

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-504008

(P2004-504008A)

(43) 公表日 平成16年2月12日(2004.2.12)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A
A 6 1 K 38/43	A 6 1 K 45/00	2 G O 4 5
A 6 1 K 45/00	A 6 1 P 25/00	4 B O 2 4
A 6 1 P 25/00	A 6 1 P 29/00	4 B O 5 0
A 6 1 P 29/00	A 6 1 P 35/00	4 B O 6 3
		4 B O 6 5
	審査請求 未請求 予備審査請求 有	(全 265 頁) 最終頁に続く

(21) 出願番号	特願2001-563585 (P2001-563585)	(71) 出願人	301005050
(86) (22) 出願日	平成13年3月1日 (2001.3.1)		インサイト・ゲノミックス・インコーポレイテッド
(85) 翻訳文提出日	平成14年9月2日 (2002.9.2)		アメリカ合衆国カリフォルニア州94304・パロアルト・ポータードライブ 3160
(86) 国際出願番号	PCT/US2001/006806	(74) 代理人	100089266
(87) 国際公開番号	W02001/064896		弁理士 大島 陽一
(87) 国際公開日	平成13年9月7日 (2001.9.7)	(72) 発明者	タング、ワイ・トム
(31) 優先権主張番号	60/186,307		アメリカ合衆国カリフォルニア州95118・サンノゼ・ランウィックコート 4230
(32) 優先日	平成12年3月1日 (2000.3.1)	(72) 発明者	リュ、デュング・アイナ・エム
(33) 優先権主張国	米国 (US)		アメリカ合衆国カリフォルニア州95123・サンノゼ・コイドライブ 233
(31) 優先権主張番号	60/192,532		最終頁に続く
(32) 優先日	平成12年3月28日 (2000.3.28)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	60/193,578		
(32) 優先日	平成12年3月30日 (2000.3.30)		
(33) 優先権主張国	米国 (US)		

(54) 【発明の名称】 ヒト酵素分子

(57) 【要約】

本発明は、ヒト酵素分子 (HEM) と、HEM を同定及びコードするポリヌクレオチドとを提供する。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提供する。更に、本発明は、HEM の異常発現に関連する疾患の診断・治療・予防方法を提供する。

【特許請求の範囲】

【請求項 1】

単離されたポリペプチドであって、

(a) SEQ ID NO: 1 乃至 SEQ ID NO: 26 (SEQ ID NO: 1 - 26) からなる群から選択されたアミノ酸配列と、

(b) SEQ ID NO: 1 - 26 からなる群から選択されたアミノ酸配列と少なくとも 90% の配列同一性を有する天然のアミノ酸配列と、

(c) SEQ ID NO: 1 - 26 からなる群から選択されたアミノ酸配列の生物学的に活性な断片と、

(d) SEQ ID NO: 1 - 26 からなる群から選択されたアミノ酸配列の免疫原性断片とで構成される群から選択されたアミノ酸配列を含むことを特徴とする単離されたポリペプチド。 10

【請求項 2】

SEQ ID NO: 1 - 26 からなる群から選択された請求項 1 の単離されたポリペプチド。

【請求項 3】

請求項 1 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 4】

請求項 2 のポリペプチドをコードする単離されたポリヌクレオチド。

【請求項 5】

SEQ ID NO: 27 - 52 からなる群から選択された請求項 4 の単離されたポリヌクレオチド。 20

【請求項 6】

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

【請求項 7】

請求項 6 の組換えポリヌクレオチドで形質転換された細胞。

【請求項 8】

請求項 6 の組換えポリヌクレオチドを含む遺伝子組換え生物。

【請求項 9】

請求項 1 のポリペプチドを生産する方法であって、 30

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

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

【請求項 10】

請求項 1 のポリペプチドに特異的に結合する単離された抗体。

【請求項 11】

単離されたポリヌクレオチドであって、 40

(a) SEQ ID NO: 27 - 52 からなる群から選択されたポリヌクレオチド配列と、

(b) SEQ ID NO: 27 - 52 からなる群から選択されたポリヌクレオチド配列と少なくとも 90% の配列同一性を有する天然のポリヌクレオチド配列と、

(c) 前記 (a) に相補的なポリヌクレオチド配列と、

(d) 前記 (b) に相補的なポリヌクレオチド配列と、

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

【請求項 12】

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

れたポリヌクレオチド。

【請求項 13】

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

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

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

10

【請求項 14】

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

【請求項 15】

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

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

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

20

【請求項 16】

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

【請求項 17】

前記ポリペプチドが、SEQ ID NO: 1 - 26 からなる群から選択されたアミノ酸配列を含むことを特徴とする請求項 16 の組成物。

【請求項 18】

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

30

【請求項 19】

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

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

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

【請求項 20】

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

40

【請求項 21】

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

【請求項 22】

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

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

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

50

【請求項 2 3】

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

【請求項 2 4】

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

【請求項 2 5】

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

10

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

【請求項 2 6】

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

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

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

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

20

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

【請求項 2 7】

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

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

(b) 前記標的ポリヌクレオチドの発現の変化を検出するステップと、

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

30

【請求項 2 8】

試験化合物の毒性を評価する方法であって、

(a) 核酸を含む生体サンプルを前記試験化合物で処理するステップと、

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

40

(c) ハイブリダイゼーション複合体の収量を定量するステップと、

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

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

【発明の詳細な説明】

【0 0 0 1】

(技術分野)

50

本発明は、ヒト酵素分子の核酸配列およびアミノ酸配列に関する。本発明はまた、自己免疫/炎症性の疾患、遺伝性疾患、神経疾患、及び癌を含む細胞増殖異常の診断・治療・予防におけるこれらの配列の利用に関する。本発明はさらに、ヒト酵素分子の核酸配列及びアミノ酸配列の発現における外来性化合物の影響についての評価にも関する。

【0002】

(発明の背景)

生合成及び生分解の細胞プロセスには、酸化還元酵素、トランスフェラーゼ、加水分解酵素、リアーゼ、イソメラーゼ、リガーゼ、及び補助因子生合成酵素を含む幾つかの主な酵素群が関与する。これらの酵素群はそれぞれ、よく制御された精密な機能を有する多数の基質特異性酵素からなる。これらの酵素は、解糖、クエン酸回路、及び脂肪酸代謝などの代謝プロセスや、アミノ酸、ステロイド、リン脂質、及びアルコール等の生合成及び生分解や、シグナル伝達、細胞増殖、炎症、及びアポトーシス等の調節を促進し、DNAの複製や修復、及び翻訳プロセスにおける重要な触媒ステップにおいて機能する。

10

【0003】

酸化還元酵素

生合成および生分解の多くの経路は、供与体または受容体補助因子の酸化または還元を伴う酸化還元酵素(脱水素酵素または還元酵素)の活性を必要とする。潜在的な補助因子には、チトクローム、酸素、ジスルフィド、鉄硫黄タンパク質、フラビンアデニンヌクレオチド(FAD)、およびニコチンアミドアデニンジヌクレオチドNADおよびNADPがある(Newsholme, E. A. および Leech, A. R. (1983) Biochemistry for the Medical Sciences, John Wiley and Sons, Chichester, U. K. pp. 779 - 793)。還元酵素活性が、補助因子の酸化と同時に起こる基質と補助因子との間の電子の移動を触媒する。脱水素酵素の逆反応は、補助因子の還元およびそれに続く基質の酸化を触媒する。酸化還元酵素は、タンパク質の幅広いスーパーファミリーであって、細菌から植物およびヒトに至る生物の全ての細胞における様々な反応を触媒する。これらの反応には、糖の代謝、肝臓における或る種の解毒反応や、脂肪酸、アミノ酸、糖質コルチコイド、エストロゲン、アンドロゲン、およびロスタグランジンの生合成および生分解が含まれる。異なったファミリーメンバーは、反応が通常に触媒される方向に従って命名されるため、酸化還元酵素、酸化酵素、還元酵素、または脱水素酵素と呼ばれ得る。更に、ファミリーメンバーは、細胞質、細胞膜、ミトコンドリア内膜および外膜、およびペルオキシソームを含むその酵素固有の細胞内部位に局在化する場合が多い。

20

30

【0004】

短鎖アルコール脱水素酵素(SCAD)は、わずか15%~30%の配列同一性を共有する脱水素酵素のファミリーであって、その類似性が主に補酵素結合ドメインおよび基質結合部位に存在する。よく知られているエタノール分解の役割に加えて、SCADは脂肪酸、ステロイド、および或る種のプロスタグランジンの生合成および生分解にも関与することから、脂質貯蔵病、筋障害、SCAD欠乏症、および或る種の遺伝病などの様々な疾患に関与する。例えば、レチノール脱水素酵素は、レチノールをレチノイン酸の前駆体であるレチナールに変換するSCADファミリーのメンバーである(Simon, A. ら (1995) J. Biol. Chem. 270 : 1107 - 1112)。分化およびアポトーシスのレギュレーターであるレチノイン酸は、細胞増殖および炎症に関与する遺伝子をダウンレギュレートすることが分かっている(Chai, X. ら (1995) J. Biol. Chem. 270 : 3900 - 3904)。更に、レチノール脱水素酵素は、小児期発生常染色体劣性重症網膜ジストロフィーなどの先天性の眼の疾患に関与する(Simon, A. ら (1996) Genomics 36 : 424 - 430)。

40

【0005】

神経インパルスの伝達、細胞増殖および分化の調節、免疫反応の誘導、および組織恒常性には、神経伝達物質の代謝が伴う(Weiss, B. (1991) Neurotox

50

icology 12 : 379 - 386 ; Collins, S. M. ら (1992) Ann. N. Y. Acad. Sci. 664 : 415 - 424 ; Brown, J. K. および Imam, H. (1991) J. Inherit. Metab. Dis. 14 : 436 - 458)。神経伝達物質の代謝の多くの経路には、 $\text{NAD}^+ / \text{NADH}$ などの補助因子の酸化または還元を伴う酸化還元酵素の活性が必要である (Newsholme, E. A. および Leech, A. R. (1983) Biochemistry for the Medical Sciences, John Wiley and Sons, Chichester, U. K. pp. 779 - 793)。カテコールアミン(エピネフリンまたはノルエピネフリン)の分解には、脳におけるアルコール脱水素酵素または末梢組織におけるアルデヒド脱水素酵素が必要である。10 NAD^+ 依存性アルデヒド脱水素酵素は、脳、血小板、肺、および肺内皮における5-ヒドロキシインドール-3-酢酸(5-ヒドロキシトリプタミン(セロトニン)の代謝産物)を酸化する(Newsholme, E. A. および Leech, A. R. 前出 p. 786)。 $\text{NAD}^+ / \text{NADH}$ 依存性酸化還元酵素活性を利用する他の神経伝達物質の分解経路には、L-DOPA(ドーパミンの前駆体、神経興奮化合物(neuronal excitatory compound))、グリシン(脳および脊髄における抑制性伝達物質)、ヒスタミン(炎症反応の際に肥満細胞から遊離)、およびタウリン(脳幹、脊髄、および網膜の抑制性伝達物質)の分解経路がある(Newsholme, E. A. および Leech, A. R. supra, pp. 790, 792)。20 神経伝達物質の代謝経路における後成的または遺伝的な障害は、パーキンソン病や遺伝性ミオクローヌスを含む様々な組織における或る範囲の病態につながり得る(McCance, K. L. および Huether, S. E. (1994) Pathophysiology, Mosby-Year Book, Inc., St. Louis, MO pp. 402 - 404 ; Gundlach, A. L. (1990) FASEB J. 4 : 2761 - 2766)。

【0006】

テトラヒドロ葉酸は担体として作用するグルタミン酸分子の誘導体であって、活性化した1炭素単位をプリン、ピリミジン、およびアミノ酸メチオニンの生合成を含む様々な生合成反応に提供する。テトラヒドロ葉酸は、テトラヒドロ葉酸脱水素酵素、テトラヒドロ葉酸シクロヒドラーゼ、およびテトラヒドロ葉酸合成酵素の3つの酵素活性を含むテトラヒ20 ドロ葉酸合成酵素と呼ばれるホロ酵素複合体の活性によって生成される。従って、テトラヒドロ葉酸脱水素酵素は、細胞増殖に必須である核酸およびアミノ酸のブロック形成において重要な役割を果たしている。

【0007】

3-ヒドロキシアシル-CoAデヒドロゲナーゼ(3HACD)は脂肪酸の代謝に関与する。3HACDは、真核生物細胞のミトコンドリアおよびペルオキシソームにおいて、 NAD の NADH への酸化が同時に生じる3-ヒドロキシアシル-CoAの3-オキソアシル-CoAへの還元を触媒する。ペルオキシソームでは、3HACDおよびエノイル-CoAヒドラーゼが2機能酵素(bifunctional enzyme)と呼ばれる酵素複合体を形成し、この複合体の欠損がペルオキシソーム2機能酵素の欠損症に関連40 する。この脂肪酸代謝の中断によって、脳、骨、および副腎の発達を阻害する超長鎖脂肪酸が生成される。この欠損症の症状をもって生まれる乳児は、通常6ヶ月以内に死亡する(Watkins, P. ら (1989) J. Clin. Invest. 83 : 771 - 777 ; Online Mendelian Inheritance in Man (OMIM), #261515)。アルツハイマー病の特徴である神経変性は、脳の或る領域における細胞外プラークを伴う。これらのプラークの主なタンパク質成分は、アミロイド前駆体タンパク質(APP)の幾つかの分解産物の1つであるアミロイドペプチド(A β)である。A β と結合することが実証されている3HACDが、アルツハイマー病に感染したニューロンで過剰に発現されている。更に、抗3HACD抗体はアルツハイマー病の培養細胞モデルにおけるA β の毒作用を遮断し得る(Yan, S. ら (50

1997) Nature 389 : 689 - 695 ; OMIM , # 602057)

【0008】

エストロゲン、テストステロン、およびコルチコステロンなどのステロイドは、共通の前駆体であるコレステロールから生成され、互いに相互変換される。幾つかの脱水素酵素を含む様々な酵素がコレステロールに作用する。ヒドロキシステロイド脱水素酵素などのステロイド脱水素酵素は、高血圧症、肥満、癌に関与する (Duax , W . L . and Ghosh , D . (1997) Steroids 62 : 95 - 100) 。このような脱水素酵素の例には、前立腺およびほかのアンドロゲン応答性組織に高発現するミクロソーム膜タンパク質である3 - オキシ - 5 - ステロイド脱水素酵素 (O S A D) が

ある。O S A Dは、テストステロンのジヒドロテストステロンへの変換を触媒する。ジヒドロテストステロンは、最も強力なアンドロゲンである。ジヒドロテストステロンは、胚発生時の男性表現型の形成、並びに適切なアンドロゲン仲介性の組織 (前立腺や男性生殖器など) の成長に必須である。テストステロンのジヒドロテストステロンへの変換を阻害するO S A Dの欠損は、外性器の不完全形成によって特徴付けられる稀に見られる男性偽雌雄同体型を引き起こす (Andersson , S . , ら (1991) Nature 354 : 159 - 161 ; Labrie , F . , ら (1992) Endocrinology 131 : 1571 - 1573 ; OMIM # 264600) 。従って、O S A Dは性分化およびアンドロゲンの生理機能に中心的な役割を果たす。

10

【0009】

17 - ヒドロキシステロイド脱水素酵素 (17 H S D 6) は、男性ホルモンであるジヒドロテストステロン (D H T T) の調節に重要な役割を果たしている。17 H S D 6は、D H T Tの前駆体である3 - ジオールを容易にグルクロン酸化して組織から除去されるアンドロステロンに酸化してD H T Tのレベルを低下させるように作用する。17 H S D 6はまた、胚腎293細胞で発現されるとアンドロゲン基質およびエストロゲン基質の両方で活性である。異なったステロイド基質に特異的であって、様々な組織における酸化反応および/または還元反応を触媒する17 H S Dの少なくとも5つの他のイソ酵素が同定された (Biswas , M . G . および Russell , D . W . (1997) J . Biol . Chem . 272 : 15959 - 15966) 。例えば17 H S D 1はエストラジールを選択的に還元し、卵巣および胎盤に豊富である。17 H S D 2はアンドロゲンの酸化を触媒し、子宮内膜および胎盤に存在する。17 H S D 3は、精巣のみに存在する還元酵素である (Geissler , W . M . ら (1994) Nature Genet . 7 : 34 - 39) 。D H T Tなどのアンドロゲンの過剰な存在は、良性前立腺肥大および前立腺癌などの或る種の病態を引き起こし得る。

20

30

【0010】

酸化還元酵素は、ミトコンドリアおよびペルオキシソームにおける脂肪酸代謝経路の構成要素である。主 酸化経路は飽和脂肪酸および不飽和脂肪酸の両方を分解し、一方、補助経路は不飽和脂肪酸の分解に必要な更なるステップを行う。補助 酸化酵素である2 , 4 - dienoyl - CoA還元酵素が、主 酸化経路に入る前に不飽和脂肪酸からの偶数個の二重結合の除去の触媒をする。この酵素はまた、不飽和脂肪酸からの奇数個の二重結合の除去もし得る (Koivuranta , K . T . ら (1994) Biochem . J . 304 : 787 - 792 ; Smeland , T . E . ら (1992) Proc . Natl . Acad . Sci . USA 89 : 6673 - 6677) 。2 , 4 - dienoyl - CoA還元酵素は、ミトコンドリアおよびペルオキシソームの両方に存在する。ミトコンドリアおよびペルオキシソームにおける 酸化酵素の遺伝性の欠損は重症疾患に関連し、そのような重症疾患の中には、生後すぐに症状が現われて、数週間以内に死亡に至るものもある。 酸化酵素の欠損は、ライ症候群、ツェルヴェーガー症候群、新生児副腎性白質ジストロフィー、乳児レフサム病、アシルCoA酸化酵素欠損症、および二機能タンパク質 (bifunctional protein) 欠損症に関連する (Suzuki , Y . ら (1994) Am . J . Hum . Genet . 5

40

50

4 : 36 - 43 ; Hoefler, 前出 ; Cotran, R. S. ら (1994) Robbins Pathologic Basis of Disease, W. B. Saunders Co., Philadelphia, PA, p. 866)。ペルオキシソーム酸化酵素は、癌性組織においてその機能が損なわれる。腫瘍性ヒト乳房上皮細胞は、正常な細胞と同数のペルオキシソームを有するが、脂肪酸アシルCoA酸化酵素活性はコントロール組織よりも低い (el Bouhtoury, F., ら (1992) J. Pathol. 166 : 27 - 35)。ヒト結腸癌は、正常な結腸組織よりもペルオキシソームが少なく、脂肪酸アシルCoA酸化酵素活性および二機能酵素 (エノイル-CoAヒドラーゼを含む) 活性が正常な組織よりも低い (Cable, S., ら (1992) Virchows Arch. B Cell Pathol. Incl. Mol. Pathol. 62 : 221 - 226)。クエン酸回路の基質であるケトグルタル酸へのイソクエン酸の変換を触媒するイソクエン酸脱水素酵素が、もう1つの重要な酸化還元酵素である。イソクエン酸脱水素酵素は、NAD依存性若しくはNADP依存性の何れかであって、サイトゾル、ミトコンドリア、およびペルオキシソームに見られる。イソクエン酸脱水素酵素の活性は、ホルモン、神経伝達物質、および成長因子によって発生的に調節される。

10

20

【0011】

グリコール酸経路におけるペルオキシソームの2-ヒドロキシ酸脱水素酵素 (2-hydroxy acid dehydrogenase) であるヒドロキシピルビン酸レダクターゼ (HPR) が、NADHおよびNADPHの両方の酸化を伴うヒドロキシピルビン酸のグリセリン酸への変換を触媒する。逆脱水素酵素反応により、NAD⁺ およびNADP⁺ が減少する。HPRは、ヌクレオチドおよび塩基をATPおよびGTPが合成される経路に戻して再利用する。ATPおよびGTPは、DNAおよびRNAを生成するため、並びにシグナル伝達およびエネルギー代謝を様々な点を調節するために用いられる。プリンヌクレオチド生合成の阻害剤が、癌およびウイルス疾患の治療に抗増殖性薬剤として以前から使用されている。HPPはまた、セリンの生合成を調節して、タンパク質の合成に利用可能な細胞内のセリンのレベルを調節する。

30

40

【0012】

ミトコンドリアの電子伝達鎖 (呼吸鎖) は、ミトコンドリア膜における一連の酸化還元酵素型酵素複合体であって、電子をNADHからこれら複体内の一連のレドックスコアを介して酸素に伝達し、この酸化がATPの合成につながる (酸化的リン酸化)。次に、ATPが、エネルギーが必要な多くの細胞の反応が行われるようにするための主なエネルギー源となる。呼吸鎖における主な複合体には、NADH:ユビキノン酸化還元酵素 (複合体I)、コハク酸:ユビキノン酸化還元酵素 (複合体II)、チトクロームc₁-b酸化還元酵素 (複合体III)、チトクロームc酸化酵素 (複合体IV)、およびATPシンターゼ (複合体V) が含まれる (Alberts, B. ら (1994) Molecular Biology of the Cell, Garland Publishing, Inc., New York, NY, p. 677 - 678)。複合体IIはサイトゾル側に存在するが、それ以外の複合体は全て、ミトコンドリア膜の内側のマトリックス側に存在する。複合体IIは、クエン酸回路で生成された電子を呼吸鎖に伝達する。クエン酸回路においてコハク酸のフマル酸への酸化により生じた電子が、複合体IIの電子伝達体を介して膜結合ユビキノン (Q) に伝達される。これら核コード遺伝子の転写調節が、呼吸酵素の生合成を調節するための主な手段であると思われる。呼吸鎖における酵素発現の欠損および変化が、様々な疾患に関連する。

【0013】

補助因子としてNADを用いる他の脱水素酵素活性も、ミトコンドリアの機能に重要である。バリンの異化反応に重要な3-ヒドロキシイソ酪酸脱水素酵素 (3HBD: 3-hydroxyisobutyrate dehydrogenase) が、ミトコンドリア内における3-ヒドロキシイソ酪酸のメチルマロン酸セミアルデヒドへのNAD依存性の酸化を触媒する。3-ヒドロキシイソ酪酸のレベルの上昇が、ケトアシドーシス、メチル

50

マロン酸血症、およびメチルマロン酸セミアルデヒド脱水素酵素に関連した他の疾患を含む様々な病態で報告された (Rougraff, P. M. ら (1989) *J. Biol. Chem.* 264 : 5899 - 5903)。

【0014】

アミノ酸代謝に重要な別のミトコンドリア脱水素酵素として、イソヴァレリル - CoA 脱水素酵素 (IVD) がある。IVD はロイシン代謝に関与し、イソヴァレリル - CoA の 3-methylcrotonyl-CoA への酸化を触媒する。ヒト IVD は四量体フラボタンパク質であって、核でコードされ、ミトコンドリア移入シグナル配列を有する 45 kDa の前駆体としてサイトゾルで合成される。IVD をコードする遺伝子の変異から生じる遺伝性疾患により、イソ吉草酸血症として知られる症状が引き起こされる。この変異により、IVD 前駆体のミトコンドリアへの移入およびプロセッシングが行われなくなる (Vockley, J. ら (1992) *J. Biol. Chem.* 267 : 2494 - 2501)。

10

【0015】

トランスフェラーゼ

トランスフェラーゼは、分子群の転移を触媒する酵素である。反応は、酸化、還元、または共有結合の開裂を伴い、基質または或る種の基質上の特定部位に特異的であることが多い。トランスフェラーゼは、細胞成分の合成及び分解などの機能、および細胞機能の調節に必須な反応に関与する。細胞機能とは、細胞シグナル伝達、細胞増殖、炎症、アポトーシス、分泌、および排泄が含まれる。トランスフェラーゼは、転移される群の型によって分類される場合が多い。例えば、メチルトランスフェラーゼは一炭素メチル基を転移し、アミノトランスフェラーゼは窒素アミノ基を転移し、同様に命名された酵素は、補酵素 A などの小さな酵素群は勿論、アルデヒド基、若しくはケトン基、アシル基、グリコシル基、アルキル基又はアリル基、イソプレニル基、saccharyl 基、リン含有基、硫黄含有基、若しくはセレンウム含有基を転移する。

20

【0016】

アシルトランスフェラーゼは、脂肪酸酸化経路に関与するペルオキシソームカルニチンオクタノイルトランスフェラーゼ、および脂肪酸の代謝及び輸送に関与する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を含む。コリン O - アセチルトランスフェラーゼは、神経伝達物質であるアセチルコリンの生合成を触媒する。

30

【0017】

アミノトランスフェラーゼは、タンパク質の合成及び分解において重要な役割を果たし、さらに別のプロセスにも貢献する。例えば、アミノトランスフェラーゼ 5 - アミノレブリン酸シンターゼは、ヘム生合成の第一ステップであるグリシンに対するスクシニル CoA の付加を触媒する。別のアミノトランスフェラーゼは、神経の機能及び代謝に重要な経路に関与する。例えば、グルタミントランスアミナーゼ K (GTK) としても知られるグルタミンフェニルピルビン酸アミノトランスフェラーゼは、ピリドキサルリン酸補助因子と共に複数の反応を触媒する。GTK は、L - グルタミンとフェニルピルビン酸を 2-oxoglutaramate 及び L - フェニルアラニンへの可逆的な変換を触媒する。GTK のその他のアミノ酸基質には、L - メチオニン、L - ヒスチジン、及び L - チロシンが含まれる。GTK はまた、キヌレニンのキヌレン酸への変換を触媒する。このキヌレン酸はトリプトファン代謝産物であり、脳における N - メチル - D - アスパラギン酸 (NMDA) 受容体のアンタゴニストであって神経調節機能として作用しうる。キヌレニン代謝経路の変化は、複数の神経疾患に関連しうる。GTK はまた、グルタチオンに結合したハロゲン化生体異物の代謝において役割を果たし、ラットでは腎毒性を、ヒトでは神経毒性を引き起こす。GTK は、腎臓、肝臓、及び脳内で発現する。ヒト GTK 及びラット GTK の双方は、推定ピリドキサルリン酸結合部位を含む (EXPASY ENZYME : EC 2.6.1.64 ; Perry, S. J. ら (1993) *Mol. Pharmacol.* 43 : 660 - 665、Perry, S. ら (1995) *FEBS Lett.* 360 : 277 - 280、及び Alberati Gianini, D. ら (

40

50

1995) J. Neurochem. 64: 1448-1455)。この経路に関連する第二のアミノトランスフェラーゼは、キヌレニンノ - アミノアジピン酸アミノトランスフェラーゼ (AadAT) である。AadATは、リジンの代謝の際、 - アミノアジピン酸及び - ケトグルタル酸塩の - ケトアジペート及びL - グルタメートへの可逆的な変換を触媒する。AadATはまた、キヌレニンからキヌレン酸へのアミノ基の転移も触媒する。サイトゾルAadATは、ラット腎臓、肝臓、及び脳内で発現する (Nakatani, Y. ら (1970) Biochim. Biophys. Acta 198: 219-228、Buchli, R. ら (1995) J. Biol. Chem. 270: 29330-29335)。

【0018】

グリコシルトランスフェラーゼは、膜結合ミクロソーム酵素のファミリーである哺乳動物UDP - glucouronosylトランスフェラーゼを含む。この膜結合ミクロソーム酵素は、薬剤、発癌物質、およびその他の外来物質の解毒および排泄において重要な役割を果たす反応において、glucouronic acidの親油性基質への転移を触媒する。もう1つの哺乳動物グリコシルトランスフェラーゼである哺乳動物UDP - ガラクトースセラミドトランスフェラーゼは、神経系のミエリン膜におけるガラクトセレブロシドの合成においてガラクトースのセラミドへの転移を触媒する。UDP - グリコシルトランスフェラーゼは、約50のアミノ酸残基の保存されたシグネチャドメインを共有する (PROSITE: PDOC00359, <http://expasy.hcu.ge.ch/sprot/prosite.html>)。

【0019】

メチルトランスフェラーゼは、様々な薬理的に重要なプロセスに関与する。ニコチンアミドN - メチルトランスフェラーゼは、ニコチンアミド及び他のピリジンのN - メチル化を触媒する。これは、薬剤及び他の外来化合物の細胞内処理における重要なステップである。フェニルエタノールアミンN - メチルトランスフェラーゼは、ノルアドレナリンのアドレナリンへの変換を触媒する。6 - O - メチルグアニン - DNAメチルトランスフェラーゼはDNAのメチル化を逆転する。これは、発癌に於ける重要なステップである。S - アデノシル - L - メチオニンからウロポルフィリノーゲンIIIへの2つのメチル基の転移を触媒するウロポルフィリン - III Cメチルトランスフェラーゼは、コバラミンの生合成における第1の特異的酵素である。このコバラミンは、悪性貧血において取り込みが減少する食物酵素である。タンパク質 - アルギニンメチルトランスフェラーゼは、タンパク質におけるアルギニン残基の翻訳後メチル化を触媒し、それによってグアニジノ基におけるアルギニンのモノメチル化またはジメチル化が生じる。基質には、ヒストンやミエリン塩基性タンパク質、更にmRNAのプロセッシング、スプライシング、および輸送に関与するヘテロ核リボ核タンパク質が含まれる。タンパク質 - アルギニンメチルトランスフェラーゼは、マイトジェンによってアップレギュレートされるタンパク質、慢性リンパ性白血病に関与するタンパク質、およびインターフェロンと相互作用することから、サイトカイン受容体シグナル伝達におけるメチル化に重要な役割を果たしていると考えられる (Lin, W. - J. ら (1996) J. Biol. Chem. 271: 15034-15044; Abramovich, C. ら (1997) EMBO J. 16: 260-266; および Scott, H. S. ら (1998) Genomics 48: 330340)。

【0020】

ホスホトランスフェラーゼは高エネルギーのリン酸基の転移を触媒する。このため、エネルギーを必要とする反応やエネルギー放出反応に重要である。代謝酵素であるクレアチンキナーゼは、クレアチン/クレアチンリン酸とATP/ADPとの間の可逆的なリン酸の転移を触媒する。グリコシアミンキナーゼはATPからguanidoacetateへのリン酸の転移を触媒し、アルギニンキナーゼはATPからアルギニンへのリン酸の転移を触媒する。システイン含有活性部位がこのファミリーに保存されている (PROSITE: PDOC00103)。

