacattaa gtcatctcta gattatttat aacacttaat 3660
acaatgtaaa tgcatagtag ttgttatacc atactggtta gtgaataatt acatgaaaa 3720
aaagagctct cacatcttca gagtttcagt cggcaatttc ttggccatgg atgtagaacc 3780
tacagataag gtgagccaac tgcatttaga aataactcta ataattctgt taattcttag 3840
agaggaaaaa ttcaaaaatc tctctcaggt attattaca actgcttta coatttttag 3900
tgtaacacag ttlaaattgt tatgataaca agtaaatag agcaagaaat ttatttctta 3960
attcaaaact atactgttg atcaaatat gtataacta aagtggtata atcaataca 4020
tgcagtgaac ataatggatt cttttataag ttattaattt ttatggttta atcagttca 4080
ttgttttgac tgtttalagaa accaaaatt ttaactgttc ttttaaggac taattttgtc 4140
aaaaactgct gttattaaat tcaactgagt tgtttaactt cttctgtgtt taagattgta 4200
attaataaatt actattttgt tatattgaaat ggttaatttt tacctataaa aaacatagat 4260
gaaatacact gtattttaga tattaagtc tatcatttgc ttttataaaa aacaaaaaaa 4320

```

<210> 16

17/20

WO 02/097060

PCT/US02/18354

<211> 1733
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7488348CB1

<400> 16
 atcggatctg agcgaacgga acgggtgctg gtgttcaaga tgtccattct caagatcacc 60
 catgcccgtg atatatctga atcccgctgg aatcccaactg ttgaggtaga cttttacact 120
 aacaggggtg gtctgttctg aagagctgct gtctctagcg gtgcctcaac tggaaatttat 180
 gaagctctcg tggaaactcg ggacaatgac aagacacgct acatgggggg gaagggtgtc 240
 tcaaaagctg ttgagcacat tatcaataaa acaattgcac cgcactgat tagcaagaat 300
 gtcaatctag tggagcaaga caagatcgat aacctgacgc tggacatgga tggatcagag 360
 aacaaatcta aatttgggtg caacgcacac ttgggtgtat ctctggctgt atgctcgaat 420
 gctgtgtcta ctgtcgagaa ggggtgtccc ttgtacgctc acattgtcga ccttctctgg 480
 aataacccag aagtcactct gcctgttccc gcttttaacg tgatcaacgg tgcctccat 540
 gctgtcaata agctggctat gcaggagttc atgatccctc cctgtgtgtc tgcagggttt 600
 aacgaacgca tccgcatttg tgcagaggtt taccacaacc tgaagaatgt catcaaggag 660
 aatatggga aagatgccac caatgtgggg gatgaagcgg ggtttgctcc caacatcctg 720
 gagaataaag aaggcctgga gctgtctgaag actgctattg ggaaagctgg ctacactgat 780
 aaggtggtca tgggcattga cgtagcggcc tccgagttct tcagggtctg gaagtatgac 840
 ctggacttca agtctccgga tgaccccagc aggtacatct cgcctgacca gctggctgac 900
 ctgtacaagg gcttctctct ggggcattgca gtgaagaact acccagtggg cgtctccatc 960
 gaggaacccc catltagcca ggaatgactgg ggtgcctgga agaagctgtt tactggcagc 1020
 ctggttggtc tccaagtggc tggagacgat ctgacctga ctaagctcga agcgcgtatt 1080
 gcaagaggtg tggagaggtg taagcctgac aactgctctc tctctctcaa ggtcaaccag 1140
 attgactctg tgacggagtc cctacaagct tgcgaagctg ccaagtccaa cggctcggggc 1200
 gtgatcccg tgagtcaccg cctctccgga gaacagaaag ataccttcat ggtgacctc 1260
 gtgtgtcggg tctgcactgg tcaagatcaa actggtccca ctgcccgaic tgaagctcta 1320
 gccaaagtaca accagctgct gaggattgaa gaggtcgagg ctggcagcaa ggcgcgttt 1380
 gctggaagaa acttcaggaa ccccgctatc aactaagctg cgtggatcga caccgcttt 1440
 ggttatgtaa gcaactagta cctacttaga ctcaaatata cttgtattag aagatgaggg 1500
 gcaggtctga gaaaagacca gtttgcaggt cctctccctc cctagatgac tctcctcac 1560
 ctcaatgtgt tccaccagct ctgactctgt acttgaacg atcgtgtctt gtagaacaa 1620
 cccagctctt gatgttttgg aggcgtgttg acttgacgga cagcaagaag gtacctacaa 1680
 aacagctagt agtgttttta catgtgataa ataaaaagca tcaaaaaaaa aaa 1733.