10

20

30

40

50

【0021】

プレニルトランスフェラーゼは、イソプレニル基の転移を触媒する、サブユニットおよびサブユニットからなるヘテロ二量体である。プレニルトランスフェラーゼの例には、哺乳動物タンパク質であるファルネシルトランスフェラーゼがある。ファルネシルトランスフェラーゼのサブユニットは、それぞれがインバリアントなトリプトファンを備えた34個のアミノ酸からなる5つの反復から構成されている (PROSITE : PDOC00703)。

【0022】

saccharylトランスフェラーゼは、様々な代謝プロセスに関与する糖化酵素である。例えば、oligosaccharylトランスフェラーゼ-48は糖化後期の最終産物の受容体である。これらの最終産物の蓄積が、糖尿病性血管合併症、大血管障害、腎不全、およびアルツハイマー病で観察される (Thornalley, P. J. (1998) Cell Mol. Biol. (Noisy Le-Grand) 44: 1013-1023)。

10

【0023】

補酵素A (CoA)トランスフェラーゼは、2つのカルボン酸間のCoAの転移を触媒する。スクシニルCoA:3-オキソ酸CoAトランスフェラーゼ (3-oxoacid CoA transferase)は、例えば、CoAをスクシニルCoAからアセト酢酸などのレシピエントに転移する。アセト酢酸は、ケトン体の代謝に必須であり、糖尿病などの代謝障害によって影響を受けた組織に蓄積される (PROSITE : PDOC00980)。

20

【0024】

加水分解酵素

加水分解とは、水の分子を導入することによって基質内の共有結合を破壊することである。この反応は、基質の標的結合における水分子の酸素原子による求核攻撃を含む。水分子は、標的結合全域で分離され、結合が破壊されて2つの分子が生成される。加水分解酵素は、細胞シグナル伝達、細胞増殖、炎症、アポトーシス、分化、排泄等の作用に不可欠な反応に必要である。加水分解酵素は、これらの作用を含む疾患プロセスにおいて鍵となるステップに関与している。加水分解酵素即ちヒドロラーゼは、基質特性によって、aminohydrolase、ホスホリパーゼ、カルボキシルエステラーゼ、ホスホジエステラーゼ、リゾチーム、グリコシダーゼ、glyoxylase、スルファターゼ、phosphohydrolase、及びserine hydrolaseを含む群に分類することができる。

30

【0025】

ホスホジエステラーゼは、ホスホジエステル化合物の2つのエステル結合の内の一方のエステル結合の加水分解を触媒する。従って、ホスホジエステラーゼは、様々な細胞のプロセスに極めて重要である。ホスホジエステラーゼは、蛋白合成及び細胞の成長及び複製に不可欠であるようなDNA及びRNAのエンドヌクレアーゼ及びエキソヌクレアーゼを含む。

40

【0026】

腓リパーゼ及びコリパーゼは、複合体を形成する。この複合体は、不溶性の長鎖トリアシルグリセロールを腸管細胞の刷子縁を越えることができる極性の高い分子へ変換することによって食事脂肪の分解において重要な役割を演じる。コリパーゼは、リパーゼのC末端ドメインに結合する。

【0027】

カルボキシルエステラーゼは、カルボン酸エステルを加水分解するタンパク質であって、A、B、Cの3つのカテゴリーに分けられる。殆どのB型カルボキシルエステラーゼは、進化論的に関連し、タンパク質ファミリーを構成するものと考えられる。タンパク質のB型カルボキシルエステラーゼファミリーには、脊椎動物アセチルコリンエステラーゼ、哺乳類肝ミクロソームカルボキシルエステラーゼ、哺乳類胆汁酸塩活性化リパーゼ、アヒル

50

脂肪アシル - CoA ヒドロラーゼが含まれる。このタンパク質ファミリーのメンバーには、触媒的に活性でないがサイログロブリンやショウジョウバエタンパク質 *neuractin* などのその他の B 型カルボキシルエステラーゼに進化論的に関連するドメインを含むものもある。

【0028】

アシル - CoA チオエステラーゼは、カルボキシルエステラーゼファミリーの別のメンバーである (Alexson, S. E. ら (1993) *Eur. J. Biochem.* 214: 719 - 727)。アシル - CoA チオエステラーゼがステロイド産生組織において調節的役割を有していることを示唆している証拠がある (Finkielstein, C. ら (1998) *Eur. J. Biochem.* 256: 60 - 66)。

【0029】

ホスホリパーゼ A₂ 阻害剤は、ヒトのロイシンリッチ₂ 糖タンパク質と 33% の配列相同性を有していることが分かった (Okumura, K. ら (1998) *J. Biol. Chem.* 273: 19469 - 19475)。ロイシンリッチリピート (LRR) コンセンサス配列はまた、ホルモン受容体、酵素、酵素阻害剤、細胞癒着のためのタンパク質、およびリポソーム結合タンパク質等の生物学的に重要なプロセスに関係するタンパク質を含む多数のタンパク質の 1 次構造においても見られる。LRR ドメインを含有する全てのタンパク質は、タンパク質 - タンパク質相互作用に関与していると考えられる。

【0030】

リゾチーム c スーパーファミリーは、通常のリゾチーム c、カルシウム結合リゾチーム c、及び ラクトアルブミンからなる (Prager, E. M. 及び P. Jolles (1996) *EXS* 75: 9 - 31)。リゾチーム c スーパーファミリーのタンパク質は、35 ~ 40% の配列相同性及び共通の 3 次元フォールドを有するが、異なる機能を有することがある。リゾチームは、糖におけるグリコシド結合の形成および切断を行う (Iyer, L. K. and Qasba, P. K. (1999) *Protein Eng.* 12 (2): 129 - 139)。リゾチーム c は、様々な組織及び分泌物に遍在し、或る種の細菌の細胞壁を溶解させることができる (McKenzie, H. A. (1996) *EXS* 75: 365 - 409)。

【0031】

glyoxylase システムは、*methyglyoxal* からの *S-Dlactoylglutathione* の生成を触媒する *glyoxylase I* と、トリオースリン酸エネルギー代謝の副次生成物と、*S-Dlactoylglutathione* を D-乳酸と還元グルタチオンに加水分解する *glyoxylase II* とからなる。*methyglyoxal* のレベルは高血糖時に上昇するが、これはトリオースリン酸エネルギー代謝の上昇による可能性が高い。*glyoxylase II* 活性の上昇が、ヒトの非インスリン依存型糖尿病及びラット同疾患モデルにおいて見られる。*glyoxylase* システムは、細菌毒の解毒と、細胞増殖及び微小管構造の調節に関与している。*glyoxylase II* の基質である *S-Dlactoylglutathione* のレベルの上昇により、HL60 細胞の成長停止及び毒性を誘導した。従って、*glyoxylase* システム、特に *glyoxylase II* は、細胞増殖や、糖尿病等の自己免疫疾患に関連するであろう。

【0032】

/ 加水分解酵素フォールド (立体構造) は、多様な系統発生的起源や触媒機能を有する幾つかの加水分解酵素に共通のタンパク質の折りたたみである。/ 加水分解酵素フォールドを有する酵素は、ヘリックスで連結された 8 枚のシートからなる共通の核構造を有する。このフォールドの最も保存された構造的特徴は、求核 - ヒスチジン - 酸触媒 3 つ組 (*nucleophile-histidine-acid catalytic triad*) のループである。求核性ループ及び酸ループは 2 つ以上のアミノ酸の型を含

むが、触媒3つ組におけるヒスチジンは完全に保存されている (Ollis, D. L. ら (1992) Protein Eng. 5:197-211)。

【0033】

スルファターゼは、広範囲にわたる配列相同性及び高い構造的類似性を共有するような高度に保存された遺伝子ファミリーのメンバーである。スルファターゼは、硫酸エステルの切断を触媒する。この機能を発揮するため、スルファターゼは小胞体において、保存されたシステイン残基の酸化を伴うユニークな翻訳後修飾を受ける。多発性スルファターゼ欠損症と称されるヒトの疾患は、スルファターゼを不活性にするようなこの翻訳後修飾段階における障害に起因する (Recksiek, M. ら (1998) J. Biol. Chem. 273:6096-6103)。

10

【0034】

phosphohydrolaseは、リン酸エステルを加水分解する酵素である。phosphohydrolaseの中には、mutTドメインシグネチャ配列を含むものもある。mutTは、DNAから酸化的障害を受けたグアニンを取り除くGOシステムに参与するタンパク質である。

【0035】

グリコシダーゼは、1個以上の糖を含む化合物であるグリコシドのhemiacetyl結合の切断を触媒する。哺乳類ガラクトシダーゼは、ガングリオシド、糖タンパク質、及びグリコサミノグリカンから末端ガラクトースを取り除く。ガラクトシダーゼは、glycosyl hydrolaseに分類されるファミリー35に属する。

20

【0036】

serine hydrolaseは、その活性部位にセリン残基を含む加水分解酵素の機能的クラスである。この酵素クラスは、多様な基質を加水分解するプロテイナーゼ、エステラーゼ、及びリパーゼを含むことから、様々な生物学的役割を有する。このスーパーファミリーのタンパク質は、基質特異性またはアミノ酸類似性に基づいて更にサブファミリーに分類することができる (Puente, X. S. 及び C. Lopez-Ont (1995) J. Biol. Chem. 270:12926-12932)。

【0037】

リアーゼ

リアーゼは、非加水分解的または非酸化的にC-C、C-O、C-N、C-S、C-(ハロゲン化物)、P-Oまたはその他の結合を切断し、少なくとも一方が二重結合を有する2つの分子を生成する反応を触媒する酵素のクラスである (Stryer, L. (1995) Biochemistry, W. H. Freeman and Co., New York NY, p. 620)。リアーゼは、脂肪酸代謝及び他の多様な酵素的プロセスを含む代謝エネルギー生産において作用する細胞生化学の重要な成分である。

30

【0038】

C-Cリアーゼ群には、carboxyl-lyase (デカルボキシラーゼ)、アルデヒドリアーゼ (アルドラーゼ)、オキソ酸リアーゼ (oxo-acid-lyase) 及びその他のリアーゼが含まれる。C-Oリアーゼ群には、多糖類に作用するリアーゼであるヒドロリアーゼ及びその他のリアーゼが含まれる。C-Nリアーゼ群には、アンモニアリアーゼ、アミジンリアーゼ、amine-lyase (デアミナーゼ) 及びその他のリアーゼが含まれる。

40

【0039】

リアーゼの適切な調節は通常 of 生理機能に必須である。例えば、変異誘導性のウロポルフィリノーゲンデカルボキシラーゼ欠損が、遺伝性疾患である家族性晩発性皮膚ポルフィリン症の感光性皮膚障害を引き起こし得る (Mendez, M. ら (1998) Am. J. Genet. 63:1363-1375)。ADA遺伝子の遺伝子変異に起因するアデノシンデアミナーゼ (ADA) 欠損は、重症複合免疫不全症 (SCID) を引き起こすことが示された (Hershfield, M. S. (1998) Semin. Hematol. 35:291-298)。

50

【0040】

イソメラーゼ

イソメラーゼは酵素の1つのクラスであって、分子内の形状または構造の変化を触媒して単一の生成物を形成する。このクラスには、エピメラーゼ、シス-トランス-イソメラーゼ、細胞内酸化還元酵素、細胞内転移酵素(ムターゼ)、および細胞内リアーゼが含まれる。イソメラーゼは細胞内生化学の重要な成分であって、タンパク質の折りたたみ、光伝達、解糖やその他の多様な酵素プロセスを含む代謝エネルギー生成において作用する(Stryer, L. (1995) Biochemistry W. H. Freeman and Co. New York, NY pp. 483-507)。

【0041】

ラセマーゼは、非対称の1つの中心を有する基質における非対称の炭素原子の周りの分子構造の反転を触媒し、それによって2つのラセマーゼを相互変換するイソメラーゼのサブセットである。エピメラーゼは、2つ以上の非対称の中心を有する基質における非対称の炭素原子の周りの構造の反転を触媒し、それによって2つのエピマーが相互変換するイソメラーゼの別のサブセットである。ラセマーゼ及びエピメラーゼは、アミノ酸及び誘導體、ヒドロキシ酸及び誘導體、炭水化物及びその誘導體に作用し得る。UDP-ガラクトースとUDP-グルコースの相互変換は、UDP-ガラクトース-4'-エピメラーゼによって触媒される。このエピメラーゼの適正な調節及び機能は、糖タンパク質及び糖脂質の合成に必須である。血中のガラクトースのレベルの上昇は、胎児のスクリーニングプログラムにおけるUDP-ガラクトース-4'-エピメラーゼの欠損に相関性があった(Gitzelmann, R. (1972) Helv. Paediat. Acta 27: 125-130)。

【0042】

フォールディング酵素(folding enzyme)の1クラスであるペプチジルプロリル・シス-トランス-イソメラーゼ(PPiase)は、タンパク質の成熟及び輸送における律速段階であると考えられるある種のプロリンイミド結合を異性化する。PPiaseは、タンパク質におけるある種のプロリンイミド結合のシス-トランス異性化を触媒する。PPiaseには、サイクロフィリンファミリー、FK506結合タンパク質ファミリー、及び新規に特徴づけられたparvulinファミリーの進化的に関連性のない3つのファミリーがある(Rahfeld, J. U. 他 (1994) FEBS Lett. 352: 180-184)。

【0043】

サイクロフィリン(CyP)は、初めはT細胞活性化のインヒビターである免疫抑制剤シクロスポリンA(CsA)の主受容体として同定された(Handschumacher, R. E. 他 (1984) Science 226: 544-547; Hardin, M. W. 他 (1986) J. Biol. Chem. 261: 8547-8555)。従って、CyPのペプチジルプロリルイソメラーゼ活性は、T細胞を活性化させるシグナル伝達経路の一部であると考えられる。その後の研究で、CyPのイソメラーゼ活性が、正確なタンパク質の折りたたみ及び/またはタンパク質の輸送に必須であり、タンパク質複合体の構築/分解並びにタンパク質の活性の調節に関与し得ることが分かった。例えば、ショウジョウバエにおいて、CyP Nina Aはロドプシンの正しい局在化に必要であり、哺乳動物CyP(Cyp40)は、ステロイド受容体を結合するHsp90/Hsp70複合体の一部である。哺乳動物CyP(CypA)は、ヒト免疫不全ウイルス1型(HIV-1)からのgagタンパク質と結合することが分かった。この相互作用は、シクロスポリンによって抑制され得る。シクロスポリンは強力な抗HIV-1活性を有するため、CypAは、HIV-1の複製において必須の機能を果たしていると考えられることができる。Cyp40は、転写因子c-Mybに結合してそれを活性化させることが分かった。この作用は、シクロスポリンによって逆にされる。この作用は、転写、形質転換、及び分化の調節にCyPが関与することを示唆するものである(Bergsma, D. J. 他 (1991) J. Biol. Chem. 266: 23204-2321

10

20

30

40

50

4; Hunter, T. (1998) Cell 92: 141-143; and Levenson, J.D. and Nessler, S.A. (1998) Mol. Cell. 1: 203-211)。

【0044】

フォールディング酵素の別のクラスには、タンパク質ジスルフィドイソメラーゼがある。タンパク質の折りたたみにおける主な律速段階の1つが、タンパク質の適正な構築に必須のチオール-ジスルフィド交換である。所定の比率の酸化チオールと還元チオールとを含む緩衝液で、還元された折りたたまれていないタンパク質をインキュベートすると天然の高次構造となるが、折りたたみの速度が遅いため、自然の高次構造への到達度がその大きさ及びそのタンパク質におけるシステインの数に比例して低下する。真核生物の小胞体及び原核生物の細胞膜周辺腔などのある種の細胞区画が、周りのサイトゾルより酸化した状態に維持される。これらの区画でジスルフィドが正しく生成される。しかしながら、その速度は通常細胞プロセスには十分であるが、分泌タンパク質の合成には不十分である。プロテインジスルフィドイソメラーゼであるチオレドキシシン及びグルタレドキシシンは、ジスルフィド結合の形成を触媒することが可能であり、細胞における酸化還元環境を調節し、必要なチオール-ジスルフィド交換を可能とする (Lofler, H. (1995) J. Biol. Chem. 270: 26178-26183)。

10

【0045】

これらのタンパク質のそれぞれの機能はやや異なっているが、それら全てがジスルフィド含有酸化還元タンパク質の群に属する。これらの酸化還元タンパク質は、真核生物及び原核生物において広く見られ、保存された活性部位配列を含む。プロテインジスルフィドイソメラーゼは、真核生物の小胞体及び原核生物の細胞周辺腔に見られる。プロテインジスルフィドイソメラーゼは、折りたたまれているペプチド鎖においてそれ自体のジスルフィドをチオールと交換して機能する。これとは対照的に、還元されたチオレドキシシン及びグルタレドキシシンは、通常は細胞質で見られ、基質タンパク質のジスルフィドを直接還元して機能する。

20

【0046】

これらの触媒分子は、ジスルフィド形成を促進するだけでなく様々な生理学的プロセスに関与する。チオレドキシシン系は、例えばリボヌクレオチドレダクターゼの水素供与体として、及び酸化還元制御によって酵素の制御因子として働く。また、NF- κ B、AP-1、およびステロイド受容体などの転写因子の活性を調節する。近年、成人T細胞白血病由来因子、3B6-インターロイキン1、T細胞ハイブリドーマ由来(MP-6) B細胞刺激因子、及び初期妊娠因子などのいくつかのサイトカインや分泌サイトカイン様因子がチオレドキシシンと同一であることが報告された (Holmgren, A. (1985) Annu. Rev. Biochem. 54: 237-271, Abate, C. 他 (1990) Science 249: 1157-1161, Tagaya, Y. 他 (1989) EMBO J. 8: 757-764, Wakasugi, H. (1987) Proc. Natl. Acad. Sci. 84: 804-808, Rosen, A. 他 (1995) Int. Immunol. 7: 625-633)。チオレドキシシンはまた、細胞成長の制御因子として及び免疫系のメディエーターとしての役割を含む様々な細胞外活性を有することが分かった (Miranda-Vizuete, A. 他 (1996) J. Biol. Chem. 271: 19099-19103, Yamauchi, A. 他 (1992) Mol. Immunol. 29: 263-270)。

30

40

【0047】

酸化還元酵素はイソメラーゼでもあり得る。酸化還元酵素は、酸化される基質から還元される基質への可逆的な電子の逆移送を触媒する。酵素のこのクラスには、脱水素酵素、水酸化酵素、酸化酵素、酸素転化酵素、過酸化酵素、及び還元酵素が含まれる。酸化還元酵素のレベルを適正に維持することは生理学的に重要である。ペントースリン酸経路は、例えば、還元剤NADPHの生成に関与し、それと同時にグルコース-6-ホスファターゼ

50

をリボース - 5 - リン酸に酸化させる酵素を利用する。NADPHは、還元的生合成が起こる反応のためのエネルギーとなる。リボース - 5 - リン酸及びその誘導体は、ATP、補酵素A、 NAD^+ 、FAD、RNA、及びDNAなどの重大な意味をもつ生物分子の一部となる。ペントースリン酸経路は、酸化的枝分かれ経路及び非酸化的枝分かれ経路の双方を有する。グルコース - 6 - リン酸脱水素酵素、ラクトナーゼ、及び6 - リングルコン酸脱水素酵素などの酵素によって触媒される酸化的枝分かれ経路で、グルコース - 6 - リン酸と NADP^+ がリブロース - 6 - リン酸とNADPHに変換される。ホスホメントースイソメラーゼ、ホスホメントースイエピメラーゼ、トランスケトラーゼ、及びトランスアルドラーゼなどの酵素によって触媒される非酸化的枝分かれ経路において、炭素数が3個、4個、5個、6個、及び7個の糖の間での相互転換が起こりうる。

10

【0048】

イソメラーゼの別のサブグループにはトランスフェラーゼ(またはムターゼ)がある。トランスフェラーゼは、化学基を1つの化合物(供与体)から別の化合物(受容体)に転移する。これらの酵素によって転移される基の中には、アシル基、アミノ基、リン酸基(ホスホトランスフェラーゼまたはホスホムターゼ)及びその他の基が含まれる。トランスフェラーゼであるカルニチンパルミトイルトランスフェラーゼは、脂肪酸代謝の重要な成分である。このトランスフェラーゼの遺伝的欠損は、筋障害を引き起こし得る(Scrivener C. R. 他(1995) The Metabolic and Molecular Basis of Inherited Disease, McGraw-Hill New York NY pp. 1501 - 1533)。

20

【0049】

イソメラーゼの別のサブグループには、ポイソメラーゼがある。ポイソメラーゼは、DNAの形態状態に影響を及ぼす酵素である。例えば、トポイソメラーゼまたはそれらの調節における障害は、通常の生理に影響を及ぼし得る。トポイソメラーゼIIのレベルの低下は、毛細血管拡張性運動失調に関連する或る種のDNAプロセスの障害に相関する(Singh, S. P. ら(1988) Nucleic Acids Res. 16 : 3919 - 3929)。

【0050】

リガーゼ

リガーゼは、2つの基質分子間の結合の形成を触媒する。このプロセスには、ATPまたは類似のエネルギー供与体におけるピロリン酸結合の加水分解が伴う。リガーゼは、炭素 - 酸素、炭素 - 硫黄、炭素 - 窒素、炭素 - 炭素、及びリン酸エステル結合を含み得る形成する結合の種類に基づいて分類される。

30

【0051】

炭素 - 酸素結合を形成するリガーゼには、翻訳に役割を果たす重要なRNA関連酵素であるアミノアシル運搬RNA(tRNA)合成酵素が含まれる。タンパク質の生合成は、適当なtRNAと結合を形成する各アミノ酸に依存する。アミノアシルtRNA合成酵素は、アミノ酸の活性化及びそのアミノ酸に対応するtRNAの接着に作用する。20のアミノアシルtRNA合成酵素が構造から2つのクラスに分類することができ、それぞれのクラスは触媒ドメインの固有の形状によって特徴づけられている。クラスI酵素は、ヌクレオチド結合ロスマンフォールドに基づいた触媒ドメインを含む。クラスII酵素は、7本鎖逆平行シートモチーフ、N末端調節ドメイン、及びC末端調節ドメインからなる中央の触媒ドメインを含む。クラスII酵素は、酵素の構造に基づいてヘテロ二量体およびホモ二量体の2つのグループに分類され、ホモ二量体のグループは、N末端調節ドメインの構造及びC末端調節ドメインの構造によって更に分類される(Hartlein, M. およびCusack, S. (1995) J. Mol. Evol. 40 : 519 - 530)。抗アミノアシルtRNA自己抗体が、皮膚筋炎及び多発性筋炎の患者によって産生され、合併症を伴う間質性肺疾患(ILD)に強い相関性がある。これらの抗体は、ウイルス感染に反応して産生されると思われ、コクサッキーウイルスを用いて動物においてウイルス筋炎を実験的に引き起こした。

40

50

【0052】

炭素 - 硫黄結合を形成するリガーゼ（酸 - チオールリガーゼ）は、炭素原子含有基質（炭素基質）の分子間輸送を伴う多数の細胞内生合成中間代謝プロセスを仲介する。このような反応の例には、クエン酸回路、脂肪酸および長鎖リン脂質の合成、アルコールおよびアルデヒドの合成、中間代謝物の合成、およびアミノ酸分解経路に参与する反応がある。これらの反応の中には、A D P若しくはA N Pとピロリン酸へのA T Pの変換の形であるエネルギーが必要な反応もある。

【0053】

多くの場合、炭素基質は、少なくとも2つの炭素原子を含む小分子に由来する。炭素基質は、細胞内で炭素基質キャリア分子として働く大きな分子に共有結合している場合が多い。上記した生合成機構では、キャリア分子は補酵素Aである。補酵素A（C o A）は、ヌクレオチドA D Pの誘導体に構造的に関連し、ホスホジエステル結合によりアデノシン3' , 5' - ニリン酸のリン酸基に結合した4' - ホスホパンテテインからなる。4' - ホスホパンテテインの末端チオール基は、炭素基質の結合を形成するための部位として作用する。細胞における生合成および中間代謝の際にキャリア分子としてC o Aを用いる主な炭素基質には、アセチル成分、サクシニル成分、およびプロピオニル成分が含まれ、総称してアシル基と呼ばれる。他の炭素基質には、脂肪酸酸化中間体として働くエノイル脂質（e n o y l l i p i d）、並びにアセチルC o A流入調節因子/ミトコンドリアアシル基輸送タンパク質として働くカルニチンが含まれる。アシルC o AおよびアセチルC o Aはそれぞれ、アシルC o A合成酵素およびアセチルC o A合成酵素によって細胞内で生合成される。

【0054】

脂肪酸の活性化は、少なくとも3つの型のアシルC o A合成酵素、すなわち（1）アセチルC o A合成酵素、（2）中鎖アシルC o A合成酵素、および（3）アシルC o A合成酵素の活性によって仲介される。アセチルC o A合成酵素は、アセテートおよび幾つかの他の低分子量カルボン酸を活性化させ、筋ミトコンドリアおよびその他の組織のサイトゾルに見られる。中鎖アシルC o A合成酵素は、主に食事に由来する4個から7個の炭素原子を含む脂肪酸を活性化させ、肝ミトコンドリアのみに存在する。アシルC o A合成酵素は、6個から20個の炭素原子を有する長鎖脂肪酸に特異的であって、ミクロソームおよびミトコンドリアに見られる。アシルC o A合成酵素活性に関連するタンパク質が、細菌、酵母、植物、マウス、およびヒトを含む様々なソースで確認された。アシルC o A合成酵素の活性は、c A M P依存性プロテインキナーゼによる酵素のリン酸化によって調節され得る。

【0055】

炭素 - 窒素結合を形成するリガーゼには、A T P加水分解のエネルギーを用いるアンモニアによるグルタミン酸のグルタミンへのアミノ化を触媒するグルタミン合成酵素（グルタミン酸アンモニアリガーゼ）などのアミド合成酵素が含まれる。グルタミンは、新規のピリミジンヌクレオチド合成、並びにプリンおよびピリミジンリボヌクレオチド相互変換に参与する様々なアミド転移反応におけるアミノ基の主な供給源である。グルタミン合成酵素の過剰な発現が、原発性肝癌で観察された（C h r i s t a , L . 他（1994）G a s t r o e n t . 106 : 1312 - 1320）。

【0056】

酸 - アミノ - 酸リガーゼ（ペプチド合成酵素）は、真核細胞およびある種の細菌における細胞内タンパク質の分解の主経路であるユビキチン結合系（U C S : u b i q u i t i n c o n j u g a t i o n s y s t e m）に関連するユビキチンプロテアーゼ（u b i q u i t i n p r o t e a s e s）によって代表される。U C Sは、異常なタンパク質の除去を仲介し、遺伝子転写および細胞周期の進行などの細胞内のプロセスを制御する重要な調節タンパク質の半減期を調節する。U C S経路では、分解の標的となったタンパク質に小さな耐熱性タンパク質であるユビキチン（U b）が結合する。U bは、まずユビキチン活性化酵素（E 1）によって活性化され、次に幾つかのU b結合酵素（E 2）の内の1

10

20

30

40

50

つに輸送される。次にE2が、Ub分子をそのC末端グリシンを介して標的タンパク質の内部のリシン(受容体リシン)に結合する。次に、ユビキチン化されたタンパク質が、大きなマルチサブユニット蛋白分解酵素複合体であるプロテアソームによって認識されて分解される。ユビキチンは、ユビキチンプロテアーゼによる再利用のために放出される。UCSは、有糸分裂サイクルキナーゼ(mitotic cyclic kinases)、オンコプロテイン、p53などの腫瘍抑制遺伝子、ウイルス蛋白、シグナル伝達に関連する細胞表面受容体、転写調節因子、および変異した或いは損傷したタンパク質の分解に関与する(Ciechanover, A. (1994) Cell 79:13-21)。マウスプロトオンコジーンであるUnpは、核ユビキチンプロテアーゼをコードする。この核ユビキチンプロテアーゼが過剰に発現すると、NIH3T3細胞が癌化する。この遺伝子のヒト相同体のレベルは、小細胞癌および肺の腺癌において一貫して上昇している(Gray, D. A. (1995) Oncogene 10:2179-2183)。

10

【0057】

シクロリガーゼおよび他の炭素-窒素リガーゼは、プリンおよびピリミジンの生合成の新規の経路に関与する様々な酵素および酵素複合体を含む。これらの経路がRNAおよびDNAの双方の複製のためのヌクレオチドの合成に極めて重要であるため、これらの酵素の多くは、癌および感染症などの細胞増殖異常の治療のための臨床薬の標的である。

【0058】

プリンの生合成は、アミノ酸であるグリシンおよびグルタミン、並びに他の小分子から新たに起こる。このプロセスにおける3つの重要な反応は、グリシンアミド-リボヌクレオチドシンターゼ(GARS)、アミノイミダゾールリボヌクレオチドシンターゼ(AIRS)、およびグリシンアミドリボヌクレオチドトランスホルミラーゼ(GART)からなる3機能酵素によって触媒される。これら3つの酵素が共同で、ribosylamine phosphateにグリシンを結合することにより、アデニル酸ヌクレオチドおよびグアニル酸ヌクレオチド双方の前駆体であるホスホリボシルアミノイミダゾールが生成される。この3機能タンパク質は、ダウン症候群の病理学に関係する(Aimi, J. 他(1990) Nucleic Acid Res. 18:6665-6672)。アデニルコハク酸シンターゼは、ATP合成経路の重要なステップであるイノシン酸をアデニルコハク酸に変換するプリン生合成の後半のステップを触媒する。この酵素はまた、尿素回路に類似の反応を触媒する別の炭素-窒素リガーゼ、すなわちアルギニノコハク酸シンターゼに類似している(Powell, S. M. 他(1992) FEBS Lett. 303:4-10)。

20

30

【0059】

新規のプリンの合成と同様に、ピリミジンヌクレオチドであるuridylylateおよびcytidylylateの新規の合成も、この場合オロト酸およびホスホリボシルピロリン酸(PPRP)に由来するオロチジル酸(orotidylylate)ヌクレオチドである共通の前駆体から起こる。3つの炭素-窒素リガーゼを含む3機能酵素は、このプロセスにおいて重要な役割を演じる。この場合、酵素であるアスパラギン酸トランスカルバミラーゼ(ATCase)、カルバミルリン酸シンターゼII、およびジヒドロオロターゼ(DHOase)が、CADと呼ばれる1つの遺伝子によってコードされる。これら3つの酵素が共同で、ピリミジン生合成の開始反応物であるグルタミン、CO₂、およびATPを結合することにより、オロト酸およびorotidylylateの前駆体であるdihydroorotateが生成される(Iwahana, H. 他(1996) Biochem. Biophys. Res. Commun. 219:249-255)。更なるステップにより、ウリジンヌクレオチドがorotidylylateから合成される。シチジンヌクレオチドは、アミノ供与体としてのグルタミンおよび酵素CTPシンターゼを用いるUTPのアミド化により、ウリジン-5'-三リン酸(UTP)から得られる。ヒトCTPシンターゼにおける調節性変異により、癌の治療に広く用いられている薬剤に対する多薬剤耐性が付与されると考えられている(Yamauchi, M. 他(1990) EMBO J. 9:2095-2099)。

40

50

【0060】

炭素 - 炭素結合を形成するリガーゼには、カルボキシラーゼであるアセチルC o Aカルボキシラーゼおよびピルビン酸カルボキシラーゼが含まれる。アセチルC o Aカルボキシラーゼは、ATP加水分解のエネルギーを利用したCO₂とH₂OによるアセチルC o Aのカルボキシル化を触媒する。アセチルC o Aカルボキシラーゼは、長鎖脂肪酸の生合成における律速段階である。アセチルC o Aカルボキシラーゼの2つのアイソフォームであるI型およびII型は、組織特異的にヒトで発現される(Ha, J. 他 (1994) Eur. J. Biochem. 219: 297 - 306)。ピルビン酸カルボキシラーゼは核でコードされるミトコンドリア酵素であって、ピルビン酸のオキサロ酢酸への変換を触媒する。オキサロ酢酸は、クエン酸回路の重要な中間体である。

10

【0061】

リン酸エステル結合を形成するリガーゼには、DNAの複製および修復に関与するDNAリガーゼが含まれる。DNAリガーゼは、ATP加水分解によるエネルギーを用いて、あるヌクレオチドの遊離5'-リン酸をまず活性化させ、次にそれを隣接するヌクレオチドの3'-OH基と反応させて、DNA鎖における隣接するヌクレオチド間のリン酸ジエステル結合を形成する。この再接合反応は、新しくDNAを複製するプロセスにおいて一過的に形成される岡崎フラグメントと呼ばれる小さなDNA断片を連結するDNA複製、並びにDNA修復の両方で起こっている。DNA修復とは、DNAの複製または転写が起こる前に、酸化障害、加水分解攻撃、またはDNAの制御のきかないメチル化によって生成される偶然の塩基の変化を修正するプロセスのことである。ブルーム症候群は遺伝性のヒト疾患であって、DNA結合の部分的な欠損が起こり、それによって癌が発生し易くなる(Alberts, B. 他 (1994) The Molecular Biology of the Cell, Garland Publishing Inc., New York, NY, p. 247)。

20

【0062】

補助因子生合成酵素

補酵素および補欠分子族を含む補助因子は、酵素の反応に必要な低分子量の無機または有機化合物である。ある特定の補助因子、すなわちモリブデン含有補助因子であるmolybdopterinは、硝酸還元酵素、亜硫酸酸化酵素、キサンチン脱水素酵素、およびアルデヒド酸化酵素を含む様々な酵素の活性に必要である遍在性の分子である。

30

【0063】

molybdopterinの合成は、2つのステップの反応経路を介して行われる。第1に、グアノシン誘導体が前駆体Zと呼ばれる中間体に変換される。次にこの前駆体Zが、molybdopterinに変換される。ヒトでは、MOC S 1転写物が、第1のステップに関与するMOC S 1 AおよびMOC S 1 Bの2つの酵素をコードする。MOC S 2も1つの転写物であって、2つの重複する読み枠で第2のステップを触媒する酵素であるmolybdopterinシターゼの2つのサブユニットをコードする(Reiss, J. 他 (1999) Am. J. Hum. Genet. 64: 706 - 11)。アスペルギルス - ニデュランスでは、Cn x A B Cがmolybdopterin合成経路における第1のステップを触媒する。ヒトにおける場合と同様にアスペルギルス - ニデュランスの場合も、第2のステップはmolybdopterinシターゼによって触媒される。加えて、変換因子であるCn x Fも、アスペルギルス - ニデュランスにおいて見出された(Appleyard, M. 他 (1998) J. Biol. Chem. 273: 14869 - 14876)。Cn x Fは、このシターゼに硫黄原子を輸送してこのシターゼが前駆体Zにdithiolene基を付加することができるようにする酵素である大腸菌Moe Bに類似している。Cn x Fも、アスペルギルス - ニデュランスにおいて同様の反応を仲介すると考えられている。Cn x Fはまた、チアミン生合成に必要な酵素であるThi F、異質細胞(heterocyst)形成に関与するHes A、および真核生物ユビキチン活性化タンパク質E 1に類似している。しかしながら、Cn x Fとこれらのタンパク質の間には明確な生理学的な関連性が存在しない。

40

50

【0064】

molybdopterinの生合成障害により、molybdopterin依存性酵素活性が消失する。molybdopterin依存性酵素の欠損により、新生児発作、精神薄弱、および水晶体転移が引き起こされる。補助因子代謝物の欠損により生じるその他の疾患には、悪性貧血およびメチルマロン酸性尿症が含まれる。

【0065】

新規のヒト酵素分子、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、自己免疫/炎症性の疾患、遺伝性疾患、神経疾患、及び癌を含む細胞増殖異常の診断・治療・予防において有用であり、また、ヒト酵素分子の核酸配列及びアミノ酸配列の発現における外来性化合物の影響についての評価にも有用である。

10

【0066】

(発明の要約)

本発明は、総称して「HEM」、個別にはそれぞれ「HEM-1」、「HEM-2」、「HEM-3」、「HEM-4」、「HEM-5」、「HEM-6」、「HEM-7」、「HEM-8」、「HEM-9」、「HEM-10」、「HEM-11」、「HEM-12」、「HEM-13」、「HEM-14」、「HEM-15」、「HEM-16」、「HEM-17」、「HEM-18」、「HEM-19」、「HEM-20」、「HEM-21」、「HEM-22」、「HEM-23」、「HEM-24」、「HEM-25」、および「HEM-26」と呼ぶヒト酵素分子である精製されたポリペプチドを提供する。本発明の一実施態様では、(a) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1-26 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含む単離されたポリペプチドを提供する。別法では、SEQ ID NO: 1-26 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。

20

【0067】

更に本発明は、(a) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドをコードする単離されたポリヌクレオチドを提供する。別法では、このポリヌクレオチドは、SEQ ID NO: 1-26 からなる一群から選択されたポリペプチドをコードする。別法では、このポリヌクレオチドは、SEQ ID NO: 27-52 からなる一群から選択される。

30

【0068】

更に、本発明は、(a) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1-26 からなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドをコードするポリヌクレオチドと機能的に結合されたプロモーター配列を含む組換えポリヌクレオチドを提供する。別法では、本発明は、この組換えポリヌクレオチドで形質転換された細胞を提供する。更なる別法では、本発明は、この組換えポリヌクレオチドを含む遺伝子組換え生物を提供する。

40

【0069】

50

また、本発明は、(a) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 26 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドの生産方法を提供する。この方法は、(a) このポリペプチドの発現に好適な条件下で、このポリペプチドをコードするポリヌクレオチドと機能的に結合されたプロモーター配列を含む組換えポリヌクレオチドで形質転換された細胞を培養するステップと、(b) このように発現したポリペプチドを回収するステップとを含む。

10

【0070】

更に、本発明は、(a) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 26 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドに特異的に結合する単離された抗体を提供する。

【0071】

更に、本発明は、(a) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と、(b) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と90%以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記(a)に相補的なポリヌクレオチド配列と、(d) 前記(b)に相補的なポリヌクレオチド配列と、(e) 前記(a)乃至(d)のRNA等化物とで構成される一群から選択されたポリヌクレオチド配列を含む単離されたポリヌクレオチドを提供する。別法では、このポリヌクレオチドは、少なくとも60個の連続するヌクレオチドを含む。

20

【0072】

更に本発明は、(a) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と、(b) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と90%以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記(a)に相補的なポリヌクレオチド配列と、(d) 前記(b)に相補的なポリヌクレオチド配列と、(e) 前記(a)乃至(d)のRNA等化物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチド配列を有する標的ポリヌクレオチドをサンプルにおいて検出する方法を提供する。この方法は、(a) 前記サンプル内の標的ポリヌクレオチドと相補的な配列を構成する少なくとも20個の連続するヌクレオチドを含むプローブと前記サンプルをハイブリダイズさせるステップであって、前記プローブと前記標的ポリヌクレオチドまたはその断片とでハイブリダイゼーション複合体が形成される条件下で、前記プローブが前記標的ポリヌクレオチドに特異的にハイブリダイズする、該ステップと、(b) 前記ハイブリダイゼーション複合体の存在するか否かを検出し、存在する場合には随意選択でその収量を測定するステップとを含む。別法では、前記

30

40

【0073】

更に本発明は、(a) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と、(b) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と90%以上の配列同一性を有する天然のポリヌクレオチド配列と、(c) 前記(a)に相補的なポリヌクレオチド配列と、(d) 前記(b)に相補的なポリヌクレオチド配列と、(e) 前記(a)乃至(d)のRNA等化物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチド配列を有する標的ポリヌクレオチドをサンプルにおいて検出する方法を提供する。この方法は、(a) ポリメラーゼ連鎖反応増幅を用いて、前記標的ポリヌクレオチドまたはその断片を増幅するステップと、

50

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

【0074】

更に本発明は、(a) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 26 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含む効果的な量のポリペプチド及び好適な医薬用賦形剤を含む組成物を提供する。一実施例では、SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列を含む組成物を提供する。更に、本発明は、患者にこの組成物を投与することを含む、機能的HEMの発現の低下に関連した疾患やその症状の治療方法を提供する。

10

【0075】

更に本発明は、(a) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 26 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドのアゴニストとして効果的な化合物をスクリーニングする方法を提供する。この方法は、(a) このポリペプチドを含むサンプルを化合物に曝露するステップと、(b) このサンプルのアゴニスト活性を検出するステップとを含む。別法では、本発明は、この方法によって同定されたアゴニスト化合物と好適な医薬用賦形剤とを含む組成物を提供する。更なる別法では、本発明は、この組成物の患者への投与を含む、機能的HEM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。

20

【0076】

更に、本発明は、(a) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 26 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドのアンタゴニストとして効果的な化合物をスクリーニングする方法を提供する。この方法は、(a) このポリペプチドを含むサンプルを化合物に曝露するステップと、(b) このサンプルのアンタゴニスト活性を検出するステップとを含む。別法では、本発明は、この方法によって同定されたアンタゴニスト化合物と好適な医薬用賦形剤とを含む組成物を提供する。更なる別法では、本発明は、この組成物の患者への投与を含む、機能的HEMの過剰な発現に関連した疾患やその症状の治療方法を提供する。

30

【0077】

更に本発明は、(a) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 26 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドに特異的に結合する化合物をスクリーニングする方法を提供する。この方法は、(a) このポリペプチドを好適な条件下で少なくとも1つの化合物と結合させるステップと、(b) このポリペプチドとこの試験化合物との結合を検出して、このポリペプチドと特異的に結合する化合物を同定するステップとを含む。

40

50

【0078】

更に本発明は、(a) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と、(b) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列と90%以上の配列同一性を有する天然のアミノ酸配列と、(c) SEQ ID NO: 1 - 26 からなる一群から選択されたアミノ酸配列の生物学的に活性な断片と、(d) SEQ ID NO: 1 - 26 とからなる一群から選択されたアミノ酸配列の免疫原性断片とで構成される一群から選択されたアミノ酸配列を含むポリペプチドの活性を調節する化合物をスクリーニングする方法を提供する。このスクリーニング方法は、(a) このポリペプチドを、その活性が許容される条件下で少なくとも1つの化合物と結合させるステップと、(b) この試験化合物の存在下でのこのポリペプチドの活性を評価するステップと、(c) この試験化合物の存在下でのこのポリペプチドの活性と、この試験化合物の不在下でのこのポリペプチドの活性とを比較するステップとを含み、この試験化合物の存在下でのこのポリペプチドの活性の変化が、このポリペプチドの活性を調節する化合物の存在を示唆するという特徴を有する。

10

【0079】

更に本発明は、SEQ ID NO: 27 - 52 からなる一群から選択された配列を含む標的ポリヌクレオチドの発現を変化させるのに効果的な化合物をスクリーニングする方法であって、(a) この標的ポリヌクレオチドを含むサンプルを化合物に曝露するステップと、(b) この標的ポリヌクレオチドの発現の変化を検出するステップとを含む、該スクリーニング方法を提供する。

20

【0080】

本発明はさらに、試験化合物の毒性を評価する方法を提供する。この方法は、(a) 核酸を含む生体サンプルを前記試験化合物で処理するステップと、(b) 処理した前記生体サンプルの核酸をプローブとハイブリダイズするステップと、(c) ハイブリダイゼーション複合体の収量を測定するステップと、(d) 前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量を、未処理の生体サンプルにおけるハイブリダイゼーション複合体の収量とを比較するステップとを含み、前記処理した生体サンプルにおけるハイブリダイゼーション複合体の収量の差異が試験化合物の毒性を示唆する。この方法における前記プローブは、(1) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と、(2) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と少なくとも90%の配列同一性を有する天然のポリヌクレオチド配列と、(3) 前記(1)に相補的なポリヌクレオチド配列と、(4) 前記(2)に相補的なポリヌクレオチド配列と、(5) 前記(1)乃至(4)のRNA等価物とで構成される一群から選択されたポリヌクレオチド配列を含むポリヌクレオチドの連続する少なくとも20個のヌクレオチドを含む。また、前記ハイブリダイゼーションは、前記プローブと前記生体サンプルの標的ポリヌクレオチドとの間で特異的なハイブリダイゼーション複合体が形成される条件下で行われる。また、前記標的ポリヌクレオチドが、(1) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と、(2) SEQ ID NO: 27 - 52 からなる一群から選択されたポリヌクレオチド配列と少なくとも90%の配列同一性を有する天然のポリヌクレオチド配列と、(3) 前記(1)に相補的なポリヌクレオチド配列と、(4) 前記(2)に相補的なポリヌクレオチド配列と、(5) 前記(1)乃至(5)のRNA等価物とを含む。代替的に前記標的ポリヌクレオチドは前記ポリヌクレオチド配列の断片である。

30

40

【0081】

(本発明の記載について)

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

50

【0082】

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

【0083】

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

10

【0084】

(定義)

用語「HEM」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HEMのアミノ酸配列を指す。

【0085】

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

20

【0086】

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

30

【0087】

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

40

【0088】

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。「

50

「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。

【0089】

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

【0090】

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

10

【0091】

用語「抗体」は、抗原決定基と結合可能なFab及びF(ab')₂、及びそれらの断片、Fv断片などの無傷の分子を指す。HEMポリペプチドと結合する抗体は、抗体を免疫する小ペプチドを含む無傷の分子またはその断片を用いて作製可能である。動物（例えば、マウス、ラット、若しくはウサギ）を免疫化するのに使用されるポリペプチド或いはオリゴペプチドは、RNAの翻訳から、或いは化学的に合成可能であり、必要に応じて担体タンパク質と結合させることも可能である。ペプチドと化学的に結合した一般に用いられる担体は、ウシ血清アルブミン、チログロブリン、及びキーホールリンペットヘモニアン（KLH）を含む。次ぎに、この結合したペプチドを用いて動物を免疫化する。

20

【0092】

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

【0093】

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

30

【0094】

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

40

【0095】

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

【0096】

「所定のポリヌクレオチド配列を含む組成物」または「所定のアミノ酸配列を含む組成物」は広い意味で、所定のヌクレオチド配列若しくはアミノ酸配列を含む任意の組成物を指

50

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

【0097】

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

10

【0098】

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

20

元の残基

保存的な置換

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

30

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

【0099】

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。

【0100】

用語「誘導体」は、化学修飾されたポリヌクレオチドまたはポリペプチドを指す。ポリヌクレオチド配列の化学修飾には、例えば、アルキル基、アシル基、ヒドロキシル基、或い

50

はアミノ基による水素の置換がある。誘導体ポリヌクレオチドは、自然分子（未修飾の分子）の生物学的或いは免疫学的機能の少なくとも1つを維持するポリペプチドをコードする。誘導体ポリペプチドとは、もとのポリペプチドの生物学的機能、或いは免疫学的機能の少なくとも1つを維持する、グリコシル化、ポリエチレングリコール化、或いは任意の同様のプロセスによって修飾されたポリペプチドのことである。

【0101】

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

【0102】

用語「断片」は、HEMまたはHEMをコードするポリヌクレオチドの固有の部分であって、その親配列（parent sequence）と同一であるがその配列より長さが短いものを指す。「断片」の最大の長さは、親配列から1つのヌクレオチド/アミノ酸残基を差し引いた長さである。例えば、ある断片は、5～1000個の連続するヌクレオチド或いはアミノ酸残基を含む。プローブ、プライマー、抗原、治療用分子、またはその他の目的に用いる断片は、少なくとも5、10、15、16、20、25、30、40、50、60、75、100、150、250若しくは500個の連続するヌクレオチド或いはアミノ酸残基の長さである。断片は、優先的に分子の特定の領域から選択される場合もある。例えば、ポリペプチド断片は、所定の配列に示された最初の250若しくは500のアミノ酸（或いは、ポリペプチドの最初の25%または50%）から選択された連続するアミノ酸の所定の長さを含み得る。これらの長さは一例であり、配列表及び表、図面を含む明細書に記載の任意の長さが、本発明の実施例に含まれ得る。

【0103】

SEQ ID NO: 27-52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: 27-52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: 27-52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: 27-52を関連ポリヌクレオチド配列から区別する類似の方法に有用である。ある断片と一致するSEQ ID NO: 27-52の正確な断片の長さや領域は、その断片の目的に基づいて当分野で一般的な技術によって日常的に測定できる。

【0104】

SEQ ID NO: 1-26のある断片は、SEQ ID NO: 27-52のある断片によってコードされる。SEQ ID NO: 1-26のある断片は、SEQ ID NO: 1-26を特異的に同定する固有のアミノ酸配列の領域を含む。例えば、SEQ ID NO: 1-26のある断片は、SEQ ID NO: 1-26を特異的に認識する抗体の開発における免疫原性ペプチドとして有用である。ある断片と一致するSEQ ID NO: 1-26の正確な断片の長さや領域は、その断片の目的に基づいて当分野で一般的な技術によって日常的に決定できる。

【0105】

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

【0106】

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

【0107】

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

10

20

30

40

50

【0108】

ポリヌクレオチド配列間の同一性のパーセントは、MEGALIGN version 3.12e 配列アラインメントプログラムに組込まれるCLUSTAL V アルゴリズムのデフォルトパラメータを用いて決定可能である。このプログラムはLASERGENE ソフトウェアパッケージの一部であり、分子生物学分析プログラム一式 (DNASTAR, Madison WI) である。このCLUSTAL V は、Higgins, D. G. 及び P. M. Sharp (1989) CABIOS 5:151-153、Higgins, D. G. 他 (1992) CABIOS 8:189-191 に記載されている。ポリヌクレオチド配列の対のアライメントの場合、デフォルトパラメータは、Ktuple=2、gap penalty=5、window=4、「diagonals saved」=4 と設定する。「重み付けされた」残基重み付け表が、デフォルトとして選択された。同一性のパーセントは、アラインメントされたポリヌクレオチド配列間の「類似性のパーセント」としてCLUSTAL V によって報告される。

【0109】

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

【0110】

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

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

【0111】

高い同一性を示さない核酸配列でも、遺伝子コードの縮重によって類似のアミノ酸配列を

コードし得る。縮重を利用して核酸配列を変え、それぞれが実質的に同じタンパク質をコードする様々な核酸配列を作製できることを理解されたい。

【0112】

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

【0113】

ポリペプチド配列間の同一性のパーセントは、MEGALIGNバージョン3.12e配列アラインメントプログラム（上記）に組込まれるCLUSTAL Vアルゴリズムのデフォルトパラメータを用いて決定可能である。CLUSTAL Vを用いる対方式のポリペプチド配列のアラインメントの場合、デフォルトパラメータは、Ktuple = 1、gap penalty = 3、window = 5、及び「diagonals saved」= 5と設定する。PAM250マトリクスが、デフォルトの残基重み付け表として選択される。ポリヌクレオチドアラインメントと同様に、アラインメントされたポリペプチド配列の対の同一性のパーセントは、「類似性のパーセント」としてCLUSTAL Vによって報告される。

【0114】

別法では、NCBI BLASTソフトウェア一式が用いられる。例えば、2つのポリペプチド配列を対で比較をする場合、ある者は、デフォルトパラメータで設定された「BLAST 2 Sequences」ツールVersion 2.0.12（Apr-21-2000）でblastpを使用するであろう。そのようなデフォルトパラメータは、例えば、以下のようにする。

【0115】

Matrix : BLOSUM62
 Open Gap : 11 及び Extension Gap : 1 penalties
 Gap x drop-off : 50
 Expect : 10
 Word Size : 3
 Filter : on

同一性のパーセントは、例えば、特定の配列番号で決められた、所定のポリペプチド配列の全長に対して測定してもよいし、それより短い長さに対して、例えば、ある大きな所定のポリペプチド配列から得られた断片、例えば、連続する少なくとも15、20または30、40、50、70、150の残基の断片の長さに対して測定してもよい。このような長さは単なる例であり、配列表及び表、図面を含む明細書に記載の配列の任意の長さの断片を用いて、同一性のパーセントが測定される長さを示すことができる。

【0116】

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

【0117】

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。

【0118】

「ハイブリダイゼーション」とは、所定のハイブリダイゼーション条件下で、ある一本鎖ポリヌクレオチドがある相補的な一本鎖と塩基対を形成するアニーリングのプロセスである。特異的なハイブリダイゼーションとは、2つの核酸配列が高い相同性を有することを意味する。アニーリングが許容される条件下で、特異的なハイブリダイゼーション複合体

10

20

30

40

50

が形成され、洗浄過程の後もハイブリダイズしたままである。洗浄過程は、ハイブリダイゼーションプロセスの厳密性即ちストリンジェンシー (stringency) の決定において特に重要であり、よりストリンジェントな条件では、非特異的な結合、即ち完全には一致しない核酸鎖間の対の結合が減少する。核酸配列間のアニーリングが許容される条件は、当業者によって日常的に決定され、ハイブリダイゼーションの間は一定であるが、洗浄過程は、目的のストリンジェンシーにするためにその最中に条件の変更が可能であり、ハイブリダイゼーション特異性が得られる。アニーリングが許容される条件は、例えば、温度が68 で、約6 x SSC、約1% (w/v) のSDS、並びに約100 µg/mlのせん断して変性したサケ精子DNAが含まれる。

【0119】

一般に、ハイブリダイゼーションのストリンジェンシーは、洗浄過程を行う際の温度によっても左右される。この洗浄温度は通常、所定のイオン強度とpHにおける特定の配列の熱融点 (T_m) より約5 ~ 20 低く選択される。このT_mは、(所定のイオン強度とpHの下) 標的の配列の50%が完全に一致するプローブとハイブリダイズする温度である。T_mを計算する式及び核酸のハイブリダイゼーションの条件は、周知であり、Sambrook, J. 他による, 1989, Molecular Cloning: A Laboratory Manual, 第2版の1-3巻, Cold Spring Harbor Press, Plainview NY; 特に2巻の9章に記載されている。

【0120】

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

【0121】

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

【0122】

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。

【0123】

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

【0124】

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

10

20

30

40

50

【0125】

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

【0126】

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。

【0127】

用語「調節」は、HEMの活性の変化を指す。例えば、調節によって、HEM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。

10

【0128】

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

【0129】

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

20

【0130】

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

【0131】

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

30

【0132】

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

40

【0133】

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

50

【0134】

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

【0135】

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

【0136】

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

うな組換え核酸は、例えば、ある細胞を形質転換するのに用いられるベクターの一部であり得る。

【0137】

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

【0138】

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

10

【0139】

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

【0140】

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

20

【0141】

用語「サンプル」は、その最も広い意味で用いられている。HEM、HEM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。

【0142】

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

30

【0143】

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

【0144】

「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。

40

【0145】

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

【0146】

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

50

【0147】

「形質転換」とは、外来DNAが受容細胞に導入されるプロセスのことである。形質転換は、当分野で周知の種々の方法により、自然或いは人工の条件下で起こり、原核宿主細胞若しくは真核宿主細胞の中に外来核酸配列を挿入する任意の周知の方法によって行うことができる。この形質転換の方法は、形質転換される宿主細胞のタイプによって選択される。この方法には、バクテリオファージまたはウイルス感染、電気穿孔法（エレクトロポレーション）、リポフェクション、及び微粒子照射が含まれるが、これらに限定されるものではない。「形質転換された」細胞には、導入されたDNAが自律的に複製するプラスミドとして或いは宿主染色体の一部として複製可能である安定的に形質転換された細胞が含まれる。さらに、限られた時間に一時的に導入DNA若しくは導入RNAを発現する細胞も含まれる。 10

【0148】

本明細書における「遺伝子組換え生物」とは、当分野で周知の遺伝子組換え技術などを用いて、人間が生物の1つ以上の細胞に異種の核酸を導入した任意の生物であり、動物及び植物を含むが、それらに限定されるものではない。微量注入や組換えウイルスに感染させるなどの慎重な遺伝子操作によって、細胞の前駆体に直接或いは間接的に異種核酸を細胞に導入する。「遺伝子操作」とは、典型的な交雑育種や*in vitro*での受精ではなく、組換えDNA分子を導入することである。本発明に従った遺伝子組換え生物には、細菌及びラン藻類、菌類、植物、動物が含まれる。本発明の単離されたDNAは、当分野で周知の、例えば、感染、形質移入、形質転換、トランス接合（*trans conjugation*）などの方法によって、宿主に導入することができる。本発明のDNAをそのような生物に導入する技術は周知であり、前出のSambrook他（1989）に記載されている。 20

【0149】

特定の核酸配列の「変異配列」とは、デフォルトパラメータ設定の「BLAST 2 Sequences」ツールVersion 2.0.9（May-07-1999）を用いるblastnによって、ある核酸配列のある長さに対する該特定の核酸配列の同一性が、少なくとも40%と決定された核酸配列のことである。このような核酸の対は、ある長さにおいて、例えば、少なくとも50%、60%、70%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或いはそれ 30
以上の同一性を示し得る。ある変異配列は、例えば、「アレル」変異配列（上述）または「スプライス」変異配列、「種」変異配列、「多型」変異配列と表すことができる。スプライス変異配列は基準分子と同一性が極めて高い可能性があるが、mRNAプロセッシング中のエキソンの択一的スプライシングによってポリヌクレオチドの数が多くなったり、少なくなったりする。対応するポリペプチドは、基準分子に存在する追加の機能ドメインを有したり、基準分子に存在するドメインが欠落したりし得る。種変異配列は、種によって異なるポリヌクレオチド配列である。得られるポリペプチドは、互いに高いアミノ酸同一性を有する。多型変異配列は、所定の種と種における特定の遺伝子のポリヌクレオチド配列が異なる。多型変異配列はまた、ポリヌクレオチド配列の1つのヌクレオチドが異なる「1ヌクレオチド多型」（SNP）も含み得る。SNPの存在は、例えば、或る集団、 40
病態、病態の性向を示唆し得る。

【0150】

特定のポリペプチド配列の「変異体」とは、デフォルトパラメータ設定の「BLAST 2 Sequences」ツールVersion 2.0.9（May-07-1999）を用いるblastpによって、あるポリペプチド配列のある長さに対する該特定のポリペプチド配列の同一性が、少なくとも40%と決定されたポリペプチド配列のことである。このようなポリペプチドの対は、ある長さにおいて、例えば、少なくとも50%、60%、70%、80%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、或いはそれ以上の同一性を示し得る。

【0151】

(発 明)

本発明は、新規のヒト酵素分子 (H E M) 及び H E M をコードするポリヌクレオチドの発見に基づき、これらの組成物を利用した自己免疫 / 炎症性の疾患、遺伝性疾患、神経疾患、及び癌を含む細胞増殖異常の診断、治療、及び予防に関する。

【 0 1 5 2 】

表 1 は、本発明のポリヌクレオチド配列及びポリペプチド配列の識別番号を示す。各ポリヌクレオチドおよびそれに対応するポリペプチドは、1つのインサイトプロジェクト識別番号 (I n c y t e P r o j e c t I D) に関連する。各ポリペプチド配列は、記載されているようにポリペプチド配列識別番号 (P o l y p e p t i d e S E Q I D N O :) およびインサイトポリペプチド配列番号 (I n c y t e P o l y p e p t i d e I D) の両方によって示されている。各ポリヌクレオチド配列は、記載されているようにポリヌクレオチド配列識別番号 (P o l y p e p t i d e S E Q I D N O :) およびインサイトポリヌクレオチドコンセンサス配列番号 (I n c y t e P o l y p e p t i d e I D) の両方によって示されている。

【 0 1 5 3 】

表 2 は、G e n B a n k タンパク質 (g e n e p t) データベースにおいて B L A S T 解析により同定された本発明のポリペプチドに相同性を有する配列を示す。列 1 および列 2 はそれぞれ、本発明の各ポリペプチドに対するポリペプチド配列識別番号 (P o l y p e p t i d e S E Q I D N O :) およびそれに対応するインサイトポリペプチド配列番号 (I n c y t e P o l y p e p t i d e I D) を示す。列 3 は、G e n B a n k の最も近い相同体の G e n B a n k の識別番号 (G e n b a n k I D N O :) を示す。列 4 は、各ポリペプチドとその G e n B a n k 相同体との間の一致を表す確率スコアを示す。列 5 は、G w n B a n k 相同体のアノテーションを示し、更に該当箇所には適当な引用文も示す。これらを引用することを以って本明細書の一部とする。