<210> 17
 <211> 2201
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5539166CB1

<400> 17
 gacagggtta agcaattttt ttaaaagcact acatatattat tgatagaata tctctccagg 60
 cagtattcac atgcttttaa gtgaagctcg tagcttgcca gtgcagttta agcaataaaa 120
 ctcaagctgt tgaacacaat gattaacagg aagcagggga ggggacatag ttactagact 180
 ctgcccggga caagaggtg aggttaaggt cttagagcgg atgacttgca gaatcgtcac 240
 ctaccccact tgcaatttca tgggtgggac aggtggaccc cctggaaatca gaacctctct 300
 gaggaacact gtttttgtgt agacacaggt tgcaggttag caggagaaca ggcaagccaa 360
 atgcaaaagg gccacttcag aatgtgtca cagaaaagtg aaatgcaac ctagtgttaa 420
 gtgaagaggg gaagaagaaa gaaaaaggac cagaacccgt aactgaaggg accgggaaca 480
 gccacagcag agcttcagcc atcacaggga tgatttcgga acctggagaa aatgtaagtt 540
 aatatatctc acactctgat cctatctcaa gagagagata ttttactcat ttcctggttg 600
 tgaatgatgg gctcttggaa gcaactgtct tttagcgtct ctttactctc tgcctgatt 660
 tttgtatttg tttaacaata tgaattatgg gagaataaac gttttctgag ggcagctctg 720
 tccaatgctc cactgttagc agaagcctgt catcagattt ttgaggggaa agttttttac 780
 ccaacagaaa atgcattgaa aactacccct gatgaagcta cctgctatga gtacatgggt 840

WO 02/097060

PCT/US02/18354

```
cgaagccact atgttaacaga aacactctct gaagaagagg ctgggttccc ttagcttac 900
acagtgacct tccacaaga ctctgggact tttagagggc tcttcggggc gatttatatg 960
ccccaaaatg tctactgtgt gaactcggat cagaaggcga cggatgcctt taaagtgca 1020
gtgaacagt tactcagctg ctcccaaat gctttctgg ctcccaaga ggagtcggtt 1080
gtctatgggg ggtctccag gctccaggct gaactgaact gcctggaaga ccttgtggcc 1140
tcctgaattc cctggaagta tctcatcaac acctcggggc aagactttcc cctgaaaacc 1200
aacggggaaa tagttcagta tctgaaggga tttaaaggga aaaaatcac ccccgagtg 1260
ctgcctcctg accacgtgt tggacggact aaatcgtcc accaagaact gttaaacac 1320
aaaaattcct acgtgattaa aacaacaaaa taaaaaact ctcctctca tgacatggtg 1380
attacattg gcacggccta cgtggctctc acaagggaact ttgtaacct cgtcctccaa 1440
gaccagctcg cacttgactt actctcctgg tcaaggaca cctacagcc cgacgaacat 1500
ttctgggtga cactcaacag gattcccggt gtctcgtgct ctatgccaaa tgcctcctgg 1560
actggaaaac tcagagctat aaagtggagt gacatggaag acagacacgg aggtgtccac 1620
ggccactatg tacatgggtat ttgtatctat ggaaacggag acttaaaagt gctgggtaat 1680
tcaccaagcc tgttlgctaa caagtltgag cttatcactt accccttac tgtggaatgc 1740
ctagaactga ggcactgga aagaacctc aatcagagt aaactgcgat acaacccagc 1800
tggtattttt gagctattca tgagctactc atgactgaag ggaactgca gctgggaaga 1860
ggagccctgt ttgtgagag acttttgctt tctaatgtt aacgtttca ggaccacgtt 1920
tatagcttca ggacctggct acgtaattat acttaaatat tccactggac actgtgaat 1980
acactaacag gatggctggg tagagcaatc tgggcacttt ggcaaatitt agtctgtgt 2040
ttctctgatg ctacactcta tattagttta ttgttaggat catgataatt caatgacctc 2100
agatctggg cgataactct ctccgggtt ttctacccg gatgtgttct ctgctgcac 2160
acatcgtcc tccgagggtc tttaacctca gccgggggcc c 2201
```

```
<210> 18
<211> 457
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> misc_feature
<223> Incyte ID No: 7500341CB1
```

```
<400> 18
ggttcgggtc cgggtggctg agcggcgggg gaaatggcca cggggacaga ccaggtggtg 60
ggactcggcc tegtgcgctg tagcctgac atcttcacct actacaccgc ctgggtgatt 120
ctcttgccat tcactgacag tcagcatgtc atccacaagt atttctgccc ccgagcctat 180
gctgtcgcca tcccaatggc tgcaggcttc ctgctgctcc tgtttgtggg cagctttcct 240
taccttgaca cagcccccaga ccccacaaag cctctggac ctggaaagcc tgggggaagg 300
ctgaacagcc ccaggacacg ccttggggct cagggcgacc accccgggcc gctgaccgac 360
tgacctctcc tcacggaggg ccagccccaa agcccccggg ctggcccggt tgggacagct 420
gcacacaaa cactgatggt gtgtttaaaa aaaaaaa 457
```

```
<210> 19
<211> 1513
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> misc_feature
<223> Incyte ID No: 7500846CB1
```

```
<400> 19
ggagctccga gtgtccacag gaaggggaact atcagctcct ggcatctgta aggatgctgt 60
ccatcctcag gacaatgacc agactctgct tctgttatt cttctctgtg gccaccagt 120
ggtgcagctg agcagcagcc tcttctcttg agatgctctc gagggaaatc gaaacctgtg 180
ccttctcctt ttcttccctg cctagaagct gcaagaaat caaggaaacg tgcctagtgt 240
caggtgatgg cctgtatttt ctccgcacca agaattggtt tgtctaccag acctctgtg 300
acatgacttc tgggggtggc ggctggaccc tggtygccag cgtgacagag aatgacatgc 360
gtgggaagt caggggtggg gatcgtggt ccagtcagca gggcaacaaa gcagactacc 420
cagaggggga tggcaactgg gccaaactca acactttgg atctgacag gggccacaga 480
gcgatgacta caagaacctg ggctactaag acatccaggg caaggacctg ggcactctgg 540
atgtgcccaa caagtccccc atgcagcatt cgccctgctg aggtaccgca 600
```

WO 02/097060

PCT/US02/18354

```

ccaacactgg ctctccacag agactgggac ataactctgt tggcatctac caggggaatt 660
tgttgcagga ttctgtccag tccgggtgtt taataacgag agggcagcca acgccctttg 720
tgttgggata aaagttaact gctgtaaacg tgagatcac tgcatcggtg gaggaggggt 780
cttccacag gcaaacccc gtcagtggtg gaattctcc gactttgact gggtggata 840
tggaaactcac gtttaagagca gctgcagtcg ggagataacg gaggccgctg tactcttgt 900
ctatagatga gacagagctc tgcggtgtca gggcgagac ccatcttcca accccggcta 960
tttggagagc gaaaaactgg aattcttaaca aggagagag gagactaaat caactcaatt 1020
tgcaaaaaaa aaaaaagggg gggccgctct aaagaacccc caaggggccc caaccttacc 1080
cttgcattga acttcaatcc tctccccccta tgtaaaaagt gagaaggggc cggggccaga 1140
acctctgggg tgcctccagc ccccggtggg tggcccgatt gaccggttct ggaagggtta 1200
accagtgggc ctttgtaaaa aggcctggag ttgcccagg tgtgctgatg ggttaggaaa 1260
agacagttag gaggatatac ttacagatgc ctgagagcg cctataaaac atccctacga 1320
agtgganaag ccgggtgtgt gaataaacac cctgcccagc tccagaaagt gcactctgta 1380
ccgctcttaa aattcgcatg gaccgatgg aacctcccc aaggggaggt tgttgttagt 1440
agggtcaaca gccccgggga atctgaagtc attcaccac tgatatccc actacotttg 1500
ttggaggtta ctc                                     1513

```

<210> 20
 <211> 2084
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3230921CB1

```

<400> 20
ggcgggacgg gcgggggtgt ggagactoga gctgggggtc gggggagaca gctgggtgtc 60
gaagccgctc gggcccaggc tgaacctgtt tgacagcaga tgtctgaaga agggccggtc 120
cagatcccaca gatccagtggt gtggggagcag gacccagaga acgtgtgtga cgtgtgtgtg 180
gctctgcccc tggtoagggc cactgtccac cgggtctgag atgtttacag tgacgccaa 240
gacagccacc cgtctgtggg ctccgctgct cgcctgtgct agaactgcgt gtgcggcctg 300
accaccctgt ccttggaaca cggccagccg ctgctcgagc acctgcagcc ccagctggcc 360
actatgaaca gcctgcctcg caggggcctg gacaagctgg aagagaagct tccctttctc 420
cagcaacctt cggagacggt ggtgacctca gccaaaggac tgggtggccag cagtgtcaag 480
ggtgtgtgtg acctggcccg gaggggccgg cgtggagcg tggagctgaa ggcctccgtg 540
agccatgctg tggatgttgt actggaaaaa tcagaggagc tggtygatca ctctctgcc 600
atgacgtggg aagagctgct ggcactggcg gctgaggtcg aaggccctga agtggtgtcg 660
gtggaggatc agaggagaca gaagggtatc tttgtggccc tgggtccctc gtcagcagcg 720
atccgccacc tggcctacga gcaactctgt gggaaactga ggcagagcaa accccgtgcc 780
caggacaccc tggccagctc gcaggagacg ctggagctga tagaccacat gcagtgtggg 840
gtgaacccca cggccccggc ccgctctggg aagtgacag agctgtgggg ggaatggggc 900
cagcgccctc cggagagccg ccgcccggag caggcagagc tggagagcgt ggtgtgtctc 960
cgacagctga cccagagagt gcaggccacg gttagagctc tggagtcag cgtccgggag 1020
ctgcccgcgg cggcccagga gaaggtggct gagggtggcg gcaagtgtga tgccctcgag 1080
accgcccttc ctgatgcctg ctgcttcagg gacgtgccag cggcccgctc ggcagagggc 1140
gggggtcgcg tggcccacgc gcacgcctgc gtggacagcg tgcgtggagt ggtgtgtcag 1200
ccgctgcccg tgccctggct ggtgggaccc ttgcgccca tccctgttga gcgacccgag 1260
ccctgcgccg acctggcgga cctggtggac gaggctcatg ggggcccga ccccgcctgg 1320
ggcaacctgg actggccggc ccagcagaga gctggggagg cagagccagc ggaaggaggt 1380
gggaatgggg atggggacag gatgggtgtt gccggggaca tctgcagaca ggaacccgag 1440
acccccagct gccgggtcaa gcaacacctg atgcccgagc tggactcttg acccatgggc 1500
cagtgagggc ggggagggaa gggcacctgc acaccccgat cctctgtgac ccttgtggc 1560
caacgttag ctggaggctt tggccttgac cctctcttga aatcaggccc aactccggat 1620
ctctgaacac ctttttggtt ttggaactct ccaatttttc ctgaaacaca tggacaaga 1680
ggcccggggg agcagggcct cgaacctatc tcaggccaac ttgagccaca agctgggttc 1740
ttcaactatg tctgtctccc tggctccatg aagcgaatcc aaatctttcc aagaggtgtg 1800
gcacagtggc tcacgctgtt aatcccagca ctttggaggt ctgaggcagg tggatcaatc 1860
gaagttagga gttccgggac atcctggcca acatgatgaa accctgtctc tacttagaaa 1920
gacagcaaaa aaaaaaaagg gggggggccc gatttaggga gctcgtggac ccggggatta 1980
aatccgggac cggcccggtg ggggtaccag ttccctaact gggagcgtat aaagctgggg 2040
aacatggcaa taagggttgt gtggaattgt tccgtcaatc caaa                                     2084

```

【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
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number
WO 2002/097060 A3

(51) International Patent Classification: **C12P 21/06**,
C12N 5/00, 15/00, 15/63, C07H 21/02, 21/04

(21) International Application Number:
PCT/US2002/018354

(22) International Filing Date: 22 May 2002 (22.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/293,768 25 May 2001 (25.05.2001) US
60/509,548 1 August 2001 (01.08.2001) US
60/514,400 23 August 2001 (23.08.2001) US
60/343,706 19 October 2001 (19.10.2001) US
60/337,999 7 December 2001 (07.12.2001) 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): SWARNAKAR,
Anita [CA/US]; 8 Locksley Avenue #5D, San Francisco,
CA 94122 (US). GORVAD, Ann, E. [US/US]; 369 Marie
Common, Livermore, CA 94550 (US). HAFALIA, April,
J., A. [US/US]; 2227 Calle de Primavera, Santa Clara,
CA 95054 (US). DUGGAN, Brendan, M. [AU/US]; 243
Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US).