【 0 1 5 4 】

表 3 は、本発明の各ポリペプチドの様々な構造的特徴を示す。列 1 および列 2 はそれぞれ、本発明の各ポリペプチドのポリペプチド配列識別番号 (S E Q I D N O :) およびそれに対応するインサイトポリペプチド配列番号 (I n c y t e P o l y p e p t i d e I D) 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、G C G 配列分析ソフトウェアパッケージの M O T I F S プログラム (G e n e t i c s C o m p u t e r G r o u p , M a d i s o n W I) によって決定された、潜在的なリン酸化部位および潜在的なグリコシル化部位を示す。列 6 は、シグネチャ (s i g n a t u r e) 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。

【 0 1 5 5 】

表 2 および表 3 は共に、本発明の各ポリペプチドの特性を一覧にしたものであって、これらの特性は請求するポリペプチドが酵素分子であることを立証するものである。例えば、S E Q I D N O : 5 は、その M 1 残基から E 3 8 4 残基において、ドブネズミ アラニンシンターゼ (G e n B a n k I D g 2 0 3 1 0 6) と 8 4 % の同一性を有することが B a s i c L o c a l A l i g n m e n t S e a r c h T o o l (B L A S T) によって示された (表 2 を参照) 。 B L A S T の確率スコアは 0 . 0 であり、これは探しているポリペプチド配列アラインメントが偶然の一致により得られる確率を示す。S E Q I D N O : 5 はまた、炭素 - 窒素加水分解酵素ドメインを有する。これは、隠れマルコフモデル (H M M) 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 P F A M 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 を参照) 。 B L I M P S 解析、更に B L A S T 解析から得られたデータによって、S E Q I D N O : 5 が加水分解酵素であることが裏付けられた。S E Q I D N O : 1、S E Q I D N O : 2、S E Q I D N O : 3、S E Q I D N O : 4、S E Q I D N O : 6、S E Q I D N O : 7、S E Q I D N O : 8、S E Q I D N O : 9、S E Q I D N O : 1 0、S E

Q I D N O : 1 1、S E Q I D N O : 1 2、S E Q I D N O : 1 3、S E Q I D N O : 1 4、S E Q I D N O : 1 5、S E Q I D N O : 1 6、S E Q I D N O : 1 7、S E Q I D N O : 1 8、S E Q I D N O : 1 9、S E Q I D N O : 2 0、S E Q I D N O : 2 1、S E Q I D N O : 2 2、S E Q I D N O : 2 3、S E Q I D N O : 2 4、S E Q I D N O : 2 5、および S E Q I D N O : 2 6 は同様の方法で解析してアノテーションを付けた。S E Q I D N O : 1 - 2 6 の解析のためのアルゴリズムおよびパラメータを表 7 に記載する。

【 0 1 5 6 】

表 4 に示されているように、本発明の完全長ポリヌクレオチド配列は、c D N A 配列、またはゲノム D N A 由来のコード (エキソン) 配列、或いはこれらの 2 種類の配列のあらゆる組み合わせを用いて組み立てた。列 1 および列 2 はそれぞれ、本発明の各ポリヌクレオチドのポリヌクレオチド配列識別番号 (P o l y n u c l e o t i d e S E Q I D N O :) およびそれに対応するインサイトポリヌクレオチドコンセンサス配列番号 (I n c y t e P o l y n u c l e o t i d e I D) を示す。列 3 は、塩基対における各ポリヌクレオチド配列の長さを示す。列 4 は、例えば、S E Q I D N O : 2 7 - 5 2 を同定するため、或いは S E Q I D N O : 2 7 - 5 2 と関連するポリヌクレオチド配列とを区別するためのハイブリダイゼーションまたは増幅技術に有用なポリヌクレオチド配列の断片を示す。列 5 は、c D N A 配列、ゲノム D N A から推定されるコード配列 (エキソン)、および / または c D N A およびゲノム D N A の両方からなる群に対応する識別番号を示す。これらの配列を用いて本発明の完全長ポリヌクレオチド配列を組み立てた。表 4 の列 6 および列 7 はそれぞれ、列 5 の配列に対応する c D N A 配列の開始ヌクレオチド (5 ') 位置および終了ヌクレオチド (3 ') 位置を示す。

【 0 1 5 7 】

表 4 の列 5 に示されている識別番号は、具体的には、例えばインサイト c D N A およびそれらに対応する c D N A ライブラリの識別番号を示す。例えば、3 5 4 0 0 8 4 H 1 はインサイト c D N A 配列の識別番号であり、S E M V N O T 0 4 はそれが由来する c D N A ライブラリである。c D N A ライブラリが示されていないインサイト c D N A は、プールされている c D N A ライブラリ (例えば、S X A D 9 0 0 8 3 V 1) に由来する。または、列 5 の識別番号は、ポリヌクレオチド配列の組み立てに用いた G e n B a n k の c D N A すなわち E S T (例えば、g 2 7 9 5 8 8 8) の識別番号の場合もある。または、列 5 の識別番号は、G e n s c a n 分析によって推定されるゲノム D N A のコード領域の場合もある。この G e n s c a n 推定コード配列は、配列を組み立てる前に編集する場合がある (実施例 4 を参照)。または、列 5 の識別番号は、“ e x o n - s t i t c h i n g ” アルゴリズムによって c D N A および G e n s c a n 推定エキソンの両方からなる群の場合もある (実施例 5 を参照)。または、列 5 の識別番号は、“ e x o n - s t r e t c h i n g ” アルゴリズムによって c D N A および G e n s c a n 推定エキソンの両方からなる群の場合もある (実施例 5 を参照)。場合によっては、列 5 に示されている配列の範囲と重複するインサイト c D N A の範囲が得られ、最終的なコンセンサス配列が決定されるが、それに相当するインサイト c D N A の識別番号は示されていない。

【 0 1 5 8 】

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

【 0 1 5 9 】

本発明はまた、H E M の変異体も含む。好適な H E M の変異体は、H E M の機能的或いは構造的特徴の少なくともどちらか一方を有し、かつ H E M アミノ酸配列に対して少なくとも約 8 0 % のアミノ酸配列同一性、或いは少なくとも約 9 0 % のアミノ酸配列同一性、更には少なくとも約 9 5 % のアミノ酸配列同一性を有する。

10

20

30

40

50

【0160】

本発明はまた、HEMをコードするポリヌクレオチドを提供する。特定の実施例において、本発明は、HEMをコードするSEQ ID NO: 27-52からなる一群から選択された配列を含むポリヌクレオチド配列を提供する。配列表に示したSEQ ID NO: 27-52のポリヌクレオチド配列は、窒素系塩基のチミンがウラシルに置換され、糖鎖の背骨がデオキシリボースではなくリボースからなる等価RNA配列を含む。

【0161】

本発明はまた、HEMをコードするポリヌクレオチド配列の変異配列を含む。詳細には、このようなポリヌクレオチド配列の変異配列は、HEMをコードするポリヌクレオチド配列と少なくとも70%のポリヌクレオチド配列同一性、或いは少なくとも85%のポリヌクレオチド配列同一性、更には少なくとも95%ものポリヌクレオチド配列同一性を有する。本発明の特定の実施形態は、SEQ ID NO: 27-52からなる一群から選択された核酸配列と少なくとも70%のポリヌクレオチド配列同一性、或いは少なくとも85%のポリヌクレオチド配列同一性、更には少なくとも95%ものポリヌクレオチド配列同一性を有するSEQ ID NO: 27-52からなる一群から選択された配列を含むポリヌクレオチド配列の変異配列を提供する。上記したポリヌクレオチド変異配列は何れも、HEM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。

【0162】

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

【0163】

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

【0164】

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

【0165】

更に本発明には、種々のストリンジェントな条件下で、請求項に記載されたポリヌクレオチド配列、特に、SEQ ID NO: 27-52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。

【0166】

10

20

30

40

50

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

【0167】

当分野で周知のPCR法をベースにした種々の方法で、部分的なヌクレオチド配列を利用して、HEMをコードする核酸配列を伸長し、プロモーターや調節エレメントなどの上流にある配列を検出する。例えば制限部位PCR法を利用する1つの方法では、一般的なプライマー及び入れ子プライマー (nested primer) を用いてクローニングベクター内のゲノムDNAから未知の配列を増幅する (例えば、Sarkar, G. (1993) PCR Methods Applic 2: 318 - 322 を参照)。逆PCR法を用いる別法では、広範な方向に伸長して環状化した鋳型から未知の配列を増幅するプライマーを用いる。この鋳型は、既知のゲノム遺伝子座及びその周辺の配列を含む制限断片に由来する (例えば、Triglia, T.ら (1988) Nucleic Acids Res 16: 8186 を参照)。キャプチャPCR法を用いる第3の方法は、ヒト及び酵母菌人工染色体DNAの既知の配列に隣接するDNA断片のPCR増幅を含む (例えば、Lagerstrom, M.他 (1991) PCR Methods Applic 1: 111 - 119 を参照)。この方法では、多数の制限酵素による消化及びライゲーションを用いて、PCRを行う前に未知の配列の領域の中に組換え二本鎖配列を挿入することが可能である。また、当分野で周知の別の方法を用いて未知の配列を得ることも可能である。 (例えば、Parker, J.D. 他 (1991) Nucleic Acids Res. 19: 3055 - 3060 を参照)。更に、PCR、ネスト化プライマー、PROMOTERFINDERライブラリ (Clontech, Palo Alto CA) を用いれば、ゲノムDNA内の歩行が可能である。この方法ではライブラリをスクリーニングする必要がなく、イントロン/エキソン接合部を探すのに有用である。全てのPCR法をベースにした方法では、プライマーは、市販のOLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) 或いは別の好適なプログラムなどを用いて、長さが22~30ヌクレオチド、GC含量が50%以上、約68~72の温度で鋳型に対してアニーリングするように設計される。

【0168】

完全長のcDNAをスクリーニングする場合は、大きなcDNAを含むようにサイズが選択されたライブラリを用いるのが好ましい。更に、オリゴd(T)ライブラリが完全な長さのcDNAを産生できない場合は、遺伝子の5'領域を有する配列を含むものが多いラ

ンダムに初回抗原刺激を受けたライブラリが有用である。ゲノムライブラリは、5'非転写調節領域への配列の伸長に有用であろう。

【0169】

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

10

【0170】

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

【0171】

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

20

【0172】

本発明のヌクレオチドを、MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; 米国特許第5,837,458号; Chang, C.-C. 他(1999) Nat. Biotechnol. 17:793-797; Christians, F.C. 他(1999) Nat. Biotechnol. 17:259-264; Cramer, A. 他(1996) Nat. Biotechnol. 14:315-319)などのDNAシャフリング技術を用いてシャフリングして、HEM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HEMの生物学的特性を変更或いは改良することができる。DNAシャフリングは、PCR法による遺伝子断片の組換えで遺伝子変異体のライブラリを作製するプロセスである。次に、このライブラリを、目的の特性を有する遺伝子変異体を同定するために選択或いはスクリーニングする。これらの好ましい変異体をプールし、DNAシャフリング及び選択/スクリーニングを繰り返す。従って、人工的な育種及び急速な分子の進化によって多様な遺伝子が作られる。例えば、ランダムな位置に変異がある1つの遺伝子の断片を、目的の特性が最適化するまで、組換え及びスクリーニング、シャフリングを実施することもできる。別法では、所定の遺伝子の断片を、同じ或いは異なる種の間で遺伝子ファミリーの相同な遺伝子の断片で組換え、それによってプロトコルに従った調節可能な方法で、多数の天然遺伝子の遺伝子多様性を最大にすることができる。

30

40

【0173】

別の実施例によれば、HEMをコードする配列は、当分野で周知の化学的方法を用いて、全体或いは一部が合成可能である(例えば、Caruthers, M.H.ら(1980) Nucl. Acids Res. Symp. Ser 7:215-223; 及びH

50

orn, T. 他 (1980) Nucl. Acids Res. Symp. Ser. 2 25-232 を参照)。別法として、化学的方法を用いてHEM自体またはその断片を合成することが可能である。例えば、ペプチド合成は種々の固相技術を用いて実行可能である(例えば、Creighton, T. (1984) Proteins. Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J. Y. ら (1995) Science 269:202-204 を参照)。また、合成の自動化は例えばABI 431Aペプチドシンセサイザー(PE Biosystems)を用いて達成し得る。更にHEMのアミノ酸配列または任意のその一部は、直接的な合成の際の変更、及び/または化学的方法を用いた他のタンパク質または任意のその一部からの配列との組み合わせにより、天然のポリペプチド配列を有するポリペプチドまたは変異体ポリペプチドを作製することが可能である。

10

【0174】

このペプチドは、分離用高速液体クロマトグラフィー(例えば、Chiez, R. M. 及び F. Z. Regnier (1990) Methods Enzymol. 182:392-421 を参照)を用いて実質的に精製可能である。合成されたペプチドの組成は、アミノ酸分析或いはシーケンシングにより確認することができる(例えば、Creighton、前出、pp28-53 を参照)。

【0175】

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

20

30

【0176】

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

40

【0177】

種々の発現ベクター/宿主系を利用して、HEMをコードする配列の保持及び発現が可能である。これらには、限定するものではないが、組換えバクテリオファージ、プラスミド、またはコスミドDNA発現ベクターで形質転換された細菌などの微生物や、酵母菌発現

50

ベクターで形質転換された酵母菌や、ウイルス発現ベクター（例えば、バキュロウイルス）に感染した昆虫細胞系や、ウイルス発現ベクター（例えば、カリフラワーモザイクウイルス、CaMV；タバコモザイクウイルス、TMV）または細菌発現ベクター（例えば、TiまたはpBR322プラスミド）で形質転換された植物細胞系や、動物細胞系などが含まれる（例えば、前出のSambrook、前出のAusubel、Van Heeke, G. および S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509、Engelhard, E.K. 他 (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227、Sandig, V. 他 (1996) *Hum. Gene Ther.* 7:1937-1945、Takamatsu, N. (1987) *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、Harrington, J.J. 他 (1997) *Nat. Genet.* 15:345-355を参照)。レトロウイルス、アデノウイルス、ヘルペスウイルスまたはワクシニアウイルス由来の発現ベクター、または種々の細菌性プラスミド由来の発現ベクターを用いて、ヌクレオチド配列を標的器官、組織または細胞集団へ輸送することができる (Di Nicola, M. 他 (1998) *Cancer Gen. Ther.* 5(6):350-356、Yu, M. 他 (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344、Buller, R.M. 他 (1985) *Nature* 317(6040):813-815; McGregor, D.P. 他 (1994) *Mol. Immunol.* 31(3):219-226、Verma, I.M. and N. Somia (1997) *Nature* 389:239-242等を参照)。本発明は使用される宿主細胞によって限定されるものではない。

【0178】

細菌系では、多数のクローニングベクター及び発現ベクターが、HEMをコードするポリヌクレオチド配列の使用目的に応じて選択可能である。例えば、HEM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、PBLUES CRIP T (Stratagene, La Jolla CA) または pSPORT1 プラスミド (GIBCO BRL) などの多機能の大腸菌ベクターを用いることができる。ベクターの多数のクローニング部位にHEMをコードする配列をライゲーションするとlacZ遺伝子が破壊され、組換え分子を含む形質転換された細菌の同定のための比色スクリーニング法が可能となる。更に、これらのベクターを用いて、クローニングされた配列の *in vitro* での転写、ジデオキシンスクリーニング、ヘルパーファージによる一本鎖の救出、入れ子状態の欠失を作り出すことが可能である (例えば、Van Heeke, G. 及び S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509. を参照)。例えば、抗体の産生のためなどに多量のHEMが必要な場合は、HEMの発現をハイレベルで誘導するベクターが使用できる。例えば、強力に発現を誘発するT5またはT7バクテリオファージプロモーターを含むベクターを使用できる。

【0179】

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

【0180】

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

10

【0181】

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

20

【0182】

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

【0183】

哺乳動物系の組換えタンパク質の長期にわたる産生のためには、株化細胞におけるHEM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、HEMをコードする配列を株化細胞に形質転換することが可能である。このような発現ベクターは、ウイルス起源の複製及び/または内在性の発現要素や、同じ或いは別のベクターの上の選択マーカー遺伝子を含む。ベクターの導入の後、細胞を選択培地に移す前に、強化培地で約1~2日の間増殖させる。選択マーカーの目的は選択的な媒介物に対する抵抗性を与えるとともに、その存在により導入された配列を確実に発現する細胞の増殖及び回収が可能となる。安定的に形質転換された細胞の耐性クローンは、その細胞型に好適な組織培養技術を用いて増殖可能である。

30

【0184】

任意の数の選択系を用いて、形質転換された細胞系を回収することが可能である。選択系には、以下のものに限定はしないが、単純ヘルペスウイルスチミジンキナーゼ遺伝子及びアデニンホスホリボシルトランスフェラーゼ遺伝子が含まれ、それぞれtk^rまたはap^r細胞において使用される。(例えば、Wigler, M. 他 (1977) *Cell* 11 : 223 - 232 ; 及びLowy, I. 他 (1980) *Cell* 22 : 817 - 823を参照)。また代謝拮抗物質、抗生物質或いは除草剤への耐性を選択のベースとして用いることができる。例えばdhfrはメトトレキセートに対する耐性を与え、neoはアミノグリコシッドネオマイシン及びG-418に対する耐性を与え、als或いはpatはクロルスルフロン(chlorosulfuron)、ホスフィンオトリシンアセチルトランスフェラーゼ(phosphinotricin acetyltransf

40

50

erase) に対する耐性を与える(例えば、Wigler, M. 他. (1980) Proc. Natl. Acad. Sci. 77:3567-3570; Colbere-Garapin, F. 他(1981) J. Mol. Biol. 150:1-14を参照)。さらに選択に利用できる遺伝子、例えば、代謝のために細胞が必要なものを変えるtrpB及びhisDが文献に記載されている(例えば、Hartman, S. C. 及びR. C. Mulligan(1988) Proc. Natl. Acad. Sci. 85:8047-51を参照)。アノトシアニン、緑色蛍光タンパク質(GFP; Clontech)、グルクロニダーゼ及びその基質GUS、ルシフェラーゼ及びその基質ルシフェリンなどの可視マーカーが用いられる。緑色蛍光タンパク質(GFP) (Clontech, Palo Alto, CA)も使用できる。これらのマーカーを用いて、トランスフォーマントを特定するだけでなく、特定のベクター系に起因する一過性或いは安定したタンパク質発現を定量することが可能である(例えば、Rhodes, C. A. 他(1995) Methods Mol. Biol. 55:121-131を参照)。

【0185】

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

【0186】

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

【0187】

特異的なポリクローナル抗体またはモノクローナル抗体のどちらかを用いるHEMの発現の検出及び計測のための免疫学的方法は、当分野で周知である。このような技法には、酵素に結合したイムノソルベントアッセイ(ELISA)、ラジオイムノアッセイ(RIA)、蛍光標示式細胞分取器(FACS)などがある。HEM上の2つの非干渉エピトープに反応するモノクローナル抗体を用いた、2部位のモノクローナルベースイムノアッセイ(two-site, monoclonal-based immunoassay)が好ましいが、競合の結合アッセイも用いることもできる。これらのアッセイ及びその他のアッセイは、当分野では十分に知られている。(例えば、Hampton, R. 他.(1990) Serological Methods, a Laboratory Manual. APS Press, St Paul, MN, Sect. IV; Coligan, J. E. 他 Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York, NY; 及びPound, J. D. (1990) Immunochemical Protocols, Humana Press, Totowa NJ)。

【0188】

種々の標識技術及び結合技術が当業者には周知であり、様々な核酸アッセイおよびアミノ酸アッセイに用いられ得る。HEMをコードするポリヌクレオチドに関連する配列を検出するための、標識されたハイブリダイゼーションプローブ或いはPCRプローブを生成する方法には、オリゴ標識化、ニックトランスレーション、末端標識化、または標識されたヌクレオチドを用いるPCR増幅が含まれる。別法として、HEMをコードする配列、ま

たはその任意の断片を mRNA プロブを生成するためのベクターにクローニングすることも可能である。当分野では周知であり市販されているこのようなベクターを、T7, T3, または SP6 などの好適な RNA ポリメラーゼ及び標識されたヌクレオチドの追加によって、in vitroでの RNA プロブの合成に用いることができる。これらの方法は、例えば、Amersham Pharmacia Biotech 及び Promega (Madison WI)、U.S. Biochemical Corp (Cleveland OH) が市販する種々のキットを用いて行うことができる。容易な検出のために用い得る好適なレポーター分子或いは標識には、基質、コファクター、インヒビター、磁気粒子、及び放射性核種、酵素、蛍光剤、化学発光剤、色素産生剤などが含まれる。

【0189】

10

HEM をコードするヌクレオチド配列で形質転換された宿主細胞は、細胞培地でのこのタンパク質の発現及び回収に好適な条件下で培養される。形質転換された細胞から産生されたタンパク質が分泌されるか細胞内に留まるかは、使用されるその配列及び/またはそのベクターによる。HEM 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 HEM 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。

【0190】

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

20

【0191】

本発明の別の実施例では、HEM をコードする自然或いは変更された、または組換えの核酸配列を上記した任意の宿主系の融合タンパク質の翻訳となる異種配列に結合させる。例えば、市販の抗体によって認識できる異種部分を含むキメラ HEM タンパク質が、HEM の活性のインヒビターに対するペプチドライブラリのスクリーニングを促進し得る。また、異種タンパク質部分及び異種ペプチド部分が、市販の親和性基質を用いて融合タンパク質の精製を促進し得る。このような部分には、グルタチオン S トラंसフェラーゼ (GST)、マルトース結合タンパク質 (MBP)、チオレドキシン (Trx)、カルモジュリン結合ペプチド (CBP)、6-His、FLAG、c-mc、赤血球凝集素 (HA) が含まれるが、これらに限定されるものではない。GST 及び MBP、Trx、CBP、6-His によって、固定されたグルタチオン、マルトース、フェニルアルシン酸化物 (phenylarsine oxide)、カルモジュリン、金属キレート樹脂のそれぞれで同族の融合タンパク質の精製が可能となる。FLAG、c-mc、及び赤血球凝集素 (HA) によって、これらのエピトープ標識を特異的に認識する市販のモノクローナル抗体及びポリクローナル抗体を用いた融合タンパク質の免疫親和性の精製ができる。また、HEM 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、HEM が精製の後に異種部分から切断され得る。融合タンパク質の発現と精製の方法は、Ausubel (1995、前出 ch 10) に記載されている。市販されている様々なキットを用いて、融合タンパク質の発現及び精製を促進できる。

30

40

【0192】

本発明の別の実施例では、TNT ウサギ網状赤血球可溶化液またはコムギ胚芽抽出系 (Promega) を用いて in vitro で放射能標識した HEM の合成が可能である。

50

これらの系は、T7またはT3、SP6プロモーターと機能的に結合したタンパク質をコードする配列の転写と翻訳をつなげる。翻訳は、例えば、³⁵Sメチオニンである放射能標識されたアミノ酸前駆体の存在の下で起こる。

【0193】

本発明のHEMまたはその断片を用いて、HEM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、HEMへの特異的な結合をスクリーニングすることが可能である。試験化合物の例には、抗体、オリゴヌクレオチド、タンパク質（例えば受容体）または小分子が挙げられる。

【0194】

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

【0195】

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

【0196】

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

【0197】

別の実施例では、胚性幹細胞（ES細胞）における相同組換えを用いて動物モデル系内で、HEMまたはその哺乳動物相同体をコードするポリヌクレオチドを「ロックアウト」する。このような技術は当技術分野において周知であり、ヒト疾患動物モデルの作製に有用である（米国特許第5,175,383号及び第5,767,337号等を参照）。例えば129/SvJ細胞株等のマウスES細胞は初期のマウス胚に由来し、培地で増殖させることができる。このES細胞は、ネオマイシンホスホトランスフェラーゼ遺伝子（neo: Capecchi, M. R. (1989) *Science* 244:1288-1292）等のマーカー遺伝子で破壊した目的の遺伝子を含むベクターで形質転換する。

このベクターは、相同組換えにより宿主ゲノムの対応する領域に組み込まれる。別法では、Cre-loxP系を用いて相同組換えを行い、組織特異的または発生段階特異的に目的遺伝子をノックアウトする(Martha, J. D. (1996) Clin. Invest. 97:1999-2002; Wagner, K. U. 他(1997) Nucleic Acids Res. 25:4323-4330)。形質転換したES細胞を同定し、例えばC57BL/6マウス系等から採取したマウス細胞胚盤胞に微量注入する。胚盤胞を偽妊娠メスに外科的に導入し、得られるキメラ子孫の遺伝形質を決め、これを交配させてヘテロ接合性系またはホモ接合性系を作製する。このようにして作製した遺伝子組換え動物は、潜在的な治療薬や毒性薬剤で検査することができる。

【0198】

HEMをコードするポリヌクレオチドを*in vitro*でヒト胚盤胞由来のES細胞において操作することが可能である。ヒトES細胞は、内胚葉、中胚葉及び外胚葉の細胞の種類を含む少なくとも8つの別々の細胞系統に分化する可能性を有する。これらの細胞系統は、例えば神経細胞、造血系統及び心筋細胞に分化する(Thomson, J. A. 他(1998) Science 282:1145-1147)。

【0199】

HEMをコードするポリヌクレオチドを用いて、ヒト疾患をモデルとした「ノックイン」ヒト化動物(ブタ)または遺伝子組換え動物(マウスまたはラット)を作製することが可能である。ノックイン技術を用いて、HEMをコードするポリヌクレオチドの或る領域を動物ES細胞に注入し、注入した配列を動物細胞ゲノムに組み込ませる。形質転換細胞を胚に注入し、胚を上記のように移植する。遺伝子組換え子孫または近交系について研究し、可能性のある医薬品で処理し、ヒトの疾患の治療に関する情報を得る。別法では、例えばHEMを乳汁内に分泌するなどHEMを過剰に発現する哺乳動物近交系は、便利なタンパク質源となり得る(Janne, J. 他(1998) Biotechnol. Annu. Rev. 4:55-74)。

【0200】

(治療)

HEMのある領域とヒト酵素分子のある領域との間に、例えば配列及びモチーフの文脈における化学的及び構造的類似性が存在する。更に、HEMの発現は、膵臓、尿性器および癌性の組織に密接に関連する。従って、HEMは、自己免疫/炎症性の疾患、遺伝性疾患、神経疾患、及び癌を含む細胞増殖異常においてある役割を果たすと考えられる。HEM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、HEMの発現または活性を低下させることが望ましい。また、HEM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、HEMの発現または活性を増大させることが望ましい。

【0201】

従って、一実施例において、HEM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HEMまたはその断片や誘導体を投与することが可能である。限定するものではないが、このような疾患には自己免疫/炎症性の疾患、遺伝性疾患、神経疾患、及び癌を含む細胞増殖異常が含まれ、自己免疫/炎症性の疾患の中には、後天性免疫不全症候群(AIDS)、日光性角化症、副腎機能不全、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー(APECED)、気管支炎、滑液包炎、胆嚢炎、硬変、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、発作性夜間血色素尿症、肝炎、過好酸球増加症、過敏性大腸症候群、リンパ球毒素性一時性リンパ球減少症、混合型結合組織病(MCTD)、多発性硬化症、重症筋無力症、心筋または心膜炎、骨髄線維症、骨関節炎、骨粗しょう症、膵炎、真性多血症、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、

10

20

30

40

50

原発性血小板血症、血小板減少症、潰瘍性大腸炎、ウェルナー症候群、癌合併症、血液透析、体外循環、ウイルス感染症、細菌感染症、真菌感染症、寄生虫感染症、原虫感染症、蠕虫感染症、外傷、並びにリンパ腫、白血病、および骨髄腫を含む造血性の癌が含まれ、遺伝性疾患の中には、副腎白質ジストロフィー、アルポート症候群、コロイデレミア、デュシェンヌ-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、マルファン症候群、筋ジストロフィー、筋緊張性異栄養、頻繁骨形成不全 (p y c n o d y s o s t o s i s)、レフサム症候群、網膜芽腫、鎌状赤血球貧血、サラセミア、ウェルナー症候群、フォン ウィルブランド病、ウィルムス腫瘍、ツェルヴェーガー症候群、ペルオキシソームアシル - C o A 酸化酵素欠損症 (p e r o x i s o m a l a c y l - C o A o x i d a s e d e f i c i e n c y)、ペルオキシソームチオラーゼ欠損症、ペルオキシソーム2官能タンパク質欠損症、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及びカルニチン欠損症、ミトコンドリア超長鎖アシル - C o A デヒドロゲナーゼ欠損症、ミトコンドリア短鎖アシル - C o A デヒドロゲナーゼ欠損症、ミトコンドリア電子伝達フラボタンパク質及び電子伝達フラボタンパク質：ユビキノンオキシドレダクターゼ欠損症、ミトコンドリア3官能タンパク質欠損症、ミトコンドリア短鎖3 - ヒドロキシアシル - C o A デヒドロゲナーゼ欠損症が含まれ、また、神経疾患の中には、新生児発作、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクローイツフェルト ヤコブ病、ゲルストマン症候群、Gerstmann - Straussler - Scheinker 症候群を含むプリオン病 (p r i o n d i s e a s e) と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫 (c e r e b e l l o r e t i n a l h e m a n g i o b l a s t o m a t o s i s)、脳3叉神経血管症候群、ダウン症候群を含む精神薄弱および中枢神経系の発達障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィー及び他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性精神障害、及び妄想性精神病と、季節性の感情の障害 (S A D)、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、トゥーレット病、進行性核上性麻痺、皮質基底変性症 (c o r t i c o b a s a l d e g e n e r a t i o n)、及び家族性前頭側頭痴呆が含まれ、また、細胞増殖異常の中には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病 (M C T D)、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、子宮の癌などが含まれる。

10

20

30

40

【0202】

別の実施例では、限定するものではないが上に列記した疾患を含むHEM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、HEMまたはその断片や誘導体を発現し得るベクターを患者に投与することも可能である。

【0203】

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

【0204】

更に別の実施例では、限定するものではないが上に列記した疾患を含むHEMの発現また

50

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、H E Mの活性を調節するアゴニストを患者に投与することも可能である。

【0205】

更なる実施例では、H E M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H E M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、上記した自己免疫/炎症性の疾患、遺伝性疾患、神経疾患、及び癌を含む細胞増殖異常が含まれる。一実施態様では、H E M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H E Mを発現する細胞または組織に薬剤を運ぶターゲティング或いは運搬機構として間接的に用いられ得る。