EMERLING, Brooke, M. [US/US]; 1735 Woodland
Avenue # 71, Palo Alto, CA 94303 (US). ISON, Craig, H.
[US/US]; 1242 Weathersfield Way, San Jose, CA 95118
(US). NGUYEN, Donnell, B. [US/US]; 1403 Ridgewood
Drive, San Jose, CA 95118 (US). LEE, Ernestine, A.
[US/US]; 20523 Crow Creek Road, Castro Valley, CA
94552 (US). YUE, Henry [US/US]; 826 Lois Avenue,
Sunnyvale, CA 94087 (US). YUE, Huibin [US/US];
1170 South Stelling Road, Cupertino, CA 95014 (US).
FORSYTHE, Ian, J. [US/US]; 308 Roble Avenue,
Redwood City, CA 94061 (US). LI, Joana, X. [US/US];
1264 Geneva Avenue, San Francisco, CA 94112 (US).
THANGAVELU, Kavitha [IN/US]; 1950 Montecito

Avenue #23, Mountain View, CA 94043 (US). CHAWLA,
Narinder, K. [US/US]; 33 Union Square, #712, Union
City, CA 94587 (US). BURFORD, Neil [GB/US]; 105
Wildwood Circle, Durham, CT 06422 (US). MASON,
Patricia, M. [US/US]; 360 Clarke Lane, Morgan Hill, CA
95014 (US). LALA, Preeth, G. [US/US]; P.O. Box 5142,
Santa Clara, CA 95056 (US). LEE, Sally [US/US]; 3643
26th Street, San Francisco, CA 94110 (US). BECHA,
Shanya, D. [US/US]; 21062 Clary Drive # 117, Castro
Valley, CA 94546 (US). TANG, Y. Tom [US/US]; 4230
Ranwick Court, San Jose, CA 95118 (US).

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

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

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

Published:

— with international search report
— before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

(88) Date of publication of the international search report:
8 July 2004

For 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: CARBOHYDRATE-ASSOCIATED PROTEINS

(57) Abstract: The invention provides human carbohydrate-associated proteins (CHOP) and polynucleotides which identify and
encode CHOP. 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 CHOP.

WO 2002/097060 A3

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

		International application No. PCT/US02/18354
A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : C12P 21/06; C12N 5/00, 15/00, 15/63; C07H 21/02, 21/04 US CL : 435/69.1, 320.1, 325, 455; 536/23.1, 23.5 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 435/69.1, 320.1, 325, 455; 536/23.1, 23.5 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched none Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Please See Continuation Sheet		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,E	HIRSCHHORN R. et al. Computer assisted cloning of human neutral alpha-glucosidase C (GANC): a new paralog in the glycosyl hydrolase gene family 31. Proc Natl Acad Sci U S A. 15 October 2002, Vol 99, No 21, pages 13642-13646. see page 13642, accession no AF5445044 and attached sequence search report regarding SEQ ID NO:11.	1, 2, 4-5, 17-18, 56
A,E	OHARA O. et al. The nucleotide sequence of a long cDNA clone isolated from human spleen EMBL/GenBank/DBJ databases. 01 June 2002. Accession No Q8TET4. see attached sequence search report regarding SEQ ID NO:1.	1, 2, 4-5, 17-18, 56
A,E	INCYTE GENOMICS INC. Carbohydrate-associated protein (CHOP)-1. 16 April 2003. Accession No AAG79779. see attached sequence search report regarding SEQ ID NO:1.	1, 2, 4-5, 17-18, 56
A,E	INCYTE GENOMICS INC. Carbohydrate-associated protein (CHOP)-1. 16 April 2003. Accession No ABA006831. see attached sequence search report regarding SEQ ID NO:11.	1, 2, 4-5, 17-18, 56
☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents:		
"A" document defining the general state of the art which is not considered to be of particular relevance	"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	
"E" earlier application or patent published on or after the international filing date	"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	
"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 reasons (as specified)	"&" document, member of the same patent family	
"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		
Date of the actual completion of the international search 30 April 2004 (30.04.2004)	Date of mailing of the international search report 27 MAY 2004	
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230	Authorized officer Sumesh Kaushal Ph.D. Telephone No. 703-308-0196	

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/18354
Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)		
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1. ☐	Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely:	
2. ☐	Claim Nos.: 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:	
3. ☐	Claim Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).	
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows: Please See Continuation Sheet		
1. ☐	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.	
2. ☐	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.	
3. ☐	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:	
4. ☒	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1, 2, 4, 5 17-18 and 56	
Remark on Protest ☐	The additional search fees were accompanied by the applicant's protest.	
☐	No protest accompanied the payment of additional search fees.	

INTERNATIONAL SEARCH REPORT

PCT/US03/354

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions, which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group 1, claim(s) 1, 2, 4-5, 17-18, 56, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:1.
 Group 2, claim(s) 1, 2, 4-5, 17-18, 57, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:2.
 Group 3, claim(s) 1, 2, 4-5, 17-18, 58, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:3.
 Group 4, claim(s) 1, 2, 4-5, 17-18, 59, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:4.
 Group 5, claim(s) 1, 2, 4-5, 17-18, 60, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:5.
 Group 6, claim(s) 1, 2, 4-5, 17-18, 61, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:6.
 Group 7, claim(s) 1, 2, 4-5, 17-18, 62, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:7.
 Group 8, claim(s) 1, 2, 4-5, 17-18, 63, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:8.
 Group 9, claim(s) 1, 2, 4-5, 17-18, 64, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:9.
 Group 10, claim(s) 1, 2, 4-5, 17-18, 65, drawn to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:10.

Group 11, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 66, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:11.
 Group 12, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 67, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:12.
 Group 13, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 68, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:13.
 Group 14, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 69, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:14.
 Group 15, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 70, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:15.
 Group 16, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 71, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:16.
 Group 17, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 72, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:17.
 Group 18, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 73, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:18.
 Group 19, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 74, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:19.
 Group 20, claim(s) 3, 6-7, 9-10, 12-13, 14-16, 29, 46-55, 75, drawn to an isolated polynucleotides selected from the amino acid sequences of SEQ ID NO:20.

Group 21, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:11.
 Group 22, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:12.
 Group 23, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:13.
 Group 24, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:14.
 Group 25, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:15.
 Group 26, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:16.
 Group 27, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:17.
 Group 28, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:18.
 Group 29, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:19.
 Group 30, claim(s) 8, drawn to a transgenic organism comprising the recombinant peptide of SEQ ID NO:20.

Group 31, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:1.
 Group 32, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:2.

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

PCT/US6 354

Group 33, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:3.

Group 34, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:4.

Group 35, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:5.

Group 36, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:6.

Group 37, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:7.

Group 38, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:8.

Group 39, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:9.

Group 40, claim(s) 11, 31-45 drawn to isolate an antibody that binds to an isolated polypeptide selected from the amino acid sequences of SEQ ID NO:10.

Group 41, claim(s) 19, drawn to a method of treating disease associated with decreased expression of NAAP.

Group 42, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:1.

Group 43, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:2.

Group 44, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:3.

Group 45, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:4.

Group 46, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:5.

Group 47, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:6.

Group 48, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:7.

Group 49, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:8.

Group 50, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:9.

Group 51, claim(s) 20-27, drawn to a method of screening compounds that modulates the activity of a polypeptide, selected from the amino acid sequences of SEQ ID NO:10.

Group 52, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:11.

Group 53, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:12.

Group 54, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:13.

Group 55, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:14.

Group 56, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:15.

Group 57, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:16.

Group 58, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:17.

Group 59, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:18.

Group 60, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:19.

Group 61, claim(s) 28, drawn to a method of screening compound that alter the expression of a polynucleotides selected from SEQ ID NO:20.

INTERNATIONAL SEARCH REPORT

PCT/US03/354

The inventions listed as Groups 1-61 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions listed as Groups 1-61 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical features of Groups 1-10 are polypeptide molecules, wherein each polypeptide molecule (SEQ ID NO:1-10) has different structure and function.

The special technical features of Groups 11-20 are nucleic acid molecule, wherein each nucleic acid molecule (SEQ ID NO:11-20) has different structure and function.

The special technical feature of Groups 21-30 is a transgenic organism characterized by an expression of a polypeptide, wherein each polypeptide (SEQ ID NO:11-20) has different structure and function.

The special technical features of Groups 31-40 is an antibody, wherein each antibody is characterized by its binding to a polypeptide that has different structure and function.

The special technical feature of Group 41 is treatment of a disease that requires in-vivo administration of a therapeutic compound.

The special technical features of Group 42-51 are compounds that modulates the activity of a polypeptide (SEQ ID NO:1-10), wherein each polypeptide has different structure and function.

The special technical feature of Group 52-61 are compounds that modulates the expression of a polynucleotides (SEQ ID NO:11-20), wherein each polynucleotides has different structure and function. In addition the method of screening compounds that interacts with a polypeptide does not share any special technical feature with method of screening compounds that interacts with a nucleic acid sequence, since these methods do not share a common core structure.

This international searching authority considers that the international application does not comply with the requirements of unity of invention (Rules 13.1, 13.2, and 13.3) for the reasons indicated below:

According to the guidelines in Section 9(i)(a) of Annex B of the PCT Administrative instruction, the special technical feature as defined by PCT Rule 13.2 shall be considered to be met when all the alternatives of a Markush-group are of similar nature. For chemical alternatives, such as the claimed sequences) the Markush group shall be regarded as being of similar nature when

(A) all alternatives have a common property or activity and

(B)(1) a common structure is present, i.e., a significant structure is shared by all of the alternatives or

(B)(2) in cases where the common structure cannot be the unifying criteria, all alternatives belong to an art recognized class of compounds in the art to which the invention pertains.

The instant application claims multiple polynucleotide sequences, the proteins encoded by them, antibodies and transgenic animals. Each of these is considered to lack unity because:

These sequences do not meet the criteria of (A), common property or activity or (B)(2) art recognized class of compounds. Each sequence behave in a different way in the context of the claimed invention. Each member of the class cannot be substituted, one for the other, with the expectation that the same intended result would be achieved.

Further, the sequences do not meet the criteria of (B)(1), as they do not share, one with another, a common core structure. Accordingly, unity of invention between the nucleotide and protein sequences is lacking and each sequence claimed is considered to constitute a special technical feature.

Accordingly, the claims are not so linked by a special technical feature within the meaning of PCT rule 13.2 so as to form a single inventive concept

Continuation of B. FIELDS SEARCHED Item 3:

Databases searched: Sequence databases in GENCORE. Modline. Search terms: sequences of SEQ ID NO:1 and SEQ ID NO:11.