【0206】

別の実施例では、限定するものではないが上に列記した疾患を含むH E M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、H E Mをコードするポリヌクレオチドの相補配列を発現するベクターを患者に投与することも可能である。

【0207】

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

【0208】

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

【0209】

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

【0210】

H E Mに対する抗体を誘発するために用いられるオリゴペプチド、ペプチド、または断片は、少なくとも約5個のアミノ酸からなり、一般的には約10個以上のアミノ酸からなるものが好ましい。これらのオリゴペプチド或いはペプチド、またはそれらの断片は、天然のタンパク質のアミノ酸配列の一部と同一であることが望ましい。H E M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。

【0211】

H E Mに対するモノクローナル抗体は、培地内の連続した細胞株によって、抗体分子を産生する任意の技術を用いて作製することが可能である。これらの技術には、ハイブリドーマ技術、ヒトB細胞ハイブリドーマ技術、及びE B V - ハイブリドーマ技術が含まれるが、これらに限定されるものではない（例えば、K o h l e r , G . ら . (1 9 7 5)

10

20

30

40

50

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

【0212】

更に、「キメラ抗体」作製のために発達したヒト抗体遺伝子にマウス抗体遺伝子をスプライシングするなどの技術が、好適な抗原特異性及び生物学的活性を備える分子を得るために用いられる(例えば、Morrisson, S. L. 他. (1984) Proc. Natl. Acad. Sci. 81-4851-4855; Neuberger, M. S. 他. (1984) Nature 312:604-608; Takeda, S. ら. (1985) Nature 314:452, 454を参照)。別法では、当分野で周知の方法を用いて、一本鎖抗体の産生のための記載された技術を適用して、HEM特異性一本鎖抗体を生成する。関連する特異性を備えるが別のイディオタイプの組成の抗体は、ランダムな組み合わせの免疫グロブリンライブラリから鎖混合によって生成することもできる(例えば、Burton D. R. (1991) Proc. Natl. Acad. Sci. 88:11120-3を参照)。

【0213】

抗体は、リンパ球集団の中の *in vivo* 産生を誘発することによって、または免疫グロブリンライブラリのスクリーニングまたは文献に示されているような、高度に特異的な結合試薬のパネルをスクリーニングすることによって、得ることもできる(例えば、Orlandi, R. 他. (1989) Proc. Natl. Acad. Sci. 86:3833-3837; Winter, G. 他. (1991) Nature 349:293-299を参照)。

【0214】

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

【0215】

種々のイムノアッセイを用いてスクリーニングし、所望の特異性を有する抗体を同定する。隔離された特異性を有するポリクローナル抗体またはモノクローナル抗体の何れかを用いる競合的な結合、または免疫放射線活性のための数々のプロトコルが、当分野では周知である。通常このようなイムノアッセイには、HEMとその特異性抗体との間の複合体調整の計測が含まれる。二つの非干渉性HEM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)。

【0216】

ラジオイムノアッセイ技術と共にScatchard分析などの様々な方法を用いて、HEMに対する抗体の親和性を評価する。親和性を結合定数Kaで表すが、このKaは、平衡状態の下でHEM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HEMエピトープに対して親和性が不均一なポリクローナル抗体医薬のKaは、HEMに対する抗体の平均親和性または結合活性を表す。特定のHEMエピトープに単一特異的なモノクローナル抗体医薬のKaは、親和性の真の測定値を表す。Ka値が $10^9 \sim 10^{12}$ L/molの高親和性抗体医薬は、HEM抗体複合体が激しい操作に耐えなければならないイムノアッセイに用いるのが好ましい。Ka値が $10^6 \sim 10^7$ L/molの低親和性抗体医薬は、HEMが抗体から最終的に活性化状態で解離する必

10

20

30

40

50

要がある免疫精製 (immunopurification) 及び類似の処理に用いるのが好ましい。(Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J. E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY)。

【0217】

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

【0218】

本発明の別の実施例では、HEMをコードするポリヌクレオチド、またはその任意の断片や相補配列が、治療目的で使用することができる。ある実施態様では、HEM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、HEM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。

【0219】

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

【0220】

本発明の別の実施例では、HEMをコードするポリヌクレオチドを、体細胞若しくは生殖細胞の遺伝子治療に用いることが可能である。遺伝子治療は、(i) 遺伝子欠損症(例えば、X染色体連鎖遺伝(Cavazzana-Calvo, M. 他(2000) *Science* 288: 669-672)によって特徴づけられる重度の複合型免疫欠損(SCID)-X1)、遺伝性アデノシン-デアミナーゼ(ADA)欠損症(Blaese, R. M. 他(1995) *Science* 270: 475-480; Bordignon, C. 他(1995) *Science* 270: 470-475)に関連する重度の複合型免疫欠損、嚢胞性繊維症(Zabner, J. 他(1993) *Cell* 75: 207-216; Crystal, R. G. 他(1995) *Hum. Gene Therapy* 6: 643-666; Crystal, R. G. 他(19

95) Hum. Gene Therapy 6:667-703)、サラセミア (thalassaemia)、家族性高コレステロール血症、第VIII因子若しくは第IX因子欠損による血友病 (Crystal, 35 R.G. (1995) Science 270:404-410; Verma, I.M. and Somia, N. (1997) Nature 389:239-242) を治療したり、(ii) 条件的致死性遺伝子産物 (例えば、細胞増殖の制御不能による癌の場合) を発現させたり、及び (iii) 細胞内の寄生虫 (例えば、ヒト免疫不全ウイルス (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschl, E. 他 (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399) や、B型若しくはC型肝炎ウイルス (HBV、HCV)、Candida albicans 及び Paracoccidioides brasiliensis 等の真菌寄生虫、Plasmodium falciparum 及び Trypanosoma cruzi 等の原虫寄生体) に対する防御機能を有するタンパク質を発現させて行うことができる。HEM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HEM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。

10

【0221】

本発明の更なる実施例では、HEMの欠損による疾患や異常症は、HEMをコードする哺乳動物発現ベクターを作製して、これらのベクターを機械的手段によってHEM欠損細胞に導入することによって治療する。in vivo 或いは ex vivo の細胞に用いる機械的な導入技術には、(i) 個々の細胞内へのDNAのマイクロインジェクション、(ii) 金粒子の打ち込み、(iii) リポソーム仲介性トランスフェクション、(iv) 受容体仲介性遺伝子導入、及び (v) DNAトランスポゾン (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Recipon (1998) Curr. Opin. Biotechnol. 9:445-450) の使用が含まれる。

20

【0222】

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

30

40

【0223】

市販のリポソーム形質転換キット (例えば、Invitrogenが販売しているPER

50

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

【0224】

本発明の別の実施例では、HEMの発現に関連する遺伝子欠損によって起こる疾患や異常症は、(i)レトロウイルス末端反復配列(LTR)プロモーター若しくは独立したプロモーターのコントロール下でHEMをコードするポリヌクレオチドと、(ii)好適なRNAパッケージングシグナルと、(iii)追加のレトロウイルス・シス作用性RNA配列及び効率的なベクターの増殖に必要なコーディング配列を伴うRev応答性エレメント(RRE)とからなるレトロウイルスベクターを作製して治療することができる。レトロウイルスベクター(例えば、PFB及びPFBNEO)はStratagene社から入手可能であり、公表データ(Riviere, I. 他.(1995) *Proc. Natl. Acad. Sci. U.S.A.* 92: 6733 - 6737)に基づいている。上記データを引用することをもって本明細書の一部とする。このベクターは、VSVg(Armentano, D. 他(1987) *J. Virol.* 61: 1647 - 1650; Bender, M. A. 他(1987) *J. Virol.* 61: 1639 - 1646; Adam, M. A. and A. D. Miller (1988) *J. Virol.* 62: 3802 - 3806; Dull, T. 他(1998) *J. Virol.* 72: 8463 - 8471; Zufferey, R. 他(1998) *J. Virol.* 72: 9873 - 9880)等の乱交雑エンベロープタンパク質若しくは標的細胞上の受容体に対する親和性を有するエンベロープ遺伝子を発現する好適なベクター産生細胞系(VPCL)において増殖される。RIGGに付与された米国特許第5,910,434号(「Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant」)において、レトロウイルスパッケージング細胞系を得るための方法が開示されており、引用することをもって本明細書の一部とする。レトロウイルスベクターの増殖、ある細胞集団(例えば、CD4⁺T細胞)の形質導入、並びに形質導入した細胞を患者に戻す方法は、遺伝子治療の分野では周知であり、多数の文献に記載されている(Ranga, U. 他.(1997) *J. Virol.* 71: 7020 - 7029; Bauer, G. 他(1997) *Blood* 89: 2259 - 2267; Bonyhadi, M. L. (1997) *J. Virol.* 71: 4707 - 4716; Ranga, U. 他(1998) *Proc. Natl. Acad. Sci. U.S.A.* 95: 1201 - 1206; Su, L. (1997) *Blood* 89: 2283 - 2290)。

【0225】

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

9 : 5 1 1 - 5 4 4 ; and Verma , I . M . and N . Somia (1 9 9 7) Nature 1 8 : 3 8 9 : 2 3 9 - 2 4 2 を参照し、引用することをもって本明細書の一部とする。

【 0 2 2 6 】

別法では、ヘルペス系遺伝子治療の送達系を用いて、HEMの発現に関連する1或いは複数の遺伝子異常を有する標的細胞にHEM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HEMを導入する際に特に重要である。ヘルペス系ベクターの作製及びパッケージングは当分野では周知である。複製適格性の単純疱疹ウイルス(HSV)I型系のベクターは、霊長類の眼にレポーター遺伝子を送達するために用いられてきた(Liu, X. 他(1999) Exp. Eye Res. 169:385-395)。HSV-1ウイルスベクターの作製は、DeLucaに付与された米国特許第5,804,413号(Herpes simplex virus swains for gene transfer)に記載されており、引用することをもって本明細書の一部とする。米国特許第5,804,413号には、ヒト遺伝子治療を含む目的のために、好適なプロモーターのコントロールの下で、細胞に導入される少なくとも1つの内在性遺伝子を含むゲノムからなる組換えHSV d9 2 についての記載がある。また上記特許には、ICP4、ICP27及びICP22のために除去される組換えHSV株の作製及び使用方法が開示されている。HSVベクターについては、Goins, W.F. 他(1999) J. Virol. 73:519-532 and Xu, H. 他(1994) Dev. Biol. 163:152-161を参照し、引用することをもって本明細書の一部とする。クローニングされたヘルペスウイルス配列の操作や、巨大ヘルペスウイルスのゲノムの異なった部分を含む多数のプラスミドをトランスフェクトした後の組換えウイルスの継代、ヘルペスウイルスの成長及び増殖、並びにヘルペスウイルスの細胞への感染は当分野で周知の技術である。

【 0 2 2 7 】

別法では、ウイルス(正の一本鎖RNAウイルス)ベクターを用いてHEMをコードするポリヌクレオチドを標的細胞に送達する。プロトタイプのウイルスであるセムリキ森林熱ウイルス(Semliki Forest Virus, SFV)の生物学的な研究が広範に行われ、遺伝子伝達ベクター(gene transfer vector)がSFVゲノムに基づいていることが分かった(Garoff, H. and K.-J. Li(1998) Cun. Opin. Biotech. 9:464-469)。ウイルスRNAの複製中に、通常はウイルスカプシドタンパク質をコードするサブゲノムRNAが作り出される。このサブゲノムRNAが完全長のゲノムRNAより高いレベルで複製されるため、酵素活性(例えばプロテアーゼ及びポリメラーゼ)を有するウイルスタンパク質に対してカプシドタンパク質が過剰に産生される。同様に、HEMをコードする配列をウイルスゲノムのカプシドをコードする領域に導入することによって、ベクター導入細胞において多数のHEMをコードするRNAが産生され、高いレベルでHEMが合成される。通常はウイルス感染は2~3日以内の細胞溶解に関係するが、シンドビスウイルス(SIN)の変異体を有するハムスターの正常な腎細胞(BHK-21)の持続的な感染を確立する能力は、ウイルスの溶解性の複製が遺伝子治療に適用できるように好適に変更することが可能であることを示唆している(Dryga, S.A. 他(1997) Virology 228:74-83)。様々な宿主にウイルスを導入できることから、様々なタイプの細胞にHEMを導入することができる。ある集団における細胞のサブセットの特定の形質導入には、形質導入する前に細胞のソーティングを必要とする場合がある。ウイルスの感染性cDNAクローンの操作、ウイルスcDNA及びRNAのトランスフェクション、並びにウイルスの感染方法は当分野で周知である。

【 0 2 2 8 】

例えば開始部位から約-10から約+10までの転写開始部位に由来するオリゴヌクレオチドを用いて、遺伝子の発現を阻害することが可能である。同様に、三重らせん塩基対合法を用いて阻害することができる。三重らせん構造は、二重らせんがポリメラーゼ、転写

10

20

30

40

50

因子、または調節分子の結合のために十分に広がるのを阻止するため有用である。三重鎖 DNA を用いる最近の治療の進歩は文献に記載されている（例えば、Gee, J. E. ら. (1994) *In: Huber, B. E. 及び B. I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY* を参照）。相補的な配列またはアンチセンス分子もまた、転写物がリボソームに結合するのを阻止することによって mRNA の翻訳を阻止するように設計できる。

【0229】

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

10

【0230】

任意の潜在的 RNA 標的内の特異的なリボザイム切断部位が、後続の配列 GUA、GUU、及び GUC を含むリボザイム切断部位に対して、標的分子をスキミングすることによって初めに同定される。一度同定されると、切断部位を含む標的遺伝子の領域に対応する 15 個から 20 個のリボヌクレオチドの短い RNA 配列を、オリゴヌクレオチドの機能を不全にする二次的な構造の特徴について評価することが可能である。候補標的の適合性も、リボヌクレアーゼ保護アッセイを用いて、相補的なオリゴヌクレオチドとのハイブリダイゼーションの容易性をテストすることによって評価することが可能である。

20

【0231】

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

30

【0232】

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

40

【0233】

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

50

ては、HEMをコードするポリヌクレオチドの発現を特異的に促進する化合物が治療上有用であり得る。

【0234】

特定のポリヌクレオチドの発現の変化の有効性を調べるために、少なくとも1個から複数個の試験化合物をスクリーニングすることができる。試験化合物は、有効な化合物の化学修飾を含む当分野で周知の任意の方法で得ることができる。このような方法は、ポリヌクレオチドの発現を変化させる場合、一般に市販されている或いは専売の天然または非天然の化合物ライブラリから選択する場合、標的ポリヌクレオチドの化学的及び/または構造的な特性に基づいて化合物を合理的にデザインする場合、更に組合せ的にまたは無作為に生成した化合物のライブラリから選択する場合に有効である。HEM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、透過処理した細胞、in vitro細胞遊離系または再構成生物学系が含まれ得る。HEMをコードするポリヌクレオチドの発現における変化は、当分野で周知の任意の方法でアッセイする。通常、HEMをコードするポリヌクレオチドの配列に相補的なヌクレオチド配列を有するプローブを用いたハイブリダイゼーションにより、特定のヌクレオチドの発現を検出する。ハイブリダイゼーションの収量を定量し、その値が1或いは複数の試験化合物に曝露される及び曝露されないポリヌクレオチドの発現の比較における基準となり得る。試験化合物に曝露されるポリヌクレオチドの発現の変化が検出される場合は、ポリヌクレオチドの発現の変化に試験化合物が有効であることを示している。特定のポリヌクレオチドの発現の変化に有効な化合物を調べるために、例えば Sc hizosaccharomyces pombe 遺伝子発現系 (Atkins, D. 他 (1999) 米国特許第5,932,435号、Arndt, G.M. 他 (2000) *Nucleic Acids Res.* 28:E15) または HeLa 細胞等のヒト細胞株 (Clarke, M.L. 他 (2000) *Biochem. Biophys. Res. Commun.* 268:8-13) を用いてスクリーニングする。本発明の特定の実施例は、特異的ポリヌクレオチド配列に対するアンチセンス活性を調べるための、各オリゴヌクレオチド (デオキシリボヌクレオチド、リボヌクレオチド、ペプチド核酸、及び修飾オリゴヌクレオチド) の組み合わせライブラリのスクリーニングを含む (Bruce, T.W. 他 (1997) 米国特許第5,686,242号、Bruce, T.W. 他 (2000) 米国特許第6,022,691号)。

10

20

30

【0235】

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

【0236】

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

40

【0237】

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

【0238】

50

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

【0239】

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

10

【0240】

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

【0241】

H E Mまたはその断片を含む高分子を直接細胞内に輸送するべく、特殊な形態に組成物が調製されるのが好ましい。例えば、細胞不透過性高分子を含むリポソーム製剤は、細胞融合及び高分子の細胞内輸送を促進し得る。別法では、H E Mまたはその断片をH I V T a t - 1タンパク質の陽イオンN末端部に結合することもできる。このようにして作製された融合タンパク質は、マウスモデル系の脳を含む全ての組織の細胞に形質導入されることが確認されている（Schwarze, S. R. 他（1999）Science 285:1569-1572）。

20

【0242】

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

30

【0243】

医学的に効果的な薬用量は、症状や容態を回復させる、たとえばH E Mまたはその断片、H E Mの抗体、H E Mのアゴニストまたはアンタゴニスト、インヒビターなどの活性処方成分の量に関連する。薬用有効度及び毒性は、たとえば、E D₅₀（服用に対して集団の50%に医薬的效果がある用量）またはL D₅₀（服用に対して集団の50%に致命的である用量）統計を計算するなど、細胞培養または動物実験における標準的な薬剤手法によって決定することができる。毒性効果と治療効果との薬用量比は治療指数であり、L D₅₀ / E D₅₀と示すことができる。高い治療指数を示す組成物が望ましい。細胞培養アッセイ及び動物実験から得られたデータが、ヒトへの適用のために、薬用量の範囲を調剤するのに用いられる。このような組成物が含まれる薬用量は、毒性を殆ど或いは全く含まず、E D₅₀を含む血中濃度の範囲であることが望ましい。薬用量は、用いられる投与形態及び患者の感受性、投与の経路によって、この範囲内で様々である。

40

【0244】

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

50

【0245】

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

【0246】

(診断)

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

【0247】

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

【0248】

本発明の別の実施例によれば、HEMをコードするポリヌクレオチドを診断のために用いることもできる。用いられるポリヌクレオチドには、オリゴヌクレオチド配列、相補的なRNA及びDNA分子、及びPNAが含まれる。このポリヌクレオチドを用いて、疾患と相関し得るHEM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、HEMの存在の有無、更に過剰な発現を調べ、治療中のHEM値の調節を監視する。

【0249】

一実施形態では、HEM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、HEMをコードする核酸配列を同定することが可能である。例えば5'調節領域である高度に特異的な領域か、例えば保存されたモチーフであるやや特異性の低い領域から作られているかのプローブの特異性と、ハイブリダイゼーション或いは増幅のストリンジェントは、プローブがHEMをコードする自然界の配列のみを同定するかどうか、或いはアレルや関連配列コードする自然界の配列のみを同定するかどうかによって決まるであろう。

【0250】

プローブはまた、関連する配列の検出に利用され、HEMをコードする任意の配列と少なくとも50%の配列同一性を有し得る。目的の本発明のハイブリダイゼーションプローブには、DNAあるいはRNAが可能であり、SEQ ID NO: 27-52の配列、或いはHEM遺伝子のプロモーター、エンハンサー、イントロンを含むゲノム配列に由来し得る。

【0251】

HEMをコードするDNAに対して特異的なハイブリダイゼーションプローブの作製方法には、HEM及びHEM誘導体をコードするポリヌクレオチド配列をmRNAプローブの

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

【0252】

HEMをコードするポリヌクレオチド配列を用いて、HEMの発現に関連する疾患を診断することが可能である。限定するものではないが、このような疾患には自己免疫/炎症性の疾患、遺伝性疾患、神経疾患、及び癌を含む細胞増殖異常が含まれ、自己免疫/炎症性の疾患の中には、後天性免疫不全症候群 (AIDS)、日光性角化症、副腎機能不全、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、自己免疫性多腺性内分泌カンジダ性外胚葉ジストロフィー (APECED)、気管支炎、滑液包炎、胆嚢炎、硬変、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、糖尿病、肺気腫、リンパ球毒素性一時性リンパ球減少症、赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、発作性夜間血色素尿症、肝炎、過好酸球増加症、過敏性大腸症候群、リンパ球毒素性一時性リンパ球減少症、混合型結合組織病 (MCTD)、多発性硬化症、重症筋無力症、心筋または心膜炎、骨髄線維症、骨関節炎、骨粗しょう症、脾炎、真性多血症、乾癬、ライター症候群、リウマチ様関節炎、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、原発性血小板血症、原発性血小板減少症、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、並びにリンパ腫、白血病、および骨髄腫を含む造血性の癌が含まれ、遺伝性疾患の中には、副腎白質ジストロフィー、アルポート症候群、コロイデミア、デュシェンヌ-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、マルファン症候群、筋ジストロフィー、筋緊張性異栄養、頻繁骨形成不全 (pseudotumor cerebri)、レフサム症候群、網膜芽腫、鎌状赤血球貧血、サラセミア、ウェルナー症候群、フォン・ウイランド病、ウィルムス腫瘍、ツェルヴェーガー症候群、ペルオキシソームアシル-CoA酸化酵素欠損症 (peroxisomal acyl-CoA oxidase deficiency)、ペルオキシソームチオラーゼ欠損症、ペルオキシソーム2官能タンパク質欠損症、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及びカルニチン欠損症、ミトコンドリア超長鎖アシル-CoAデヒドロゲナーゼ欠損症、ミトコンドリア短鎖アシル-CoAデヒドロゲナーゼ欠損症、ミトコンドリア電子伝達フラボタンパク質及び電子伝達フラボタンパク質：ユビキノンオキシドレダクターゼ欠損症、ミトコンドリア3官能タンパク質欠損症、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損症が含まれ、また、神経疾患の中には、新生児発作、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクローイツフェルトヤコブ病、ゲルストマン症候群、Gerstmann-Straussler-Scheinker症候群を含むプリオン病 (prion disease) と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫 (cerebellar hemangioblastomatosis)、脳3叉神経血管症候群、ダウン症候群を含む精神薄弱および中枢神経系の発達障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィー及び他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及

び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性精神障害、及び妄想性精神病と、季節性の感情の障害（SAD）、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、トゥレット病、進行性核上性麻痺、皮質基底変性症（corticobasal degeneration）、及び家族性前頭側頭痴呆が含まれ、また、細胞増殖異常の中には日光性角化症及びアテローム性動脈硬化、滑液包炎、硬変、肝炎、混合型結合組織病（MCTD）、骨髄線維症、発作性夜間ヘモグロビン尿症、真性多血症、乾癬、原発性血小板血症、並びに腺癌及び白血病、リンパ腫、黒色腫、骨髄腫、肉腫、及び奇形癌、具体的には、副腎、膀胱、骨、骨髄、脳、乳房、頸部、胆嚢、神経節、消化管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、子宮の癌などが含まれる。HEMをコードするポリヌクレオチド配列は、サザン法やノーザン法、ドットプロット法、或いはその他の膜系の技術、PCR法、ディップスティック（dipstick）、ピン（pin）、ELISA式アッセイ、及び変異HEMの発現を検出するために患者から採取した体液或いは組織を利用するマイクロアレイに使用することが可能である。このような質的或いは量的方法は、当分野では周知である。

10

【0253】

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

20

【0254】

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

30

【0255】

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

40

【0256】

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

【0257】

HEMをコードする配列から設計されたオリゴヌクレオチドのさらなる診断への利用には、PCRの利用が含まれ得る。このようなオリゴマーは、化学的な合成、酵素を用いた生成、或いは *in vitro* で生成され得る。オリゴマーは、好ましくはHEMをコードするポリヌクレオチドの断片、或いはHEMをコードするポリヌクレオチドと相補的なボ

50

リヌクレオチドの断片を含み、最適な条件の下、特定の遺伝子や条件を識別するために利用される。また、オリゴマーは、やや緩いストリンジェントな条件の下、近縁のDNA或いはRNA配列の検出及び/または定量のため用いることが可能である。

【0258】

或る実施態様において、HEMをコードするポリヌクレオチド配列由来のオリゴヌクレオチドプライマーを用いて、一塩基多型(SNP)を検出し得る。SNPは、ヒトの先天性または後天性遺伝病の原因となる場合が多いヌクレオチドの置換、挿入及び欠失である。限定するものではないが、SNPの検出方法には、一本鎖立体構造多型(SSCP)及び蛍光SSCP(fSSCP)法が含まれる。SSCPでは、HEM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。このDNAは、例えば病変或いは正常な組織、生検サンプル、体液等に由来し得る。このDNA内のSNPは、一本鎖形状のPCR産物の2次及び3次構造に差異を生じさせる。この差異は非変性ゲル中でのゲル電気泳動法を用いて検出可能である。fSSCPでは、オリゴヌクレオチドプライマーを蛍光標識することによって、DNAシーケンシング装置などのハイスループット機器でアンプリマー(amplimer)の検出をすることが可能になる。更に、インシリコSNP(in silico SNP: isSNP)と呼ばれる配列データベース分析法は、共通のコンセンサス配列の構築に用いられる個々の重複するDNA断片の配列を比較することによって、多型を同定することができる。これらのコンピュータベースの方法は、DNA配列クロマトグラムの自動分析及び統計モデルを用いたシーケンシングエラーや研究室でのDNAの調整に起因する配列のばらつきを排除する。別法では、例えばハイスループットの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。

10

20

【0259】

HEM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(例えば、Melby, P. C.ら(1993) J. Immunol. Methods, 159: 235-44; Duplaa, C.ら(1993) Anal. Biochem. 229-236を参照)。多数のサンプルの定量速度は、ハイスループット型のアッセイを用いることで速くなるであろう。このアッセイでは、目的のオリゴマーやポリヌクレオチドが様々な希釈液中に含まれ、分光光度法或いは非色応答によって定量が迅速である。

30

【0260】

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

40

【0261】

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

【0262】

50

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

10

【0263】

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

【0264】

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

20

30

40

【0265】

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

【0266】

別の実施例は、本発明のポリペプチド配列を用いて組織または細胞型のプロテオームを分

50

析することに関連する。「プロテオーム」という用語は、或る特定の組織または細胞型におけるタンパク質発現の包括的パターンを指す。プロテオームを構成する各タンパク質は、個々に更なる分析をすることができる。プロテオーム発現パターン即ちプロファイルは、所定の条件下で所定の時間に発現したタンパク質の数及びそれらの相対的な存在量を定量することにより分析する。従って、ある細胞のプロテオームのプロファイルは、特定の組織または細胞型のポリペプチドを分離及び分析することにより作成し得る。或る実施例では、このような分離は2次元ゲル電気泳動によって行う。この2次元ゲル電気泳動法では、まず、1次元の等電点電気泳動によりサンプルからタンパク質を分離し、次に、2次元のドデシル硫酸ナトリウムスラブゲル電気泳動により分子量に従って分離する（前出の Steiner and Anderson）。これらのタンパク質は、通常クーマシーブルーまたはシルバーまたは蛍光染色などの染色剤を用いてゲルを染色して、分散した個別の位置にあるスポットとしてゲル中で可視化される。各タンパク質スポットの光学密度は、通常サンプル中のタンパク質レベルに比例する。異なるサンプル、例えば試験化合物または治療薬で処理済みまたは未処理のいずれかの生体サンプルから得られる等位置にあるタンパク質スポットの光学密度を比較し、処理に関連するタンパク質スポット密度の変化を調べる。スポット内のタンパク質は、例えば化学的または酵素的に切断した後、質量分析する標準的な方法を用いて部分的にシーケンシングする。スポット内のタンパク質の同一性は、好適には少なくとも5個の連続するアミノ酸残基であるその部分的な配列を、本発明のポリペプチド配列と比較することにより決定し得る。場合によっては、決定的なタンパク質同定のための更なる配列が得られる。

10

20

【0267】

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

30

【0268】

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

40

【0269】

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

【0270】

別の実施例では、試験化合物の毒性は、タンパク質を含む生体サンプルをその試験化合物

50

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

【0271】

当分野で周知の方法でマイクロアレイを準備して使用し、分析する。(例えば、Brennan, T.M. 他 (1995) 米国特許第5,474,796号; Schena, M. 他 (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler 他 (1995) PCT出願番号WO95/251116; Shalon, D. 他 (1995) PCT出願番号WO95/35505; Heller, R.A. 他 (1997) Proc. Natl. Acad. Sci. 94:2150-2155; 及び Heller, M.J. 他 (1997) 米国特許第5,605,662号を参照)。様々なタイプのマイクロアレイが周知であり、詳細については、DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, Londonに記載されている。また、この文献を引用することを以って本明細書の一部とする。

【0272】

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

【0273】

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

【0274】

染色体標本のin situハイブリダイゼーション、及び確定した染色体マーカーを用いた結合分析などの物理的マッピング技術を用いて、遺伝子地図を拡張することもできる。マウスなどの別の哺乳動物の染色体上に遺伝子を配置させることにより、たとえ正確なヒト染色体の位置が分かっていなくても、関連するマーカーが明らかになる場合が多い。この情報は、位置クローニング或いは別の遺伝子発見技術を用いて遺伝的疾患の研究をして

いる研究者にとって価値がある。疾患や症候群に關与する1つ或いは複数の遺伝子の位置が、例えば血管拡張性失調症の11q22-23などの特定の遺伝子領域に遺伝子結合によって大まかに決定されると、その領域に対するどの配列マッピングも、さらなる調査のための關連する遺伝子或いは調節遺伝子を表す(例えば、Gatti, R. A. 他による(1988) Nature 336:577-580を参照)。また、目的の本発明のヌクレオチド配列を用いて、正常者、保有者、即ち感染者の間の、転位置、反転などによる染色体位置の違いを検出することもある。

【0275】

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

10

【0276】

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

20

【0277】

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

【0278】

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

30

【0279】

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

【0280】

前述した及び以下に記載する全ての特許出願、特許、刊行物、特に米国特許出願第60/186,307号および同第60/192,532号に言及することをもって本明細書の一部とする。

【0281】

(実施例)

40

1 cDNAライブラリの作製

インサイトcDNAはLIFESEQ GOLD データベース (Incyte Genomics, Palo Alto CA) に含まれているcDNAライブラリに由来し、表4の列5に示されている。組織の一部をホモジナイズしてグアニジニウムイソチオシアネート溶液に溶解する一方、この組織の別の一部をホモジナイズしてフェノールに溶解するか、或いはTRIZOL (Life Technologies)、グアニジニウムイソチオシアネート及びフェノールの単相溶液などの好適な変性剤の混合液に溶解した。この溶解物を塩化セシウムにおいて遠心分離によって、或いはクロロホルムで抽出した。イソプロパノール或いは酢酸ナトリウムのどちらかとエタノール、或いは別の方法でこの溶解

50

物からRNAを沈殿させた。

【0282】

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

【0283】

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

【0284】

2 cDNAクローンの単離

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

【0285】

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

10

20

30

40

50

【0286】

3 シークエンシング及び分析

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

【0287】

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

リヌクレオチド配列およびポリペプチド配列のアラインメントを、アラインメントした配列間のパーセント同一性も計算するMEGALIGNマルチシーケンスアラインメントプログラム(DNASTAR)に組み込まれたCLUSTALアルゴリズムによって指定されたデフォルトパラメータを用いて作成した。

【0288】

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

10

【0289】

完全長のポリヌクレオチド配列およびポリペプチド配列の組み立て及び分析に用いる上記のプログラムは、SEQ ID NO: 27-52のポリヌクレオチド配列断片の同定にも利用できる。ハイブリダイゼーション及び増幅技術に有用である約20~約4000ヌクレオチドの断片を表4の列4に示した。

【0290】

4 ゲノムDNAからのコード配列の同定および編集

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

20

30

40

【0291】

5 cDNA配列データを用いたゲノム配列データの組み立て

ステッチ配列(Stitched Sequence)

部分的なcDNA配列を、実施例4に記載したGenscan遺伝子同定プログラムによって推定されたエキソンで伸長した。実施例3に記載されたように組み立てられた部分的なcDNAをゲノムDNAにマッピングし、1或いは複数のゲノム配列から推定されるG

50

e n s c a n エキソンおよび関連 c D N A を含む複数のクラスターに分けた。各クラスターを、グラフ理論および動的計画法に基づいたアルゴリズムを用いて、c D N A およびゲノム情報を統合して分析し、可能性のあるスプライスバリエーションを生成して、続いて確認、編集、または伸長して完全長の配列を作製した。区間の全長が或るクラスターの2つ以上の配列に存在する配列区間を同定し、このように同定した区間は移行によるもので同等と考える。例えば、或る区間が c D N A および2つのゲノム配列のそれぞれに存在する場合、これら3つ全ての区間を同等と考える。この方法によって、関連性がないが連続するゲノム配列を c D N A 配列によって繋ぎ合わせる。このようにして同定された区間を、親配列 (p a r e n t s e q u e n c e) に沿って現われるようにステッチアルゴリズムを用いて順に縫い合わせ、可能な最長の配列および変異配列を作製する。或るタイプの親配列 (c D N A と c D N A 、またはゲノム配列とゲノム配列) に沿って連結される区間と区間との繋ぎ合わせは、親配列のタイプが異なる (c D N A とゲノム配列) 連結より好ましい。得られたステッチ配列を翻訳し、B L A S T 解析で g e n p e p t および g b p r i 公共データベースにおける配列と比較した。G e n s c a n によって推定された不適当なエキソンを、g e n e p t において B L A S T で最もヒットした配列と比較して修正する。このような配列を更なる c N D A 配列で伸長し、必要に応じてゲノム D N A で検査した。

10

【0292】

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

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

20

30

【0293】

6 H E M をコードするポリヌクレオチドの染色体マッピング

S E Q I D N O : 27 - 52 を組み立てるために用いた配列を、B L A S T 及び S m i t h - W a t e r m a n アルゴリズムを用いて、インサート L I F E S E Q 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。S E Q I D N O : 27 - 52 と一致するこれらのデータベースの配列を、P h r a p (表7) などの構築アルゴリズムを使用して、連続及び重複した配列のクラスターに組み入れた。S t a n f o r d H u m a n G e n o m e S e q u e n c e C e n t e r (S H G C) 、W h i t e h e a d I n s t i t u t e f o r G e n o m e R e s e a r c h (W I G R) 及び G e n e t h o n などの公共の情報源から入手できる放射線ハイブリッド (r a d i a t i o n h y b r i d) 及び遺伝子マッピングのデータを用いて、クラスター化した配列がすでにマッピングされているかを調べる。クラスターにマッピングされた配列が含まれている場合は、そのクラスターの全ての配列 (特定の S E Q I D N O を含む) をそのマッピング位置に割り当てた。

40

【0294】

遺伝子地図の位置は、範囲、区間、またはヒト染色体によって表される。センチモルガンで示したマッピング位置の範囲は、染色体の短腕 (p) の末端から測定した (センチモルガン (c M)) は、同一染色体上の遺伝子間の乗換え率に基づいた距離を表す単位である。

50

平均すると、1 cMはヒトの染色体の1メガベースに概ね等しいが、組換え率の高い部分と低い部分があるため、大きく変化し得る)。距離cMは、配列がそれぞれのクラスターに含まれている放射線ハイブリッドマーカーの境界を検出できるGeneThonによってマッピングされた遺伝子マーカーに基づいている。NCBI「GeneMap99」(<http://www.ncbi.nlm.nih.gov/genemap>)などの公衆が入手可能なヒト遺伝子マップおよびその他の情報源を用いて、上記した区間が既に同定されている疾患遺伝子マップ内若しくは近傍に位置するかを決定できる。

【0295】

このようにして、SEQ ID NO: 32が、第20染色体の11.0~20.9センチモルガンの範囲内にマッピングされた。SEQ ID NO: 35は、第2染色体の175.0~180.6センチモルガンおよび190.8~196.8センチモルガンの範囲内にマッピングされた。SEQ ID NO: 41は、第1染色体の235.8~237.2センチモルガンおよび243.3~245.2センチモルガンの範囲内にマッピングされた。SEQ ID NO: 47は、第2染色体の118.0~127.4センチモルガンの範囲内にマッピングされた。SEQ ID NO: 35およびSEQ ID NO: 41では複数のマッピング位置を示したが、これは異なったマッピング位置が1つのクラスター内に集められたことを表す。例えば、高い類似性を有するが完全には同一でない配列が1つのクラスター内に集められた場合に生じる。

10

【0296】

7 ポリヌクレオチド発現の分析

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

20

【0297】

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

30

【0298】

【数1】

(BLASTスコア×配列一致率)

5 × (長さ(配列1), 長さ(配列2))の最小値

として定義される積スコアである。積スコアは、0~100の標準化された値であり、以下のように求める。BLASTスコアにヌクレオチド配列の一致率を乗じ、その積を2つの配列の短い方の長さの5倍で除する。高スコアのセグメントの対(HSP)において一致する各塩基に+5のスコアを割り当て、各不適性塩基対に-4を割り当てることにより、BLASTスコアを計算する。2つの配列は、2以上のHSPを共有し得る(ギャップにより離隔される)。2以上のHSPがある場合には、最高BLASTスコアの塩基対を用いて積スコアを計算する。積スコアは、BLASTアラインメントの断片的重複と質とのバランスを表す。例えば積スコア100は、比較した2つの配列の短い方の長さ全体にわたって100%一致する場合にのみ得られる。積スコア70は、100%の同一性で重畳が70%であるか、或いは88%の同一性で重畳が100%であるかの何れかの場合である。積スコア50は、100%の同一性で重畳が50%であるか、或いは79%の同一性で重畳が100%であるかの何れかの場合である。

40

【0299】

50

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

【0300】

8 HEMをコードするポリヌクレオチドの伸長

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

【0301】

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

【0302】

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

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

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

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

ステップ7 4 で保管。

【0303】

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

【0304】

伸長したヌクレオチドを脱塩及び濃縮してから384ウェルプレートに移し、CviJIコレラウィルスエンドヌクラーゼ(Molecular Biology Research, Madison WI)で消化し、pUC18ベクター(Amersham Pharmacia Biotech)に再連結する前に音波処理またはせん断を行った。ショットガンシーケンシングのために、消化したヌクレオチドを低濃度(0.6~0.8%)のアガロースゲル上に分離して断片を切断し、寒天をAgar ACE (Promega)で消化した。T4リガーゼ(New England Biolabs, Beverly MA)を用いて伸長したクローンをpUC18ベクター(Amersham Pharmacia Biotech)に再連結し、Pfu DNAポリメラーゼ(Stratagene)で制限部位の伸び出しを処理してコンピテント大腸菌細胞に形質移入した。形質移入した細胞を選択して抗生物質を含む培地に移し、それぞれのコロニーを切りとってLB/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。 20

【0305】

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

ステップ1 94 で3分間

ステップ2 94 で15秒 30

ステップ3 60 で1分間

ステップ4 72 で2分間

ステップ5 ステップ2、3、及び4を29回繰り返す

ステップ6 72 で5分間

ステップ7 4 で保管。

上記したようにPICO GREEN試薬(Molecular Probes)でDNAを定量化した。DNA回収率の悪いサンプルは、上記した条件で再び増幅した。サンプルを20%のジメチルサルホサイド(dimethylsulphoxide)(1:2, v/v)で希釈し、DYENAMIC DIRECTキット(Amersham Pharmacia Biotech)またはABI PRISM BIGDYE Terminator cycle sequencing ready reactionキット(Applied Biosystems)を用いてシーケンシングした。 40

【0306】

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

【0307】

9 個々のハイブリダイゼーションプローブの標識化及び使用法

SEQ ID NO: 27-52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 50

なるオリゴヌクレオチドの標識について特に記すが、より大きな cDNA フラグメントの場合でも基本的に同じ手順を用いる。オリゴヌクレオチドを、OLIGO 4.06 ソフトウェア (National Bioscience) のような最新式のソフトウェアを用いてデザインし、50 pmol の各オリゴマーと、250 μ Ci の [³²P] アデノシン三リン酸 (Amersham, Chicago, IL) 及び T4 ポリヌクレオチドキナーゼ (DuPont NEN, Boston MA) とを組み合わせて用いることにより標識する。標識されたオリゴヌクレオチドを、SEPHADEX G-25 超精細排除デキストランビードカラム (Amersham Pharmacia Biotech) を用いて実質的に精製する。毎分 10⁷ カウントの標識されたプローブを含むアリコットを、次のエンドヌクレアーゼ、Ase I、Bgl II、Eco RI、Pst I、Xba I あるいは Pvu II (DuPont NEN) の 1 つを用いて切断したヒトゲノム DNA の典型的な膜ベースのハイブリダイゼーション解析において用いる。

10

【0308】

各切断物からの DNA を、0.7% アガロースゲル上で分画して、ナイロン製メンブラン (Nytran Plus, Schleicher & Schuell, Durham NH) に転写する。ハイブリダイゼーションは 40°C で 16 時間かけて行う。非特異的シグナルを取り除くため、例えば、最大 0.1 M クエン酸ナトリウム食塩水及び 0.5% ドデシル硫酸ナトリウムの条件の下、プロットを順次室温にて洗浄する。ハイブリダイゼーションパターンをオートラジオグラフィー或いは別のイメージ化手段で視覚化して比較する。

20

【0309】

10 マイクロアレイ

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

30

【0310】

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

40

【0311】

組織または細胞サンプルの調製

50

グアニジウムチオシアネート法を用いて組織サンプルから全RNAを単離し、オリゴ(dT)セルロース法を用いてポリ(A)⁺RNAを精製する。各ポリ(A)⁺RNAサンプルは、MMLV逆転写酵素、0.05 pg/μlのオリゴ(dT)プライマー(21mer)、1×第1鎖緩衝液、0.03単位/μlのRNアーゼインヒビター、500 μM dATP、500 μM dGTP、500 μM dTTP、40 μM dCTP、40 μM dCTP-Cy3(BDS)またはdCTP-Cy5(Amersham Pharmacia Biotech)を用いて逆転写する。この逆転写反応は、GEMBRIGHTキット(Incyte)を用いて、200 ngのポリ(A)⁺RNAを含む25 ml容量で行う。特異的なコントロールポリ(A)⁺RNAは、in vitro転写により非コーディング酵母ゲノムDNAから合成する。370 で2時間インキュベートした後、各反応サンプル(一方はCy3標識、他方はCy5標識)は、2.5 mlの0.5 M 水酸化ナトリウムで処理し、850 で20分間インキュベートし、反応を停止させてRNAを変性する。サンプルは、2つの連続するCHROMA SPIN 30ゲル濾過スピンカラム(CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA)を用いて精製する。結合後、2つの反応サンプルを、1 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、Speed VAC(Savant Instruments Inc., Holbrook NY)を用いて乾燥して仕上げ、14 μl 5×SSC/0.2% SDS中で再懸濁する。

10

20

【0312】

マイクロアレイの準備

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

【0313】

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

30

【0314】

米国特許第5,807,522号に記載されている方法を用いて、コーティングしたガラス基板にアレイエレメントを付加する。この特許に引用することを以って本明細書の一部とする。平均濃度が100 ng/μlのアレイエレメントDNA 1 μlを高速機械装置により開放型キャピラリープリンティングエレメント(open capillary printing element)に充填する。次にこの装置が、スライド毎に約5 nlのアレイエレメントサンプルを分注する。

40

【0315】

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

50

【0316】

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、 $5 \times \text{SSC}$ 、 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 $0.2 \mu\text{g}$ 含む $9 \mu\text{l}$ のサンプル混合体を含めたものである。サンプル混合液を、 65°C で5分間加熱し、マイクロアレイ表面上に一定量分注してから 1.8 cm^2 のカバーガラスで覆う。このアレイを、顕微鏡スライドより僅かに大きいキャビティを有する防水チャンバーに移す。チャンバーの角に $140 \mu\text{l}$ の $5 \times \text{SSC}$ を加えて、チャンバー内を湿度 100% に保持する。このアレイを含むチャンバーを、 60°C 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 ($1 \times \text{SSC}$, 0.1% SDS) において 45°C で10分間、第2洗浄緩衝液中 ($0.1 \times \text{SSC}$) において 45°C で10分間それぞれ3回洗浄し、その後乾燥させる。

10

【0317】

検出

レポーター標識されたハイブリダイゼーション複合体は、Cy3を励起するための 488 nm 、及びCy3を励起するための 632 nm のスペクトル線を生成し得る Innova 70 混合ガス10 Wレーザー (Coherent, Inc., Santa Clara CA) を備えた顕微鏡で検出する。20倍の顕微鏡対物レンズ (Nikon, Inc., Melville NY) を用いて、アレイ上に励起レーザー光を集中させる。このアレイを含むスライドを顕微鏡のコンピュータ制御 X-Y ステージに置き、対物レンズを通してラスタスキャンする。本実施例で用いた $1.8 \text{ cm} \times 1.8 \text{ cm}$ のアレイは、 $20 \mu\text{m}$ の解像度でスキャンする。

20

【0318】

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

30

【0319】

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

40

【0320】

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、リニア 20 色変換を用いてシグナル強度が青色 (低シグナル) から赤色 (高シグナル) までの擬似カラー範囲にマッピングされるイメージとして表示される。データはまた、定量的に分析される。2つの異なる蛍光体を同時に励起して測定する場合には、各蛍光体の発光スペクトルを用いて、先ずデータは蛍光体間の光学的漏話 (重複発光スペクトルに起因する) に対して補正される。

【0321】

50

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

【0322】

1.1 相補的ポリヌクレオチド

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

【0323】

1.2 HEMの発現

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

【0324】

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

10

20

30

40

50

いる。これらの方法で精製したHEMを直接用いて以下の実施例16および17のアッセイを行うことができる。

【0325】

1.3 機能のアッセイ

HEMの機能は、哺乳動物細胞培養系において生理学的に高められたレベルでのHEMをコードする配列の発現によって評価する。cDNAを、cDNAを高いレベルで発現する強いプロモーターを含む哺乳動物発現ベクターにサブクローニングする。このようなベクターには、pCMVSPORTTM (Life Technologies.)及びpCR3.1 (Invitrogen, Carlsbad, CA)が含まれ、どちらもサイトメガロウイルスプロモーターを含んでいる。5~10 μ gの組換えベクターを、例えば内皮由来か造血由来のヒト細胞株にリポソーム製剤或いは電気穿孔法によって一時的に形質移入する。更に、標識タンパク質をコードする配列を含む1~2 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とを区別できる。また、標識タンパク質の発現によって、cDNAの組換えベクターからの発現を正確に予想できる。このような標識タンパク質には、緑色蛍光タンパク質(GFP; Clontech)、及びCD64またはCD64-GFP融合タンパク質が含まれる。レーザー光学に基づいた技術を利用した自動流動細胞計測法(FCM)を用いて、GFPまたはCD64-GFPを発現する形質移入された細胞を同定し、その細胞のアポトーシス状態や他の細胞特性を評価する。また、FCMで、先行した或いは同時の細胞死の現象を診断する蛍光分子の取り込みを検出して計量する。これらの現象には、プロビジウムヨウ化物でのDNAの染色によって計測される核DNA内容物の変化と、プロモデオキシウリジンの取り込み量の低下によって計測されるDNA合成の下方調節と、特異的な抗体との反応性によって計測される細胞表面及び細胞内のタンパク質の発現の変化と、蛍光複合アネキシンVタンパク質の細胞表面への結合によって計測される原形質膜組成の変化とが含まれる。流動細胞計測法は、Ormerod, M. G.による(1994) Flow Cytometry Oxford, New York, NY. に記載されている。

【0326】

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

【0327】

1.4 HEMに特異的な抗体の作製

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

【0328】

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

【0329】

通常、約15残基の長さのオリゴペプチドを、Applied BiosystemsのABI 431Aペプチドシンセサイザー(PE Biosystems)を用いてfmoc法のケミストリにより合成し、N-マレイミドベンゾイル-N-ヒドロキシスクシニミドエステル(MBS)を用いた反応によりKLH(Sigma-Aldrich, St. Louis MO)に結合させて、免疫原性を高める(例えば、前出のAusubel, 1995を参照)。フロイントの完全アジュバントにおいてオリゴペプチド-KLH複合体を用いてウサギを免疫化する。得られた抗血清の抗ペプチド活性及び抗HEM活性を検査するには、ペプチドまたはHEM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。

10

【0330】

15 特異的抗体を用いる天然HEMの精製

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

【0331】

HEMを含む培養液をイムノアフィニティークラムに通し、HEMを優先的に吸着できる条件で(例えば、界面活性剤の存在下において高イオン強度のバッファーで)そのカラムを洗浄する。そのカラムを、抗体とHEMとの結合を切るような条件で(例えば、pH2~3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、HEMを回収する。

20

【0332】

16 HEMと相互作用する分子の同定

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

30

【0333】

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

【0334】

HEMはまた、ハイスループット型の酵母2ハイブリッドシステムを使用する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)。

40

【0335】

17 HEM活性の実証

補助因子生合成酵素活性

HEMのmolybdopterinシンターゼスルフリラーゼ活性は、例えば、アスペルギルス ニデュランスcnxF変異体をHEMのポリヌクレオチド配列を含む適切な発現ベクターで形質転換し、アスペルギルス ニデュランス細胞が硝酸塩またはヒポキサン

50

チンを含む培地で成長できるか否かを観察して測定する（前出の Appleyard, M. らを参照）。molybdopterinsinターゼスルフリラーゼ活性がアスペルギルス ニデュランス cnx F 変異体に存在せず、この活性が硝酸塩またはヒポキサンチンを役立たせるために必要であるため、この変異体をHEMのポリヌクレオチド配列を含む適切な発現ベクターで救出することは、HEMがmolybdopterinsinターゼスルフリラーゼ活性を有することを意味する。また、molybdopterinsinおよび生合成経路における中間体のレベルを、HPLC分析で測定する（前出の Appleyard, M. らを参照）。

【0336】

イソメラーゼ活性

HEMのイソメラーゼ活性は、様々な特異的酵素活性によって実証される。そのうちの幾つかを以下に記載する。

【0337】

ペプチジルプロリルシス - トランスイソメラーゼ活性

HEMのペプチジルプロリルシス - トランスイソメラーゼ活性は、記載のようにアッセイすることができる（Rahfeld, J. U. ら（1994）FEBS Lett. 352 : 180 - 184）。このアッセイは、35 mM HEPESバッファ（pH 7.8、キモトリプシン（0.5 mg/ml）および様々な濃度のHEMを含む）で10で行う。このアッセイでは、基質は、4つの疎水性残基を含むペプチドである。このペプチドは、N末端にコハク酸基、C末端にnitroanilide基を含む。この基質は、プロリル結合に対して、80~95%がトランス構造で、5~20%がシス構造で平衡状態である。ジメチルスルホキシド（10 mg/ml）に溶解した基質のアリコット（2 μ l）を上記したように反応液に加える。基質のシス異性体のみが、キモトリプシンによる切断のための基質である。従って、基質がHEMにより異性化すると、生成物がキモトリプシンにより切断され4-nitroanilideが生成される。4-nitroanilideは、390 nmでその吸光度を測定して検出される。4-nitroanilideの生成は、時間およびHEMの濃度に依存する。

【0338】

別法では、HEMのペプチジルプロリルシス - トランスイソメラーゼ活性は、キモトリプシンを用いる結合アッセイで色素生産性ペプチドを用いて測定できる（Fischer, G. ら（1984）Biomed. Biochim. Acta 43 : 1101 - 1111）。

【0339】

チオレドキシシン活性

HEMチオレドキシシン活性は、記載のようにアッセイする（Luthman, M.（1982）Biochemistry 21 : 6628 - 6633）。チオレドキシシンは、ジスルフィド結合の形成を触媒し、細胞中のレドックス環境を調節してチオール：ジスルフィド交換を可能にする。チオール：ジスルフィド交換を測定する1つの方法は、混合液（0.1 Mリン酸カリウム、pH 7.0、2 mM EDTA、0.16 μ Mインスリン、0.33 mM DTT、および0.48 mM NADPH）におけるインスリンの減少を測定することである。様々な濃度のHEMを反応液に加えて、340 nMでNADPHの酸化をモニタリングして反応速度を測定する。

【0340】

トランスフェラーゼ活性

HEMトランスフェラーゼ活性は、供与体基質と受容体基質との間の標識したメチル基の転移を測定するメチルトランスフェラーゼアッセイによって測定する（Bokar, J. A. ら（1994）J. Biol. Chem. 269 : 17697 - 17704）。最終容量50 μ lの反応液は、15 mM HEPES、pH 7.9、1.5 mM $MgCl_2$ 、10 mMジチオスレイトール、3%ポリビニルアルコール、供与体基質（1.5 μ Ci [メチル-³H] AdoMet（0.375 μ M AdoMet）（

10

20

30

40

50

DuPont-NEN)、0.6 μg HEM、および受容体基質(最終濃度1 mMまでの6-メルカプトプリン(6-MP)または0.4 μg [³⁵S] RNA)を含む。反応液を30℃で30分間インキュベートし、次に65℃で5分間インキュベートする。生成物は、クロマトグラフィー若しくは電気泳動法で分離し、メチルトランスフェラーゼ活性のレベルを回収したメチル-³H-RNA若しくはメチル-³H-6MPを定量して測定する。

【0341】

ヒドロラーゼ活性

例を示す目的で、HEM分子のβ-グルコシダーゼ活性を測定するアッセイを記載する。様々な量のHEMを、50 mM酢酸ナトリウムバッファ(pH 5.0)において様々な時間(通常は1~5分間)、1 mM 4-nitrophenyl-D-glycopyranoside(基質)と共にインキュベートする。反応は、100℃で2分間加熱して停止させる。分光光度計を用いて410 nmで吸光度を測定する。この吸光度の変化がサンプル中のHEMの活性に比例する(Hrmova, M.ら(1998) J. Biol. Chem. 273: 11134-11143)。

【0342】

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

【0343】

(表の簡単な説明)

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

【0344】

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

【0345】

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

【0346】

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

【0347】

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。

【0348】

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

【0349】

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

【表1】

表1

インサイト プロジェクトID	ポリペプチド SEQ ID NO:	インサイト ポリペプチドID	ポリスクレオチド SEQ ID NO:	インサイト ポリスクレオチドID
2783442	1	2783442CD1	27	2783442CB1
2116390	2	2116390CD1	28	2116390CB1
5785224	3	5785224CD1	29	5785224CB1
1870996	4	1870996CD1	30	1870996CB1
138841	5	138841CD1	31	138841CB1
1485405	6	1485405CD1	32	1485405CB1
2024617	7	2024617CD1	33	2024617CB1
4721827	8	4721827CD1	34	4721827CB1
5406614	9	5406614CD1	35	5406614CB1
1252792	10	1252792CD1	36	1252792CB1
1299305	11	1299305CD1	37	1299305CB1
1385190	12	1385190CD1	38	1385190CB1
2133162	13	2133162CD1	39	2133162CB1
254567	14	254567CD1	40	254567CB1
3903488	15	3903488CD1	41	3903488CB1
5391816	16	5391816CD1	42	5391816CB1
5886989	17	5886989CD1	43	5886989CB1
723432	18	723432CD1	44	723432CB1
1332963	19	1332963CD1	45	1332963CB1
1559410	20	1559410CD1	46	1559410CB1
1752587	21	1752587CD1	47	1752587CB1
1911509	22	1911509CD1	48	1911509CB1
2210170	23	2210170CD1	49	2210170CB1
234664	24	234664CD1	50	234664CB1
2884114	25	2884114CD1	51	2884114CB1
4103559	26	4103559CD1	52	4103559CB1

10

20

30

40

【表2】

表2-1

ポリペプチド SEQ_ID_NO:	インサイト ポリペプチドID	GenBank ID NO:	確率スコア	GenBank 相同体
1	2783442CD1	g1825606	1.00E-130	no1ybdppter in 合成MOEBタンパク質に類似 [線虫] Wilson, R.ら (1994) Nature 368:32-38
2	2116390CD1	g2984222	1.2e-12	Fuculose-1-phosphate aldolase [Aquilifex aeolicus] Deckert, G.ら (1998) Nature 392:353-358
		g10177247	1.00E-55	エノラーゼ-ホスファターゼ 遺伝子[id:K1917.1]に類似性を有する [シロイヌナズナ] Nakamura, Y.ら (1997) DNA Res. 4:401-414
3	5785224CD1	g1799782	2e-61	胆汁酸誘導性オベロンタンパク質F [大腸菌] Yamamoto, Y.ら (1997) DNA Res. 4:91-113
4	1870996CD1	g1788642	1.00E-19	推算上の糖スクレオチドエピメララーゼ [大腸菌K12] Blattner, F.R.ら (1997) Science 277:1453-1474
5	138841CD1	g203106	0	β-アラニンシンターゼ [ドブネズミ] Kvalnes-Krick, K.L.ら (1993) J. Biol. Chem. 268:5686-5693
6	1485405CD1	g19137	6.4e-33	アスパラギナーゼ [Impinus arboreus] Lough, T.J.ら (1992) Plant Mol Biol 19:391-399
7	2024617CD1	g1196614	2.70E-278	L-グルタミンアミドヒドロラーゼ [ドブネズミ] Smith, E.M.ら (1990) J. Biol. Chem. 265:10631-10636
8	4721827CD1	g11546051	1.00E-115	dJ104916.1.1 (K1A1247 (グルコサミン-6- スルファターゼおよび K1A1077)に類似) [ヒト]

10

20

30

40

表2-2

ポリペプチド SEQ ID NO:	インサイト ポリペプチド ID	GenBank ID NO:	確率スコア	GenBank 相同体
9	5406614CD1	95524927	1.30E-161	DNA分解酵素 III (DNase III) [ヒト] Hoss, M. ら (1999) EMBO J. 18:3868-3875
10	1252792CD1	97576450	0	ALEX1 [ヒト] Kurochkin, I.V. ら (2001) Biochem. Biophys. Res. Commun. 280:340-347
11	1299305CD1	911545670	0	セリンβラクタマーゼ様タンパク質LACT-1 [ハツカネズミ]
15	3903488CD1	94103135	6.00E-66	推奨上の転写因子Fet5 [Schizosaccharomyces pombe] Shpakovskii, G.V. ら (1997) Bioorg. Khim. 23:234-237
16	5391816CD1	92330815	1.7e-86	プリンスクレオチド結合タンパク質 Fet5 [Schizosaccharomyces pombe]
18	723432CD1	96459813	2.20E-61	"エンド-1,3-1,4-β-D-グルカナーゼ [トウモロコシ] Thomas, B.R. ら (1998) Plant Physiol. 117:1525
19	1332963CD1	910173877	3.00E-82	ArgE/DapE/AcyIファミリーアミノシラーゼ タンパク質[マイココッカス・ラジオデュランス] White, O. ら (1999) Science 286:1571-1577
				(5-methylaminomethyl-2- thiouridylate)-methyl transferase [Bacillus halodurans] Takami, H. ら (1999) Extremophiles 3:21-28
		93328701	2.70E-69	PP-ループスーパーファミリー-ATPase [トラコーマ病原体] Stephens, R.S. ら (1998) Science 282:754-759

10

20

30

40

表2-3

ポリペプチド SEQ ID NO:	インサイト ポリペプチドID	GenBank ID NO:	確率スコア	GenBank 相同体
20	1559410CD1	g1943807	8.00E-57	ATPaseと低い類似性 [線虫] The C. elegans Genome Sequencing Consortium (1998) Science 282:2012- 2018
21	1752587CD1	g2225906	2.20E-68	PIG-L deacetylase [ドブネズミ] Nakamura, N. (1997) J. Biol. Chem. 272:15834-15840
22	1911509CD1	g1825778	1.10E-57	アセチルトランスフェラーゼに類似 [線虫] Wilson, R. (1994) Nature 368:32-38
23	2210170CD1	g2853081	5.00E-58	ATP結合蛋白質 [シロイヌナズナ]
24	234664CD1	g1101009	5.00E-26	エンドセリン変換酵素-2 [Bos taurus] Emoto, N. (1995) J. Biol. Chem. 270:15262-15268
25	2884114CD1	g2318131	3.00E-90	histone deacetylase [シロイヌナズナ]
		g6132878	0	histone deacetylase 8; HDAC8 [ヒト]
26	4103559CD1	g3098603	1.30E-53	katanin p60 サブユニット [Strongylocentrotus purpuratus]