Form PCT/ISA/210 (second sheet) (July 1998)

フロントページの続き

(51)Int.Cl.⁷

F I

テーマコード(参考)

A 6 1 P 1/12
 A 6 1 P 1/16
 A 6 1 P 1/18
 A 6 1 P 3/04
 A 6 1 P 3/06
 A 6 1 P 3/08
 A 6 1 P 3/10
 A 6 1 P 3/12
 A 6 1 P 5/14
 A 6 1 P 5/38
 A 6 1 P 5/40
 A 6 1 P 7/00
 A 6 1 P 7/04
 A 6 1 P 7/06
 A 6 1 P 7/08
 A 6 1 P 9/06
 A 6 1 P 9/10
 A 6 1 P 9/12
 A 6 1 P 11/00
 A 6 1 P 13/00
 A 6 1 P 13/02
 A 6 1 P 13/08
 A 6 1 P 13/12
 A 6 1 P 15/00
 A 6 1 P 15/08
 A 6 1 P 15/10
 A 6 1 P 17/00
 A 6 1 P 17/02
 A 6 1 P 17/06
 A 6 1 P 19/02
 A 6 1 P 19/04
 A 6 1 P 19/06
 A 6 1 P 19/10
 A 6 1 P 21/00
 A 6 1 P 21/04
 A 6 1 P 25/00
 A 6 1 P 25/02
 A 6 1 P 25/08
 A 6 1 P 25/14
 A 6 1 P 25/16
 A 6 1 P 25/18
 A 6 1 P 25/22
 A 6 1 P 25/24
 A 6 1 P 25/28
 A 6 1 P 27/02
 A 6 1 P 27/12
 A 6 1 P 27/16

A 6 1 P 1/16
 A 6 1 P 1/18
 A 6 1 P 3/04
 A 6 1 P 3/06
 A 6 1 P 3/08
 A 6 1 P 3/10
 A 6 1 P 3/12
 A 6 1 P 5/14
 A 6 1 P 5/38
 A 6 1 P 5/40
 A 6 1 P 7/00
 A 6 1 P 7/04
 A 6 1 P 7/06
 A 6 1 P 7/08
 A 6 1 P 9/06
 A 6 1 P 9/10
 A 6 1 P 9/12
 A 6 1 P 11/00
 A 6 1 P 13/00
 A 6 1 P 13/02
 A 6 1 P 13/08
 A 6 1 P 13/12
 A 6 1 P 15/00
 A 6 1 P 15/08
 A 6 1 P 15/10
 A 6 1 P 17/00
 A 6 1 P 17/02
 A 6 1 P 17/06
 A 6 1 P 19/02
 A 6 1 P 19/04
 A 6 1 P 19/06
 A 6 1 P 19/10
 A 6 1 P 21/00
 A 6 1 P 21/04
 A 6 1 P 25/00
 A 6 1 P 25/02
 A 6 1 P 25/08
 A 6 1 P 25/14
 A 6 1 P 25/16
 A 6 1 P 25/18
 A 6 1 P 25/22
 A 6 1 P 25/24
 A 6 1 P 25/28
 A 6 1 P 27/02
 A 6 1 P 27/12
 A 6 1 P 27/16
 A 6 1 P 29/00

4 B 0 6 5
 4 C 0 8 4
 4 C 0 8 5
 4 H 0 4 5

A 6 1 P 29/00	A 6 1 P 29/00	1 0 1
A 6 1 P 31/04	A 6 1 P 31/04	
A 6 1 P 31/10	A 6 1 P 31/10	
A 6 1 P 31/12	A 6 1 P 31/12	
A 6 1 P 33/02	A 6 1 P 33/02	
A 6 1 P 33/10	A 6 1 P 33/10	
A 6 1 P 35/00	A 6 1 P 35/00	
A 6 1 P 35/02	A 6 1 P 35/02	
A 6 1 P 37/02	A 6 1 P 37/02	
A 6 1 P 37/06	A 6 1 P 37/06	
A 6 1 P 37/08	A 6 1 P 37/08	
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 M 1/00	C 1 2 M 1/00	A
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 N 5/10	C 1 2 P 21/02	C
C 1 2 P 21/02	C 1 2 P 21/08	
C 1 2 P 21/08	C 1 2 Q 1/68	A
C 1 2 Q 1/68	G 0 1 N 33/15	Z
G 0 1 N 33/15	G 0 1 N 33/50	Z
G 0 1 N 33/50	G 0 1 N 33/53	D
G 0 1 N 33/53	G 0 1 N 33/53	M
G 0 1 N 33/566	G 0 1 N 33/566	
	C 1 2 N 15/00	F
	C 1 2 N 5/00	A
	A 6 1 K 37/02	

(31)優先権主張番号 60/343,706

(32)優先日 平成13年10月19日(2001.10.19)

(33)優先権主張国 米国(US)

(31)優先権主張番号 60/337,999

(32)優先日 平成13年12月7日(2001.12.7)

(33)優先権主張国 米国(US)

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

(72)発明者 ハファリア、エープリル・ジェイ・エイ

アメリカ合衆国カリフォルニア州 9 5 0 5 4 ・ サンタクララ ・ コーレデブリマベール 2 2 2 7

(72)発明者 ダガン、ブレンダン・エム

アメリカ合衆国カリフォルニア州 9 4 0 8 6 ・ サニーベイル ・ # 3 0 6 ・ プエナビスタアベニュー 2 4 3

(72)発明者 エマーリング、ブルック・エム

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 ・ # 7 1 ・ ウッドランドアベニュー 1

7 3 5

- (72)発明者 アイソン、クレイグ・エイチ
アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・サンノゼ・ウェザーズフィールドウェイ 1 2 4 2
- (72)発明者 ニュエン、ダニエル・ビー
アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・サンノゼ・リッジウッドドライブ 1 4 0 3
- (72)発明者 リー、アーンステーン・エイ
アメリカ合衆国カリフォルニア州 9 4 5 5 2 ・カストロバレー・クロウクリークロード 2 0 5 2
3
- (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 ユエ、ヒュイピン
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・サウスステリングロード 1 1
7 0
- (72)発明者 フォーサイス、イアン・ジェイ
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・ローブルアベニュー 3 0 8
- (72)発明者 リー、ジョアナ・エックス
アメリカ合衆国カリフォルニア州 9 4 1 1 2 ・サンフランシスコ・ジュネーブアベニュー 1 2 6
4
- (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 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 バーフォード、ニール
アメリカ合衆国コネチカット州 0 6 4 2 2 ・ダラム・ワイルドウッドサークル 1 0 5
- (72)発明者 メイソン、パトリシア・エム
アメリカ合衆国カリフォルニア州 9 5 0 1 4 ・モーガンヒル・クラークレーン 3 6 0
- (72)発明者 ラル、ブリーティ・ジー
アメリカ合衆国カリフォルニア州 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 リー、サリー
アメリカ合衆国カリフォルニア州 9 4 1 1 0 ・サンフランシスコ・トゥエンティシックスストリ
ート 3 6 4 3
- (72)発明者 ベチャ、シャニア・ディー
アメリカ合衆国カリフォルニア州 9 4 5 4 6 ・カストロバレー・# 1 1 7 ・ゲイリードライブ 2
1 0 6 2
- (72)発明者 タング、ワイ・トム
アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・サンノゼ・ランウィックコート 4 2 3 0

F ターム(参考) 2G045 AA35 AA40 BA11 BB50 DA13 DA36 FB02 FB03

4B024 AA01 AA11 BA80 CA04 EA04 GA14 HA14

4B029 AA07 BB20 CC03 FA15

4B063 QA01 QA05 QA11 QQ42 QR32 QR48 QR55 QS25 QS34 QX02

4B064 AG31 CA10 CA19 CC24 CE12 DA01 DA13

4B065 AA90X AA90Y AA91X AA93X AA95X AB01 BA02 BA03 CA24 CA25

CA27 CA44 CA46

4C084 AA02 AA03 AA07 AA17 BA01 BA08 BA22 BA23 CA18 MA13

MA52 MA56 MA57 MA58 MA60 MA66 NA14 ZA022 ZA052 ZA062

ZA122 ZA152 ZA162 ZA182 ZA202 ZA222 ZA332 ZA342 ZA382 ZA392

ZA402 ZA422 ZA452 ZA512 ZA522 ZA532 ZA552 ZA592 ZA662 ZA682

ZA702 ZA732 ZA752 ZA812 ZA892 ZA942 ZA962 ZA972 ZB052 ZB082

ZB092 ZB112 ZB132 ZB152 ZB262 ZB272 ZB332 ZB352 ZB382 ZB392
ZC062 ZC082 ZC212 ZC312 ZC332 ZC352 ZC552
4C085 KA03 KA04 KA05 LL01 LL09 LL11 LL12 LL13 LL15 LL17
LL18
4H045 AA10 AA11 AA20 AA30 BA10 BA53 CA40 DA86 DA89 EA20
EA50 FA74 GA26

专利名称(译)	碳水化合物结合蛋白		
公开(公告)号	JP2005501526A	公开(公告)日	2005-01-20
申请号	JP2003500229	申请日	2002-05-22
[标]申请(专利权)人(译)	洞察Genomics公司		
申请(专利权)人(译)	洞察基因组公司		
[标]发明人	スウォーナカールアニータ ゴーバッドアンイー ハファリアエープリルジェイエイ ダガンプレンダンエム エマーリングブルックエム アイソクレイグエイチ ニュエンダニエルビー リーアーンステイーンエイ ユエヘンリー ユエヒュイビン フォーサイスイアンジェイ リージョアナエックス サンガベルカピサ チョーラナリンダーケイ バーフォードニール メイソンパトリシアエム ラルプリーティジー リーサリー ベチャシャニアディー タングワイトム		
发明人	スウォーナカール、アニータ ゴーバッド、アン・イー ハファリア、エープリル・ジェイ・エイ ダガン、ブレンダン・エム エマーリング、ブルック・エム アイソン、クレイグ・エイチ ニュエン、ダニエル・ビー リー、アーンステイーン・エイ ユエ、ヘンリー ユエ、ヒュイビン フォーサイス、イアン・ジェイ リー、ジョアナ・エックス サンガベル、カピサ チョーラ、ナリンダー・ケイ バーフォード、ニール メイソン、パトリシア・エム ラル、プリーティ・ジー リー、サリー ベチャ、シャニア・ディー タング、ワイトム		
IPC分类号	G01N33/50 A61K38/00 A61K45/00 A61K49/00 A61P1/04 A61P1/12 A61P1/16 A61P1/18 A61P3/04 A61P3/06 A61P3/08 A61P3/10 A61P3/12 A61P5/14 A61P5/38 A61P5/40 A61P7/00 A61P7/04 A61P7/ 06 A61P7/08 A61P9/06 A61P9/10 A61P9/12 A61P11/00 A61P13/00 A61P13/02 A61P13/08 A61P13		

	/12 A61P15/00 A61P15/08 A61P15/10 A61P17/00 A61P17/02 A61P17/06 A61P19/02 A61P19/04 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25 /16 A61P25/18 A61P25/22 A61P25/24 A61P25/28 A61P27/02 A61P27/12 A61P27/16 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P33/02 A61P33/10 A61P35/00 A61P35/02 A61P37/02 A61P37 /06 A61P37/08 C07K1/22 C07K14/47 C07K16/18 C12M1/00 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12P21/02 C12P21/08 C12Q1/68 G01N33/15 G01N33/53 G01N33/566
CPC分类号	A61K38/00 A61P1/04 A61P1/12 A61P1/16 A61P1/18 A61P3/04 A61P3/12 A61P11/00 A61P13/00 A61P13/02 A61P13/08 A61P13/12 A61P15/00 A61P15/08 A61P15/10 A61P17/00 A61P17/02 A61P17 /06 A61P19/02 A61P19/04 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25/16 A61P25/18 A61P25/22 A61P25/24 A61P25/28 A61P27/02 A61P27 /12 A61P27/16 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P33/02 A61P33/10 A61P35/00 A61P35/02 C07K14/47