10

20

30

40

表 3-1

SEQ ID NO.	インサイト ポリペプチド ID	アミノ酸 残基数	潜在的リソシ化部位	潜在的 グリコシル化部位	シグネチャ (signature) 配列、モチーフ、 及びドメイン	分析方法及び データベース
1	2783442CDL	404	S66 T152 S180 T201 S350 S381 T386 S393 T134 T243 S296	N150 N356	Thif ファミリー (細菌KeeB/Thif/HesAファミリ- 及びユビキチン活性化酵素E1) における反復ドメイン): R72-P221 molybdopterin 生合成 (MoeBタンパク質): PD035666: G222-I368 ユビキチン活性化酵素: DM00412 A49890 1-238: Y53-P232 クラスIIアドラーゼ及びAdducin N末端ドメイン アドラーゼ II: R25-V242 クラスIIアドラーゼ及びAdducin N末端ドメイン PF00596: T44-I50, I56-D77, R93-S118, V190-A207 タンパク質イソメラゼ亜鉛有差Adducin 細胞骨格調節カルモジュリン結合代替 I390: E30-A207	HMMER-PFAM BLAST-PRODOM BLAST-DMO HMMER-PFAM BLIMPS-PFAM BLAST-PRODOM
2	2116390CDL	242	Y57 Y216 S13 S49 S89 T106 S223 T148 S2 T136 T167 T209		タンパク質イソメラゼ亜鉛有差Adducin 細胞骨格調節カルモジュリン結合代替 I390: E30-A207	BLAST-PRODOM
3	5785224CDL	404	T155 S136 T192 S293 T301		タンパク質イソメラゼ亜鉛有差Adducin 細胞骨格調節カルモジュリン結合代替 I390: E30-A207 PD002990: P6-T127 セクシヨントンパク質アルシルCoAの転移酵素 PD150845: K262-E380 タンパク質イソメラゼ亜鉛有差Adducin 細胞骨格調節カルモジュリン結合代替 I390: E30-A207	BLAST-PRODOM
4	1870996CDL	126	T57 S113		シグナルペプチド: M1-A59	SPSCAN BLAST-PRODOM BLAST-DMO

【表 6】

10

20

30

40

表3-2

SEQ. ID NO.	インサイト ID	アミノ酸残基数	潜在的リン酸化部位	潜在的グリコシル化部位	シグネチャ(signature)配列、モチーフ、及びドメイン	分析方法及びデータベース
5	13884ICD1	384	T147 T194 S324 T359 S8 S48 S141 T299 Y57 Y312	N206	炭素-窒素加水分解酵素: I71-Y365 ニトリラーゼ/シアノ化物加水分解酵素: R80-A126, H198-I252 ニトリラーゼ脂肪族アミダーゼ: V89-A258 N-カルバミルアミド加水分解酵素: M1-I71	HMMER-PFAM BLIMPS-BLOCKS BLAST-PRODOM
6	1485405CD1	164	T141 S43 T71 S80		アスパラギナーゼ: M1-K164 加水分解酵素M4-前駆体: I81-A126 前駆体加水分解酵素シグナル: V5-H114 グリコシルアスパラギナーゼ鎖: L38-K152	HMMER-PFAM BLIMPS-PRODOM BLAST-PRODOM ELAST-DOMO
7	2024617CD1	602	S3 T50 S313 S444 T79 S189 S255 S313 S508 T580 T584 Y78 Y322	N308 N484	アンキリン反復: D518-R551, W552-Q585 アンキリン反復タンパク質: L523-L538, G553-Q562	HMMER-PFAM BLIMPS-PFAM
8	4721827CD1	434	S68 T344 S31 T47 S51 S68 S79 S83 T93 T174 T178 S198 Y208	N2, N186 N336 N346	加水分解酵素グルタミナーゼ: N383-L490 グルタミナーゼアイソフォーム: 肝臓: M68-L502 アンキリン反復: F504-E540	BLAST-PRODOM BLAST-DOMO
9	5406614CD1	369	S200 S222 S224 S245 S318 T366 T283 S293 T311		糖加水分解酵素ファミリー: D417-L423 N-アセチルグルコサミン-6-スルファターゼ前駆体: C329-G400	BLIMPS-PRINTS BLAST-PRODOM
10	1252792CD1	453	T114 S203 S328 T391 S408 T54 T114 S120 S166 S177 S366 S407 S408	N135 N273 N314 N353	イキソスクレアラゼ: I70-G78, V126-A145	BLIMPS-PFAM
11	1299305CD1	511	T436 S74 T231 T323 T367 Y200 T9 S74 T173 S268 S356 S439	N495	ATP/GTP 結合部位モチーフ (Pループ): G162-K168 シグナル切断: M1-A20 ロイシンジッパー: L365-L386 β-ラクタマーゼクラス-A、-B、-C、及J-D 活性部位: G106-V153 タンパク質前駆体加水分解酵素: E73-A211	MOTIFS SFSCAN MOTIFS PROFILESKAN ELAST-PRODOM

【表7】

10

20

30

40

表3-3

SEQ ID NO:	インサイトポリペプチドID	アミノ酸残基数	潜在的リン酸化部位	潜在型ペプチド	シグネチャ(signature)配列、モチーフ、及びドメイン	分析方法及びデータベース
12	1385190CD1	520	S510 S180 T215 T491 S117 S193 T228 S254 S277 T298 Y299 S355 S436	N30 N251	ATP/GTP 結合部位モチーフ (Pループ): G211-K217	MOTIFS
13	2133162CD1	687	S257 T632 S2 T242 T252 S256 T383 T402 T422 S431 S527 T547 S595 S608 T622 T641 S667 S40 T90 S117 S232 S234 S236 S501 S503 S518 S519 T552 S585	N88 N173 N283	ATP/GTP 結合部位モチーフ (Pループ): A615-K621	MOTIFS
14	254567CD1	190	S187 S188 T86	N168	ATP/GTP 結合部位モチーフ (Pループ): G9-T16	MOTIFS
15	3903488CD1	284	Y196 S14 S68 S148 S180 S208 T215 T174 S197 S202 S236 S272	N31 N47	CBX/GOXスーパーファミリーシグネチャ: H4-I19 ATP結合タンパク質: V5-Q186 ATP/GTP 結合部位モチーフ (Pループ): G10-S17	BLIMPS-PRINTS BLAST-PRODOM MOTIFS
16	5391816CD1	245	T191 Y165 S37 S223		シグナル切断: M1-A29 ATP NP 結合: M1-D268 エポキシド加水分解酵素シグネチャ: I128-I141	SFSCAN BLAST-PRODOM BLAST_DOMO BLIMPS-PRINTS
17	5866989CD1	231	S12 T99 T120 T166 T185 T87 T194 S149		ATP/GTP 結合部位モチーフ (Pループ): G8-K14 シグナル切断: M1-G13	MOTIFS SFSCAN
					フリジシキナーゼ: M1-I18	BLIMPS-PRINTS
					ATP結合3'領域: F2-G180	BLAST-PRODOM
					ATP NP 結合: M1-G180	BLAST_DOMO

【表 8】

10

20

30

40

SEQ ID NO.	インサイト ポリペプチド ID	アミノ酸 残基数	潜在的リジン酸化部位	潜在的 グリコシル化部位	シグネチャ (signature) 配列、モチーフ、及びドメイン	分析方法及びデータベース
18	723432CDI	475	T117 S130 T150 S168 S223 S235 S249 T270 S370 T423 S439 S446 Y377 S223 S299 S341 T362 T428 T183		TRPAS 残基を含む遺伝子領域: N157-Y456 β-トランスデューシンプアミリー-tp-asp 反復: A3-Q472	BLAST-PRODOM BLAST-DOMO
19	1332963CDI	421	S41 T131 S132 S169 S246 S406 S410 S169 T255 S284 T307 S380 Y58 S218 S178 S20	N153 N269	シグナルペプチド: M1-A20	SIGPEPT
20	1559410CDI	310		N300	ATP/GTP 結合部位モチーフ (Pループ): G16-T23 ロイシンジッパー: L81-L102 アデニル酸キナーゼシグネチャ: A13-C26 ATP結合HAMI-FM2遺伝子領域: I15-L182	BLAST-PRODOM MOTIFS
21	1752587CDI	194	S25 T87		ATP NP 結合: G11-R269 シグナルペプチド: M1-W20 シグナル切断: M1-G17 膜貫通ドメイン: M4-D23	BLAST-DOMO SIGPEPT SPSCAN TMAP
22	1911509CDI	349	S122 S134 S238 T244 S298 T49 T65 S73 T100 S115 S163 S310 S237	N303 N71 N265	無β加水分解酵素: R102-C345 α/β加水分解酵素ファミリー α/β加水分解酵素ファミリーシグネチャ: M101-S116, L149-162, S163-V176 加水分解酵素: K52-F181	BLAST-PRODOM BLAST-DOMO
23	2210170CDI	245	S14 T44 S74 S97 T238 S163 T209	N95	加水分解酵素: H70-P182 ATP/GTP 結合部位モチーフ (Pループ): G75-K81 シグナルペプチド: M1-A23 Hrp (ATP-結合) ファミリータンパク質: V66-L92, I101-V113, V154-V195 収束上のATP結合タンパク質: I165-V234	MOTIFS SIGPEPT BLIMPS-BLOCKS BLAST-PRODOM
					N15F/PRXC ファミリー: V66-M224	BLAST-DOMO

【表 9】

冊 5

SEQ ID NO:	インサイト ポリペプチド ID	アミノ酸 残基数	潜在的リン酸化部位	潜在的 グリコシル化部位	シグネチャ(signature)配列、モチーフ、 及びドメイン	分析方法及び データベース
24	234664CD1	255	S47 S194 T85 S244	N83	Hydrogen uptake protein: I62-Y74 エンドセリン変換加水分解酵素: M97-E160	BLIMPS-PRINTS BLAST-PRODOM
25	288414CD1	377	S26 S43 S63 Y174 T192 T311 S351 S39	N136	ヒストンデアセチラーゼファミリー: V17-G324 ヒストンデアセチラーゼファミリー: R65-I79, D101-G117, G139-I162, R171-M196, F208- Y240, I243-I284, I300-R313 ヒストンデアセチラーゼ: R37-G324	HMMER-PFAM BLIMPS-PFAM
26	4103559CD1	403	T7 T108 S215 T297 S330 S359 S372 S93 S94 T155 S173 S365 Y343 S174	N113 N316	PP3 ヒストンデアセチラーゼ/AQUOアセトイン利用: F35-V321 ATP/GTP 結合部位モチーフ (Pループ): G293-K299 AAA 様々な細胞活性に関連するATP分解酵素: G288-Q403 エンドペプチダーゼ LA (LON) S: G293-T312 ATP-依存性GTPプロテアーゼATP結合サブユニット: L289-A307 AAA-タンパク質ファミリータンパク質: W252-P272, W186-A307, A319-R361 ATP-結合プロテアーゼサブユニット相同体: T300-E402 AAA-タンパク質ファミリー: P248-W403	BLAST-PRODOM BLAST-PRODOM MOTIFS HMMER-PFAM BLIMPS-PRINTS BLIMPS-BLOCKS BLAST-PRODOM BLAST-PRODOM

10

20

30

40

【表 10】

表4-1

ポリスクレオチド SEQ ID NO:	インサイト ポリスクレオチドID	配列長さ	選択された断片	配列断片	5' 位置	3' 位置
27	2783442CB1	1478	1-537	3540084HL (SEMVNOT04)	1	290
				1619664T6 (BRAITUT13)	807	1478
				2084618R6 (PANCNOT04)	255	882
28	2116390CB1	1183	1-62	897793R1 (BRSTNOT05)	531	1142
				190070F1 (SYNORAB01)	404	1183
				776806R1 (COLNNOT05)	1	565
29	5785224CB1	1572	1-656	644802R6 (BRSTTUT02)	401	855
				5785224HL (FLERTXS07)	699	994
				2506827F6 (CONUTT01)	1	516
30	1870996CB1	1335	1-170, 255- 332	1212335T6 (BRSTTUT01)	929	1572
				1795181R6 (PROSTTUT05)	1331	1335
				1398049F1 (BRAITUT08)	802	1335
31	138841CB1	1341	1-46	SKAD90083V1	1	463
				2182959HL (SININOT01)	378	770
				SKAF00571V1	651	1335
32	1485405CB1	755	1-1688	3507430HL (CONCNOT01)	464	790
				138841R6 (LIVRNOT01)	1	480
				167939R6 (LIVRNOT01)	820	1341
33	2024617CB1	2390	1-1688	SCFA05578V1	515	1094
				SCFA04593D1	392	1061
				831524H1 (PROSTTUT04)	538	755
				1971451F6 (UCMCL5T01)	370	747
				3075952HL (BONRUF01)	1	292
				1970651X301D1 (UCMCL5T01)	213	684
				1444713T6 (THYRNOT03)	1785	2368
				92795888	1619	2375
				4144157F6 (BRSTTUT01)	983	1524
				9872924	1948	2390
				2026172X303D1 (KERANOT02)	705	1218
				2024617X304B1 (KERANOT02)	1794	2372
				2024617X302D1 (KERANOT02)	48	567
				2026172X308D1 (KERANOT02)	1	481
				2024617R6 (KERANOT02)	1356	1770
1444713T2 (THYRNOT03)	1746	2364				
G3431607	1942	2378				

10

20

30

40

表4-2

ポリヌクレオチド SEQ ID NO:	インサイト ポリヌクレオチドID	配列長さ	選択された断片	配列断片	5' 位置	3' 位置
34	4721827CBL	2940	1-265, 856- 2240	1730047T6 (BRSTTUT08) 5202590T6 (STOMNOT08) 1730047F6 (BRSTTUT08) 4935430HL (BRSTTUT20) 827193HL (PROSNOT06) 2107214R6 (BRAITUT03) 1819064F6 (PROSNOT20) 2844607T6 (DRGLNOT01) 1538938T1 (SINTTUT01) 5515657H1 (LIVRDIR01) 2844607F6 (DRGLNOT01) 1636138F6 (UTRSNOT06) 1291844F1 (PGANNOT03) 875716R1 (LUNGAST01) SEVA00533V1	1747 200 777 942 574 2558 1238 1686 2257 1 1174 930 265 557 1	2273 835 1222 1231 846 2940 1703 2268 2916 257 1686 1417 863 1173 455
35	5406614CBL	1417	1-840	1252792F6 (LUNGPEF03) 618039R6 (PGANNOT01) 1255336F2 (MENITUT03) 1571325F6 (UTRSNOT05) 2586114HL (BRAITUT22) 2688507HL (LUNGNOT23)	1599 854 210 1017 1 718	2133 1401 769 1624 243 961
36	1252792CBL	2133	1-863, 2111-2133	1299305F6 (BRSTNOT07) SXYA00951V1 1299305X349V1 (BRSTNOT07) 218730F1 (STCMNOT01) 1299305X352V1 (BRSTNOT07) SXYA02187V1 SXYA01827V1 3692261HL (HEANOT01)	783 249 1297 1256 1309 624 298 1	1306 772 1828 1826 1829 1046 798 301
37	1299305CBL	1829	1-750			

10

20

30

40

表4-3

ポリヌクレオチド SEQ ID NO:	インサイト ポリヌクレオチドID	配列長さ	選択された断片	配列断片	5'位置	3'位置				
38	1385190CB1	2323	1-1607	4560082H1 (KERATX01)	605	847				
				1464459R1 (PANCN04)	771	1300				
				866833R1 (BRAITUT03)	1746	2323				
				817831H1 (QVARTUT01)	492	737				
				4621649H1 (ENDVNOT01)	1421	1691				
				1436421F1 (PANCN08)	1380	1645				
				140593F1 (TYMINOR01)	1688	2310				
				4384773H1 (BRAVUT02)	1	247				
				476121R6 (MPLR2DT01)	107	614				
				2124443T6 (BRSTN07)	1660	2308				
				2622173H1 (KERAN02)	2367	2543				
				3605342H1 (LUNGNC03)	727	1042				
				040203R6 (TELYNOT01)	2018	2480				
39	2133162CB1	2592	1-107, 901- 1515, 2398- 2441	1573212F6 (LINDN03)	839	1282				
				1829765F6 (THPLAZT01)	2448	2981				
				960713H1 (BRSTTUT03)	2751	2992				
				2539058F6 (BONRTUT01)	1	766				
				2174569F6 (ENDCN03)	1216	1726				
				SASA00952F1	1378	1897				
				1573212T6 (LINDN03)	1733	2260				
				3878345H1 (SPLNNOF1)	1979	2268				
				3142967H1 (HNTAZZS07)	1	312				
				2192041H1 (THYRTUT03)	595	850				
				1397586T1 (BRAITUT08)	405	847				
				1859156T6 (PROSN0F8)	190	835				
				3240016H1 (COLAUC01)	1336	1472				
40	254567CB1	850	1-58, 1387- 1450	1353041T6 (LATRTUT02)	480	1019				
				1216572R1 (BRSTTUT01)	1	586				
				2286017T6 (BRALN0N01)	675	1375				
				827147R1 (PROSN0T06)	988	1561				
				1283691F6 (COLNNOF6)	904	1522				
				118023R1 (RUSCN0T01)	416	342				
				2608576F6 (EONTN0T01)	1	516				
				41	3903488CB1	1472	1-58, 1387- 1450	1353041T6 (LATRTUT02)	480	1019
								1216572R1 (BRSTTUT01)	1	586
				42	5391816CB1	1561	1-58, 1387- 1450	2286017T6 (BRALN0N01)	675	1375
								827147R1 (PROSN0T06)	988	1561

10

20

30

40

表4-4

ポリヌクレオチド SEQ ID NO:	インサイト ポリヌクレオチドID	配列長さ	選択された断片	配列断片	5'位置	3'位置
43	5886989CB1	1949	566-754, 1-245	1963212R6 (BRSTN0T04) 1504160F1 (BRAITUT07) 931121H1 (CRKVN0T01) SBR05080D1 3335260F6 (BRAIFET01) 870085R6 (LUNGAST01) 1985045R6 (LUNGAST01) 723432R1 (SYNCOAT01) 5637189H1 (UPRSNMR01) 1426405K13F1 (SINTBST01) 2072529X11C1 (LSLWNT01) SAFC01647F1 SAFC01486F1	1319 541 1714 655 1 1128 1629 816 1 275 142 1344 885	1749 1107 1949 1287 646 1726 2221 1394 262 842 687 1317 1418
44	723432CB1	2221	28-117	SEMA02284F1 859640R1 (BRAITUT03) 2955002F6 (KIDNFEF01) 1576265T1 (LNODN0T03) 1811269F6 (PROSTUT12) 4637615H1 (MYBPTXT01) 769274H1 (COLNCRU01) 1798960F6 (COLNCRU01) 3520838H1 (LUNGN0N03) 4062190H1 (BRAINOT21) 3593730H1 (FIBPN0T01) SCMA05260V1 098895H1 (PIPNOR01) 150073F6 (SINTBST01) SCMA05704V1 SCMA00926V1 4864824T6 (PROSTUT09) SCMA03952V1 SCMA04987V1	255 553 1 981 617 500 1333 960 33 358 1 870 1 1665 237 1317 1575 45 607	924 1133 518 1619 1147 740 1448 1447 364 529 207 1460 234 2225 719 1802 2203 497 1211
45	1332963CB1	1619	1-24, 735-1025			
46	1559410CB1	1448	1-227			
47	1752587CB1	2225	1-44, 562-1515			

10

20

30

40

【表 1 4】

表4-5

ポリスクレオチド SEQ ID NO:	インサイト ポリスクレオチドID	配列長さ	選択された断片	配列断片	5'位置	3'位置
48	1911509CBI	1400	1-28	2148490F6 (LUNGNOT18) 1404974F6 (LATRPTUT02) 1963768T6 (BRSHNOT04) 2149013H1 (BRALNOT09) 4897260H1 (LIVRPTUT12) 1911509F6 (CONNTPT01)	448 1197 726 1131 380 1	919 1400 1384 1393 689 445
49	2210170CBI	1361	739-1103	SBUA01233D1 2210170F6 (SINTFET03) SBUA00454D1 2210170T6 (SINTFET03)	1 500 895	840 372 993 1361
50	234664CBI	993	1-34	SXCA00325V1 SCLA05768V1 SCLA00227V1 2172996T6 (ENDCMOT03)	1 611 525 433	497 993 984 979
51	2884114CBI	1367	774-1367	2047322F6 (THPLT7F01) 4288169X302D1 (LIVRDIR01) 1446241F6 (PLACNOT02) 5616107F1 (THYMNOR02) 4794983H1 (LIVRPTUT09) 4288169H1 (LIVRDIR01) 91894065 3356420H1 (PROSTUT16) 3356420T6 (PROSTUT16) 92784609 3356420F6 (PROSTUT16)	166 1 979 949 611 800 356 1 587 868	652 381 1367 1268 870 1108 735 270 1274 1317 555
52	4103559CBI	1317	1-33, 76-355		1	270

10

20

30

40

表5

ポリヌクレオチド SEQ ID NO:	インサइट プロジェクトID	代表的ライブラリ
27	278342CB1	PANCTUT02
28	211639CB1	BLADTUT04
29	5785224CB1	CONUTUT01
30	1870996CB1	PROSTUT05
31	138841CB1	LIVRNUT01
32	1885405CB1	UCNCL5T01
33	2024617CB1	LIVRNUT04
34	4721827CB1	BRSTTUT08
35	5406614CB1	LUNGAST01
36	1252793CB1	BRALTUT01
37	1299305CB1	MMLRIDF01
38	1385190CB1	LUNGAST01
39	2133182CB1	UYMNUT06
40	254567CB1	PROGNUT18
41	3903488CB1	LUNGNOT09
42	5321816CB1	LUNGFET03
43	5886989CB1	CORPNUT02
44	723432CB1	BRALTUT02
45	1332963CB1	BRALTUT03
46	1559410CB1	PROSTUT12
47	1752587CB1	SIMFRST01
48	1911509CB1	SIMWNOT02
49	2210170CB1	SIMPFET03
50	234664CB1	KIDNTUT13
51	2884114CB1	THELT7T01
52	4103559CB1	PROSTUT16

10

20

30

40

【表 16】

ライブラリ	ヘクター	ライブラリの説明
BLADTUT04	PINCY	このライブラリは、60歳の白人男性から根治膀胱切除、前立腺切除、及び精管切除術の際に採取した膀胱腫瘍組織から単離したRNAを用いて作製した。病理学的には、グレード3の左膀胱壁移行性細胞癌であった。細胞癌 <i>in situ</i> がドーム及び三角で見つかった。患者には喫煙の習慣があった。家族歴には、1型糖尿病、胃の悪性腫瘍、アテローム硬化性冠動脈疾患、及び急性心筋梗塞があった。
BRAITUT01	PSPORT1	このライブラリは、50歳の白人女性から前頭葉切除の際に採取した脳腫瘍組織から単離したRNAを用いて作製した。病理学的には、果状壊死及び広汎性石灰化を伴う再発性のグレード3オリゴ星状細胞腫を示していた。患者の病歴には、言語障害及び癲癇が含まれていた。患者の脳は、放射線治療を受けており、その総照射量は 5,082cGy (フラクシオン 8) であった。家族歴には、脳腫瘍が含まれていた。
BRAITUT02	PSPORT1	このライブラリは、58歳の白人男性の前葉から脳髄膜の病巣切除の際に採取した脳腫瘍組織から単離したRNAを用いて作製した。病理学的には、グレード2転移性副腎腫を示していた。患者の既往症には、グレード2腎細胞癌、不眠症、及び慢性気道閉塞があった。家族歴には、腎臓の悪性新生物があった。
BRAITUT03	PSPORT1	このライブラリは、17歳の白人女性から脳髄膜病巣切除の際に左前葉から採取した脳腫瘍組織から単離したRNAを用いて作製した。病理学的には、グレード4の纖維状の巨細胞及び小細胞星状細胞腫を示していた。家族歴には、良性高血圧症及び脳血管障害が含まれていた。
BRSTTUT08	PINCY	このライブラリは、45歳の白人女性から片側拡大単純乳房切除中に採取した乳房腫瘍組織から単離したRNAを用いて作製した。病理学的には、浸潤性核グレード2~3の管型の腺癌であって、23のリンパ節の内3つが転移性であった。腫瘍体積の50%以上は、 <i>in situ</i> 面胞型及び非面胞型である。免疫染色では、エストロゲン/プロゲステロンレセプターに対して陽性であり、非関連組織は増殖の変化を示していた。患者は、腹式全子宮摘出術も同時に受けた。既往症には、置換にやらない僧帽弁形成術、リウマチ性僧帽弁閉鎖不全、及びリウマチ性心疾患があった。家族歴には、急性心筋梗塞、アテローム硬化性冠動脈疾患、及び11型糖尿病が含まれていた。

【表 17】

【表 1 8】

ライブラリ	ベクター	ライブラリの説明
CONUT01	pINCY	このライブラリは、61歳の白人女性の腹式全子宮摘出術および所属リンパ節切除を伴う両側卵管卵巣摘出術の際に採取したS状腸間膜腫瘍組織から単離したRNAを用いて作製した。病理学的には、S状腸間膜の2ヶ所における転移性グレード4の悪性ミューラー混合腫瘍であった。
CORPN02	pINCY	このライブラリは、アルツハイマー病で死亡した74歳の白人男性の脳から採取した病変脳梁組織から単離したRNAを用いて作製した。
KIDNTUT13	pINCY	このライブラリは、51歳の白人女性から腎尿管切除の際に採取した腎臓腫瘍組織から単離したRNAを用いて作製した。病理学的には、グレード3腎細胞癌を示していた。患者の既往歴には、抑うつ障害、低血糖症、及び子宮内臓症が含まれる。家族歴には、腎臓結石、結腸癌、及びII型糖尿病があった。
LIVRNOT01	PBLUESCRIPT	このライブラリは、49歳の男性の肝臓より単離したRNAを用いて、Stratageneで作製された。
LIVRTUT04	pINCY	このライブラリは、50歳の白人男性の部分肝切除の際に採取した腎臓腫瘍組織から単離したRNAを用いて作製した。病理学的には、グレード3のヘパトームであって、腫瘤を形成している。患者の既往歴には、良性高血圧症、及び肝炎があった。B型肝炎コア抗原及びB型肝炎表面抗原が患者に認められていた。
LUNCAST01	PSPORT1	このライブラリは、頭部外傷で死亡した17歳の白人男性の肺組織から単離したRNAを用いて作製した。患者の病歴には喘息があった。
LUNGFET03	pINCY	このライブラリは、妊娠20週目で死亡した白人女胎児の肺組織から単離したRNAを用いて作製した。
LUNGNOT09	pINCY	ライブラリは、23週目の白人男性胎児の肺組織より単離したRNAを用いて作製した。乳児は、超音波診断により乳児型嚢胞腎と診断された後に中絶された。
MHLRIDT01	PSPORT1	このライブラリは、男性及び女性から得たドナープールに由来する接着単核細胞から単離したRNAを用いて作製した。この細胞は、Ficol1-Hypaque 遠心分離の後24時間培養した。
PANCTUT02	pINCY	このライブラリは、45歳の白人女性から根治的痔十二指腸切除術の際に採取した痔腫瘍腫瘍組織から単離したRNAを用いて作製した。病理学的には、グレード4未分化癌を示していた。家族歴には、良性高血圧症、高脂血症、及びアテローム硬化性冠動脈疾患が含まれていた。

【表 19】

ライブラリ	ベクター	ライブラリの説明
PROSN0T18	pINCY	このライブラリは、58歳白人男性の根治前立腺切除及び根治前立腺切除、胃フィステル形成の際に採取した病変前立腺組織から単離したRNAを用いて作製した。病理学的には、腺線維筋腫性過形成であり、この組織はグレード3の移行性細胞癌に関連する。患者の既往歴には、口峽炎及び気腫があった。家族歴には、急性心筋梗塞、アテローム硬化性冠動脈疾患、及びびい型糖尿尿病があった。
PROSTUT05	PSPORT1	このライブラリは、69歳白人男性の根治前立腺切除の際に採取した前立腺腫瘍組織から単離したRNAを用いて作製した。病理学的には、腺癌 (Gleason グレード 3+4) であった。腺線維筋腫性過形成もあった。家族歴には、うつ病、血性心不全、多発性骨髄腫、高脂血症、リュウマチ様関節炎が含まれていた。
PROSTUT12	pINCY	このライブラリは、65歳白人男性の根治前立腺切除中に採取した前立腺腫瘍組織から単離したRNAを用いて作製した。病理学的には、腺癌 (Gleason グレード 2+2) を示していた。腺線維筋腫性過形成も示していた。患者には、前立腺特異的抗原 (PSA) の上昇が見られた。
PROSTUT16	pINCY	このライブラリは、55歳の白人男性から採取した前立腺腫瘍組織から単離したRNAを用いて作製した。病理学的には、Gleason グレード 5+4 の腺癌を示していた。また、腺線維筋腫性過形成もあった。患者には、前立腺特異的抗原 (PSA) の上昇が見られた。患者の既往歴には腎臓結石があった。家族歴には、肺癌及び乳癌が含まれていた。
SINTBST01	pINCY	このライブラリは、18歳の白人女性の腸吻合の際に回腸組織を採取し、その回腸組織から単離したRNAを用いて作製した。病理学的には、15cmの小腸を含む回腸のクロロノーム病である。家族歴には、脳血管疾患及びアテローム硬化性冠動脈疾患が含まれる。
SINTFET03	pINCY	このライブラリは、妊娠20週目で死亡した白人女性胎児から採取した小腸組織から単離したRNAを用いて作製した。
SINTNOT02	PBLUESCRIPT	このライブラリは、クモ膜下出血によって死亡した55歳の白人女性の小腸の小腸から単離したRNAを用いて作製した。血清学的には、サイトメガロウイルス (CMV) に陽性であった。患者は過去に子宮摘出術を受けていた。

冊 6-4

ライブラリ	ベクター	ライブラリの説明
THP1T7T01	pINCY	このライブラリは、インサイト社で開発されたインサイト社が所有権を有するT7増幅法で増幅した50,000の培養THP-1細胞から単離されたRNAを用いて作製した。THP-1 (ATCC TIB 202)は、急性単球性白血病の1歳の白人男児の末梢血液由来のヒト前単核細胞系である (Int. J. Cancer (1980) 26:171を参照)。アンチセンスRNA (aRNA) を、1回目のDNA合成の後にT7RNAポリメラーゼを用いて生成した。次に、aRNAをランダムプライミングして、EcoRIアダプターに結合した二本鎖cDNAを生成し、pINCYベクター(インサイト社)に無方向性にクローニングした。このライブラリの作製に用いた増幅技術は、配列存在量を極端に偏らせることなく、転写イメージング目的に使用できる。
TLYMN0T06	pINCY	このライブラリは、活性化されたTh2細胞から単離したポリA RNAを0.5μg用いて作製した。これらの細胞は、抗IL-12抗体及びB7導入COS細胞の存在下でIL4により誘導CD4 T細胞から分化し、抗CD3抗体及び抗CD28抗体で6時間活性化された。cDNA合成を、NotI-オリゴ(dT)プライマーを用いて開始した。二本鎖cDNAを平滑化してEcoRIアダプターに結合し、NotIで消化し、大ききさで選択し、pINCYベクター(インサイト社)のNotI部位及びEcoRI部位の中にクローニングした。
UCMCL5T01	PBLUESCRIPT	このライブラリは、12名の臍帯血から得られた単核細胞から単離されたRNAを用いて作製した。RNAをプールのライセートから単離する前に細胞を12日間IL-5と共に培養した。

10

20

30

【表 20】

表7-1

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

【表 2 1】

表7-2

プログラム名	説明	引用文献	パラメーター関連
ProfileScan	Prosites で定義された配列パターンと一致するタンパク質配列における構造及び配列のモチーフを検索するアルゴリズム。	Gribskov, M. 他 (1988) CABIOS 4:61-66; Gribskov, 他 (1989) Methods Enzymol. 183:146-159; Bairoch, A. 他 (1997) Nucleic Acids Res. 25:217-221	標準化された質のスコアを特定 の Prosites モチーフに対する GC 指定 "HIGH" 値 一般に、スコア=1.4-2.1
Phred	高い感度及び確率で自動配列決定機のトレースを調べる塩基読み出しアルゴリズムである。	Ewing, B. 他 (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	Smith-Waterman アルゴリズムの効率的なインプリメンテーションに基づく SWAT や GrossMatch を含む、Phrap Revised Assembly プログラムであって、配列相同性の検索及び DNA 配列の構築に有用である。	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; Green, P., University of Washington, Seattle, WA.	スコア=120 以上 一致長さ=56 以上
Consed	Phrap で構築したものの表示及び編集をするためのグラフィックツールである。	Gordon, D. 他 (1998) Genome Res. 8:195-202.	
SPScan	タンパク質配列をスキヤンして分泌シグナルペプチドの存在を調べる重み付けマトリクス解析プログラムである。	Nielson, H. 他 (1997) Protein Engineering 10: 1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	スコア=3.5 以上
TMAP	重み付けマトリクスを用いたタンパク質配列上の膜貫通セグメントの明確化および向きを決定するためのプログラム	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
THMMER	隠れマルコフモデル (HMM) を用いたタンパク質配列上の膜貫通セグメントの明確化および向きを決定するためのプログラム。	Sonnhammer, E.L. 他 (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow 他 eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	Prosites で定義された配列と一致したパターンについてアミノ酸配列を検索するプログラムである。	Bairoch 他、(1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

【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
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64896 A2(51) International Patent Classification: C12N 15/52,
9/00, 5/10, A01K 67/027, C12P 21/00, C07K 16/40, C12Q
1/68, G01N 33/57, C12Q 1/00CA 95125 (US); **RANDMAN, Olga** [US/US]; 366 Anna
Avenue, Mountain View, CA 94043 (US); **YUE, Henry**
[US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US);
AZIMZAI, Yalida [US/US]; 5518 Boulder Canyon Drive,
Castro Valley, CA 94552 (US); **J.A.L. Presto** [US/US]; P.O.
Box 5142, Santa Clara, CA 95056 (US); **BURFORD, Neil**
[GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US);
BALIGHN, Mariah, R. [US/US]; 14244 Santiago Road,
San Leandro, CA 94577 (US).