FI分类号	C12N15/00.ZNA.A A61K45/00 A61K49/00.A A61P1/04 A61P1/12 A61P1/16 A61P1/18 A61P3/04 A61P3/06 A61P3/08 A61P3/10 A61P3/12 A61P5/14 A61P5/38 A61P5/40 A61P7/00 A61P7/04 A61P7 /06 A61P7/08 A61P9/06 A61P9/10 A61P9/12 A61P11/00 A61P13/00 A61P13/02 A61P13/08 A61P13 /12 A61P15/00 A61P15/08 A61P15/10 A61P17/00 A61P17/02 A61P17/06 A61P19/02 A61P19/04 A61P19/06 A61P19/10 A61P21/00 A61P21/04 A61P25/00 A61P25/02 A61P25/08 A61P25/14 A61P25 /16 A61P25/18 A61P25/22 A61P25/24 A61P25/28 A61P27/02 A61P27/12 A61P27/16 A61P29/00 A61P29/00.101 A61P31/04 A61P31/10 A61P31/12 A61P33/02 A61P33/10 A61P35/00 A61P35/02 A61P37/02 A61P37/06 A61P37/08 C07K1/22 C07K14/47 C07K16/18 C12M1/00.A C12N1/15 C12N1 /19 C12N1/21 C12P21/02.C C12P21/08 C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 C12N15/00.F C12N5/00.A A61K37/02
F-TERM分类号	2G045/AA35 2G045/AA40 2G045/BA11 2G045/BB50 2G045/DA13 2G045/DA36 2G045/FB02 2G045 /FB03 4B024/AA01 4B024/AA11 4B024/BA80 4B024/CA04 4B024/EA04 4B024/GA14 4B024/HA14 4B029/AA07 4B029/BB20 4B029/CC03 4B029/FA15 4B063/QA01 4B063/QA05 4B063/QA11 4B063 /QQ42 4B063/QR32 4B063/QR48 4B063/QR55 4B063/QS25 4B063/QS34 4B063/QX02 4B064/AG31 4B064/CA10 4B064/CA19 4B064/CC24 4B064/CE12 4B064/DA01 4B064/DA13 4B065/AA90X 4B065 /AA90Y 4B065/AA91X 4B065/AA93X 4B065/AA95X 4B065/AB01 4B065/BA02 4B065/BA03 4B065 /CA24 4B065/CA25 4B065/CA27 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA03 4C084/AA07 4C084/AA17 4C084/BA01 4C084/BA08 4C084/BA22 4C084/BA23 4C084/CA18 4C084/MA13 4C084 /MA52 4C084/MA56 4C084/MA57 4C084/MA58 4C084/MA60 4C084/MA66 4C084/NA14 4C084/ZA022 4C084/ZA052 4C084/ZA062 4C084/ZA122 4C084/ZA152 4C084/ZA162 4C084/ZA182 4C084/ZA202 4C084/ZA222 4C084/ZA332 4C084/ZA342 4C084/ZA382 4C084/ZA392 4C084/ZA402 4C084/ZA422 4C084/ZA452 4C084/ZA512 4C084/ZA522 4C084/ZA532 4C084/ZA552 4C084/ZA592 4C084/ZA662 4C084/ZA682 4C084/ZA702 4C084/ZA732 4C084/ZA752 4C084/ZA812 4C084/ZA892 4C084/ZA942 4C084/ZA962 4C084/ZA972 4C084/ZB052 4C084/ZB082 4C084/ZB092 4C084/ZB112 4C084/ZB132 4C084/ZB152 4C084/ZB262 4C084/ZB272 4C084/ZB332 4C084/ZB352 4C084/ZB382 4C084/ZB392 4C084/ZC062 4C084/ZC082 4C084/ZC212 4C084/ZC312 4C084/ZC332 4C084/ZC352 4C084/ZC552 4C085/KA03 4C085/KA04 4C085/KA05 4C085/LL01 4C085/LL09 4C085/LL11 4C085/LL12 4C085 /LL13 4C085/LL15 4C085/LL17 4C085/LL18 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/BA53 4H045/CA40 4H045/DA86 4H045/DA89 4H045/EA20 4H045/EA50 4H045 /FA74 4H045/GA26
优先权	60/293768 2001-05-25 US 60/309548 2001-08-01 US 60/314400 2001-08-23 US 60/343706 2001-10-19 US 60/337999 2001-12-07 US
其他公开文献	JP2005501526A5
外部链接	Espacenet
摘要(译)	

本发明提供了鉴定和编码CHOP的人碳水化合物相关蛋白（CHOP）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CHOP异常表达有关的疾病的方法。

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	ポリヌクレオチド SEQ ID NO	Incyte ポリヌクレオチド ID
7247177	1	7247177CD1	11	7247177CB1
5596327	2	5596327CD1	12	5596327CB1
4179651	3	4179651CD1	13	4179651CB1
7482109	4	7482109CD1	14	7482109CB1
2867618	5	2867618CD1	15	2867618CB1
7488348	6	7488348CD1	16	7488348CB1
5539166	7	5539166CD1	17	5539166CB1
7500341	8	7500341CD1	18	7500341CB1
7500846	9	7500846CD1	19	7500846CB1
3230921	10	3230921CD1	20	3230921CB1