(21) International Application Number: PCT/US01/06806

(22) International Filing Date: 1 March 2001 (01.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).(30) Priority Data:
60/186,307 1 March 2000 (01.03.2000) US
60/192,532 28 March 2000 (28.03.2000) US
60/193,578 30 March 2000 (30.03.2000) US(81) Designated States (national): AR, AT, AU, AM, AI, AL,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KR, KG, KP, KZ, LC, LK, LR,
LS, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY,
NO, NZ, PL, PT, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(63) Related by continuation (CIN) or continuation-in-part
(CIP) to earlier applications:US 60/186,307 (CIP)
Filed on 1 March 2000 (01.03.2000)
US 60/192,532 (CIP)
Filed on 28 March 2000 (28.03.2000)
US 60/193,578 (CIP)
Filed on 30 March 2000 (30.03.2000)(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZW); Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM); European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR); OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NR, SN, TD, TG).(71) Applicant (for all designated States except US): **INCYTE
GENOMICS, INC.**, [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).Published:
without international search report and to be republished
upon receipt of that report

(72) Inventors; and

(75) Inventors/Applicants (for US only): **TANG, Y., Tom**
[US/US]; 4250 Ranvie Court, San Jose, CA 95118 (US);
LU, Dyuung, Aina, M. [US/US]; 233 Coy Drive, San Jose.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.

WO 01/64896 A2

(54) Title: HUMAN ENZYME MOLECULES

(57) Abstract: The invention provides human enzyme molecules (HEM) and polynucleotides which identify and encode HEM. 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 HEM.

WO 01/64896

PCT/US01/06806

HUMAN ENZYME MOLECULES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human enzyme molecules and to the use of these sequences in the diagnosis, treatment, and prevention of autoimmune/inflammation disorders, genetic disorders, neurological disorders, and cell proliferative disorders including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human enzyme molecules.

BACKGROUND OF THE INVENTION

The cellular processes of biogenesis and biodegradation involve a number of key enzyme classes including oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases, and cofactor biosynthetic enzymes. These enzyme classes are each comprised of numerous substrate-specific enzymes having precise and well regulated functions. These enzymes function by facilitating metabolic processes such as glycolysis, the tricarboxylic cycle, and fatty acid metabolism; synthesis or degradation of amino acids, steroids, phospholipids, alcohols, etc.; regulation of cell signalling, proliferation, inflammation, apoptosis, etc., and through catalyzing critical steps in DNA replication and repair, and the process of translation.

Oxidoreductases

Many pathways of biogenesis and biodegradation require oxidoreductase (dehydrogenase or reductase) activity, coupled to the reduction or oxidation of a donor or acceptor cofactor. Potential cofactors include cytochromes, oxygen, disulfide, iron-sulfur proteins, flavin adenine dinucleotide (FAD), and the nicotinamide adenine dinucleotides NAD and NADP (Newsholme, E.A. and Leech, A.R. (1983) Biochemistry for the Medical Sciences, John Wiley and Sons, Chichester, U.K. pp. 779-793). Reductase activity catalyzes the transfer of electrons between substrate(s) and cofactor(s) with concurrent oxidation of the cofactor. The reverse dehydrogenase reaction catalyzes the reduction of a cofactor and consequent oxidation of the substrate. Oxidoreductase enzymes are a broad superfamily of proteins that catalyze numerous reactions in all cells of organisms ranging from bacteria to plants to humans. These reactions include metabolism of sugar, certain detoxification reactions in the liver, and the synthesis or degradation of fatty acids, amino acids, glucocorticoids, estrogens, androgens, and prostaglandins. Different family members are named according to the direction in which their reactions are typically catalyzed; thus they may be referred to as oxidoreductases, oxidases, reductases, or dehydrogenases. In addition, family members often have distinct cellular localizations, including the cytosol, the plasma membrane, mitochondrial inner or outer membrane, and

WO 01/64896

PCT/US01/06806

peroxisomes.

Short-chain alcohol dehydrogenases (SCADs) are a family of dehydrogenases that only share 15% to 30% sequence identity, with similarity predominantly in the coenzyme binding domain and the substrate binding domain. In addition to the well-known role in detoxification of ethanol, SCADs are also involved in synthesis and degradation of fatty acids, steroids, and some prostaglandins, and are therefore implicated in a variety of disorders such as lipid storage disease, myopathy, SCAD deficiency, and certain genetic disorders. For example, retinol dehydrogenase is a SCAD-family member (Simon, A. et al. (1995) *J. Biol. Chem.* 270:1107-1112) that converts retinol to retinal, the precursor of retinoic acid. Retinoic acid, a regulator of differentiation and apoptosis, has been shown to down-regulate genes involved in cell proliferation and inflammation (Chai, X. et al. (1995) *J. Biol. Chem.* 270:3900-3904). In addition, retinol dehydrogenase has been linked to hereditary eye diseases such as autosomal recessive childhood-onset severe retinal dystrophy (Simon, A. et al. (1996) *Genomics* 36:424-430).

Propagation of nerve impulses, modulation of cell proliferation and differentiation, induction of the immune response, and tissue homeostasis involve neurotransmitter metabolism (Weiss, B. (1991) *Neurotoxicology* 12:379-386; Collins, S.M. et al. (1992) *Ann. N.Y. Acad. Sci.* 664:415-424; Brown, J.K. and Irmay, H. (1991) *J. Inher. Metab. Dis.* 14:436-458). Many pathways of neurotransmitter metabolism require oxidoreductase activity, coupled to reduction or oxidation of a cofactor, such as NAD⁺/NADH (Newsholme, E.A. and Leech, A.R. (1983) *Biochemistry for the Medical Sciences*, John Wiley and Sons, Chichester, U.K. pp. 779-793). Degradation of catecholamines (epinephrine or norepinephrine) requires alcohol dehydrogenase (in the brain) or aldehyde dehydrogenase (in peripheral tissue). NAD⁺-dependent aldehyde dehydrogenase oxidizes 5-hydroxyindole-3-acetate (the product of 5-hydroxytryptamine (serotonin) metabolism) in the brain, blood platelets, liver and pulmonary endothelium (Newsholme, E.A. and Leech, A.R. (supra) p. 786). Other neurotransmitter degradation pathways that utilize NAD⁺/NADH-dependent oxidoreductase activity include those of L-DOPA (precursor of dopamine, a neuronal excitatory compound), glycine (an inhibitory neurotransmitter in the brain and spinal cord), histamine (liberated from mast cells during the inflammatory response), and taurine (an inhibitory neurotransmitter of the brain stem, spinal cord and retina) (Newsholme, E.A. and Leech, A.R. supra, pp. 790, 792). Epigenetic or genetic defects in neurotransmitter metabolic pathways can result in a spectrum of disease states in different tissues including Parkinson disease and inherited myoclonus (McCance, K.L. and Huether, S.E. (1994) *Pathophysiology*, Mosby-Year Book, Inc., St. Louis, MO pp. 402-404; Gundlach, A.L. (1990) *EASEB J.* 4:2761-2766).

Tetrahydrofolate is a derivatized glutamate molecule that acts as a carrier, providing activated one-carbon units to a wide variety of biosynthetic reactions, including synthesis of purines,

WO 01/64896

PCT/US01/06806

- pyrimidines, and the amino acid methionine. Tetrahydrofolate is generated by the activity of a holoenzyme complex called tetrahydrofolate synthase, which includes three enzyme activities: tetrahydrofolate dehydrogenase, tetrahydrofolate cyclohydrolase, and tetrahydrofolate synthetase. Thus, tetrahydrofolate dehydrogenase plays an important role in generating building blocks for nucleic and amino acids, crucial to proliferating cells.
- 3-Hydroxyacyl-CoA dehydrogenase (3HACD) is involved in fatty acid metabolism. It catalyzes the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA, with concomitant oxidation of NAD to NADH, in the mitochondria and peroxisomes of eukaryotic cells. In peroxisomes, 3HACD and enoyl-CoA hydratase form an enzyme complex called bifunctional enzyme, defects in which are associated with peroxisomal bifunctional enzyme deficiency. This interruption in fatty acid metabolism produces accumulation of very-long chain fatty acids, disrupting development of the brain, bone, and adrenal glands. Infants born with this deficiency typically die within 6 months (Watkins, P. et al. (1989) *J. Clin. Invest.* 83:771-777; Online Mendelian Inheritance in Man (OMIM), #261515). The neurodegeneration that is characteristic of Alzheimer's disease involves development of extracellular plaques in certain brain regions. A major protein component of these plaques is the peptide amyloid- β (A β), which is one of several cleavage products of amyloid precursor protein (APP). 3HACD has been shown to bind the A β peptide, and is overexpressed in neurons affected in Alzheimer's disease. In addition, an antibody against 3HACD can block the toxic effects of A β in a cell culture model of Alzheimer's disease (Yan, S. et al. (1997) *Nature* 389:689-695; OMIM, #602057).
- Steroids, such as estrogen, testosterone, corticosterone, and others, are generated from a common precursor, cholesterol, and are interconverted into one another. A wide variety of enzymes act upon cholesterol, including a number of dehydrogenases. Steroid dehydrogenases, such as the hydroxysteroid dehydrogenases, are involved in hypertension, fertility, and cancer (Duax, W.L. and Ghosh, D. (1997) *Steroids* 62:95-100). One such dehydrogenase is 3-oxo-5- α -steroid dehydrogenase (OASD), a microsomal membrane protein highly expressed in prostate and other androgen-responsive tissues. OASD catalyzes the conversion of testosterone into dihydrotestosterone, which is the most potent androgen. Dihydrotestosterone is essential for the formation of the male phenotype during embryogenesis, as well as for proper androgen-mediated growth of tissues such as the prostate and male genitalia. A defect in OASD that prevents the conversion of testosterone into dihydrotestosterone leads to a rare form of male pseudohermaphroditism, characterized by defective formation of the external genitalia (Andersson, S., et al. (1991) *Nature* 354:159-161; Labrie, F., et al. (1992) *Endocrinology* 131:1571-1573; OMIM #264600). Thus, OASD plays a central role in sexual differentiation and androgen physiology.
- 17 β -hydroxysteroid dehydrogenase (17 β HSD6) plays an important role in the regulation of

WO 01/64896

PCT/US01/06806

the male reproductive hormone, dihydrotestosterone (DHTT). 17 β HSD6 acts to reduce levels of DHTT by oxidizing a precursor of DHTT, 3 α -diol, to androsterone which is readily glucuronidated and removed from tissues. 17 β HSD6 is active with both androgen and estrogen substrates when expressed in embryonic kidney 293 cells. At least five other isozymes of 17 β HSD have been identified that catalyze oxidation and/or reduction reactions in various tissues with preferences for different steroid substrates (Biswas, M.G. and Russell, D.W. (1997) *J. Biol. Chem.* 272:15959-15966). For example, 17 β HSD1 preferentially reduces estradiol and is abundant in the ovary and placenta. 17 β HSD2 catalyzes oxidation of androgens and is present in the endometrium and placenta. 17 β HSD3 is exclusively a reductive enzyme in the testis (Geissler, W.M. et al. (1994) *Nature Genet.* 7:34-39). An excess of androgens such as DHTT can contribute to certain disease states such as benign prostatic hyperplasia and prostate cancer.

Oxidoreductases are components of the fatty acid metabolism pathways in mitochondria and peroxisomes. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids, while the auxiliary pathway performs additional steps required for the degradation of unsaturated fatty acids. The auxiliary beta-oxidation enzyme 2,4-dienoyl-CoA reductase catalyzes the removal of even-numbered double bonds from unsaturated fatty acids prior to their entry into the main beta-oxidation pathway. The enzyme may also remove odd-numbered double bonds from unsaturated fatty acids (Koivuranta, K.T. et al. (1994) *Biochem. J.* 304:787-792; Smeland, T.E. et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6673-6677). 2,4-dienoyl-CoA reductase is located in both mitochondria and peroxisomes. Inherited deficiencies in mitochondrial and peroxisomal beta-oxidation enzymes are associated with severe diseases, some of which manifest themselves soon after birth and lead to death within a few years. Defects in beta-oxidation are associated with Reye's syndrome, Zellweger syndrome, neonatal adrenoleukodystrophy, infantile Refsum's disease, acyl-CoA oxidase deficiency, and bifunctional protein deficiency (Suzuki, Y. et al. (1994) *Am. J. Hum. Genet.* 54:36-43; Hoeffler, *supra*; Cotran, R.S. et al. (1994) *Robbins Pathologic Basis of Disease*, W.B. Saunders Co., Philadelphia, PA, p.866). Peroxisomal beta-oxidation is impaired in cancerous tissue. Although neoplastic human breast epithelial cells have the same number of peroxisomes as do normal cells, fatty acyl-CoA oxidase activity is lower than in control tissue (el Bouhtoury, F., et al. (1992) *J. Pathol.* 166:27-35). Human colon carcinomas have fewer peroxisomes than normal colon tissue and have lower fatty-acyl-CoA oxidase and bifunctional enzyme (including enoyl-CoA hydratase) activities than normal tissue (Cable, S., et al. (1992) *Virchows Arch. B Cell Pathol. Incl. Mol. Pathol.* 62:221-226). Another important oxidoreductase is isocitrate dehydrogenase, which catalyzes the conversion of isocitrate to α -ketoglutarate, a substrate of the citric acid cycle. Isocitrate dehydrogenase can be either NAD or NADP dependent, and is found in the cytosol, mitochondria, and peroxisomes. Activity of isocitrate dehydrogenase is regulated developmentally, and by

WO 01/64896

PCT/US01/06806

hormones, neurotransmitters, and growth factors.

Hydroxyypyruvate reductase (HPR), a peroxisomal 2-hydroxyacid dehydrogenase in the glycolate pathway, catalyzes the conversion of hydroxyypyruvate to glycerate with the oxidation of both NADH and NADPH. The reverse dehydrogenase reaction reduces NAD⁺ and NADP⁺. HPR
5 recycles nucleotides and has back into pathways leading to the synthesis of ATP and GTP. ATP and GTP are used to produce DNA and RNA and to control various aspects of signal transduction and energy metabolism. Inhibitors of purine nucleotide biosynthesis have long been employed as
antiproliferative agents to treat cancer and viral diseases. HPR also regulates biochemical synthesis of serine and cellular serine levels available for protein synthesis.

10 The mitochondrial electron transport (or respiratory) chain is a series of oxidoreductase-type enzyme complexes in the mitochondrial membrane that is responsible for the transport of electrons from NADH through a series of redox centers within these complexes to oxygen, and the coupling of this oxidation to the synthesis of ATP (oxidative phosphorylation). ATP then provides the primary
source of energy for driving a cell's many energy-requiring reactions. The key complexes in the
15 respiratory chain are NADH:ubiquinone oxidoreductase (complex I), succinate:ubiquinone oxidoreductase (complex II), cytochrome c₁-b oxidoreductase (complex III), cytochrome c oxidase (complex IV), and ATP synthase (complex V) (Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, Inc., New York, NY, p. 677-678). All of these complexes are located on
the inner matrix side of the mitochondrial membrane except complex II, which is on the cytosolic
20 side. Complex II transports electrons generated in the citric acid cycle to the respiratory chain. The electrons generated by oxidation of succinate to fumarate in the citric acid cycle are transferred through electron carriers in complex II to membrane bound ubiquinone (Q). Transcriptional
regulation of these nuclear-encoded genes appears to be the predominant means for controlling the
biogenesis of respiratory enzymes. Defects and altered expression of enzymes in the respiratory
25 chain are associated with a variety of disease conditions.

Other dehydrogenase activities using NAD as a cofactor are also important in mitochondrial function. 3-hydroxyisobutyrate dehydrogenase (3HBD), important in valine catabolism, catalyzes the
NAD-dependent oxidation of 3-hydroxyisobutyrate to methylmalonate semialdehyde within
mitochondria. Elevated levels of 3-hydroxyisobutyrate have been reported in a number of disease
30 states, including ketoacidosis, methylmalonic acidemia, and other disorders associated with deficiencies in methylmalonate semialdehyde dehydrogenase (Rougraff, P.M. et al. (1989) J. Biol. Chem. 264:5899-5903).

Another mitochondrial dehydrogenase important in amino acid metabolism is the enzyme isovaleryl-CoA-dehydrogenase (IVD). IVD is involved in leucine metabolism and catalyzes the
35 oxidation of isovaleryl-CoA to 3-methylcrotonyl-CoA. Human IVD is a tetrameric flavoprotein that

WO 01/64896

PCT/US01/06806

is encoded in the nucleus and synthesized in the cytosol as a 45 kDa precursor with a mitochondrial import signal sequence. A genetic deficiency, caused by a mutation in the gene encoding TVD, results in the condition known as isovaleric acidemia. This mutation results in inefficient mitochondrial import and processing of the TVD precursor (Vockley, J. et al. (1992) J. Biol. Chem. 267:2494-2501).

Transferases

Transferases are enzymes that catalyze the transfer of molecular groups. The reaction may involve an oxidation, reduction, or cleavage of covalent bonds, and is often specific to a substrate or to particular sites on a type of substrate. Transferases participate in reactions essential to such functions as synthesis and degradation of cell components, regulation of cell functions including cell signaling, cell proliferation, inflammation, apoptosis, secretion and excretion. Transferases are involved in key steps in disease processes involving these functions. Transferases are frequently classified according to the type of group transferred. For example, methyl transferases transfer one-carbon methyl groups, amino transferases transfer nitrogenous amino groups, and similarly denominated enzymes transfer aldehyde or ketone, acyl, glycosyl, alkyl or aryl, isoprenyl, saccharyl, phosphorous-containing, sulfur-containing, or selenium-containing groups, as well as small enzymatic groups such as Coenzyme A.

Acyl transferases include peroxisomal carnitine octanoyl transferase, which is involved in the fatty acid beta-oxidation pathway, and mitochondrial carnitine palmitoyl transferases, involved in fatty acid metabolism and transport. Choline O-acetyl transferase catalyzes the biosynthesis of the neurotransmitter acetylcholine.

Amino transferases play key roles in protein synthesis and degradation, and they contribute to other processes as well. For example, the amino transferase 5-aminolevulinic acid synthase catalyzes the addition of succinyl-CoA to glycine, the first step in heme biosynthesis. Other amino transferases participate in pathways important for neurological function and metabolism. For example, glutamine-phenylpyruvate amino transferase, also known as glutamine transaminase K (GTK), catalyzes several reactions with a pyridoxal phosphate cofactor. GTK catalyzes the reversible conversion of L-glutamine and phenylpyruvate to 2-oxoglutarate and L-phenylalanine. Other amino acid substrates for GTK include L-methionine, L-histidine, and L-tyrosine. GTK also catalyzes the conversion of kynurenine to kynurenic acid, a tryptophan metabolite that is an antagonist of the N-methyl-D-aspartate (NMDA) receptor in the brain and may exert a neuromodulatory function. Alteration of the kynurenine metabolic pathway may be associated with several neurological disorders. GTK also plays a role in the metabolism of halogenated xenobiotics conjugated to glutathione, leading to nephrotoxicity in rats and neurotoxicity in humans. GTK is

WO 01/64896

PCT/US01/06806

expressed in kidney, liver, and brain. Both human and rat GTKs contain a putative pyridoxal phosphate binding site (ExPASy ENZYME: EC 2.6.1.64; Perry, S.J. et al. (1993) *Mol. Pharmacol.* 43:660-665; Perry, S. et al. (1995) *FEBS Lett.* 360:277-280; and Alberati-Giani, D. et al. (1995) *J. Neurochem.* 64:1448-1455). A second amino transferase associated with this pathway is

5 kynurenine/ α -aminoadipate amino transferase (AadAT). AadAT catalyzes the reversible conversion of α -aminoadipate and α -ketoglutarate to α -ketoacid and L-glutamate during lysine metabolism. AadAT also catalyzes the transamination of kynurenine to kynurenic acid. A cytosolic AadAT is expressed in rat kidney, liver, and brain (Nakatani, Y. et al. (1970) *Biochim. Biophys. Acta* 198:219-228; Buchli, R. et al. (1995) *J. Biol. Chem.* 270:29330-29335).

10 Glycosyl transferases include the mammalian UDP-glucuronosyl transferases, a family of membrane-bound microsomal enzymes catalyzing the transfer of glucuronic acid to lipophilic substrates in reactions that play important roles in detoxification and excretion of drugs, carcinogens, and other foreign substances. Another mammalian glycosyl transferase, mammalian UDP-galactose-ceramide galactosyl transferase, catalyzes the transfer of galactose to ceramide in the synthesis of

15 galactocerebrosides in myelin membranes of the nervous system. The UDP-glycosyl transferases share a conserved signature domain of about 50 amino acid residues (PROSITE: PDOC00359, <http://expasy.hcuge.ch/spot/prosite.html>).

Methyl transferases are involved in a variety of pharmacologically important processes. Nicotinamide N-methyl transferase catalyzes the N-methylation of nicotinamides and other pyridines,

20 an important step in the cellular handling of drugs and other foreign compounds. Phenylethanolamine N-methyl transferase catalyzes the conversion of noradrenalin to adrenalin. 6-O-methylguanine-DNA methyl transferase reverses DNA methylation, an important step in carcinogenesis. Uroporphyrin-III C-methyl transferase, which catalyzes the transfer of two methyl groups from S-adenosyl-L-methionine to uroporphyrinogen III, is the first specific enzyme in the

25 biosynthesis of cobalamin, a dietary enzyme whose uptake is deficient in pernicious anemia. Protein-arginine methyl transferases catalyze the posttranslational methylation of arginine residues in proteins, resulting in the mono- and dimethylation of arginine on the guanidino group. Substrates include histones, myelin basic protein, and heterogeneous nuclear ribonucleoproteins involved in

mRNA processing, splicing, and transport. Protein-arginine methyl transferase interacts with

30 proteins upregulated by mitogens, with proteins involved in chronic lymphocytic leukemia, and with interferon, suggesting an important role for methylation in cytokine receptor signaling (Lin, W.-J. et al. (1996) *J. Biol. Chem.* 271:15034-15044; Abramovich, C. et al. (1997) *EMBO J.* 16:260-266; and Scott, H.S. et al. (1998) *Genomics* 48:330-340).

Phospho transferases catalyze the transfer of high-energy phosphate groups and are important

35 in energy-requiring and -releasing reactions. The metabolic enzyme creatine kinase catalyzes the

WO 01/64896

PCT/US01/06806

reversible phosphate transfer between creatine/creatine phosphate and ATP/ADP. Glycoylamine kinase catalyzes phosphate transfer from ATP to guanidoacetate, and arginine kinase catalyzes phosphate transfer from ATP to arginine. A cysteine-containing active site is conserved in this family (PROSITE: PDOC00103).

5 Prenyl transferases are heterodimers, consisting of an alpha and a beta subunit, that catalyze the transfer of an isoprenyl group. An example of a prenyl transferase is the mammalian protein farnesyl transferase. The alpha subunit of farnesyl transferase consists of 5 repeats of 34 amino acids each, with each repeat containing an invariant tryptophan (PROSITE: PDOC00703).

Saccharyl transferases are glycosylating enzymes involved in a variety of metabolic processes.

10 Oligosaccharyl transferase-48, for example, is a receptor for advanced glycation endproducts. Accumulation of these endproducts is observed in vascular complications of diabetes, macrovascular disease, renal insufficiency, and Alzheimer's disease (Thornalley, P.J. (1998) Cell Mol. Biol. (Noisy-Le-Grand) 44:1013-1023).

Coenzyme A (CoA) transferase catalyzes the transfer of CoA between two carboxylic acids.

15 Succinyl CoA:3-oxoacid CoA transferase, for example, transfers CoA from succinyl-CoA to a recipient such as acetoacetyl. Acetoacetyl is essential to the metabolism of ketone bodies, which accumulate in tissues affected by metabolic disorders such as diabetes (PROSITE: PDOC00980).

Hydrolases

20 Hydrolysis is the breaking of a covalent bond in a substrate by introduction of a water molecule. The reaction involves a nucleophilic attack by the water molecule's oxygen atom on a target bond in the substrate. The water molecule is split across the target bond, breaking the bond and generating two product molecules. Hydrolases participate in reactions essential to functions such as cell signaling, cell proliferation, inflammation, apoptosis, secretion and excretion. Hydrolases are involved in key steps in disease processes involving these functions. Hydrolases, or hydrolytic enzymes, may be grouped by substrate specificity into classes including aminohydrolases, phospholipases, carboxyl-esterases, phosphodiesterases, lysozymes, glycosidases, glyoxalases, sulfatases, phosphohydrolases, and serine hydrolases.

30 Phosphodiesterases catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. Phosphodiesterases are, therefore, crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endo- and exo-nucleases, which are essential to cell growth and replication as well as protein synthesis.

Pancreatic lipase and colipase form a complex that plays a key role in dietary fat digestion by converting insoluble long chain triacylglycerols into more polar molecules able to cross the brush border of intestinal cells. Colipase binds to the C-terminal domain of lipase.

WO 01/64896

PCT/US01/06806

Carboxylesterases are proteins that hydrolyze carboxylic esters and are classified into three categories- A, B, and C. Most type-B carboxylesterases are evolutionarily related and are considered to comprise a family of proteins. The type-B carboxylesterase family of proteins includes vertebrate acetylcholinesterase, mammalian liver microsomal carboxylesterase, mammalian bile-salt-activated lipase, and duck fatty acyl-CoA hydrolase. Some members of this protein family are not catalytically active but contain a domain related evolutionarily to other type-B carboxylesterases, such as thyroglobulin and *Drosophila* protein neuractin.

Acyl-CoA thioesterase is another member of the carboxylesterase family (Alexson, S.B., et al. (1993) *Eur. J. Biochem.* 214(3): 719-727). Evidence suggests that acyl-CoA thioesterase has a regulatory role in steroidogenic tissues (Finkielstein, C., et al. (1998) *Eur. J. Biochem.* 255(1): 60-66).

A phospholipase A₂ inhibitor has been identified that has 33% sequence homology with human leucine-rich α_2 -glycoprotein (Okumura, K., et al. (1998) *J. Biol. Chem.* 273(31): 19469-19475). Leucine-rich repeat (LRR) consensus sequences have also been found in the primary structure of many proteins, including proteins that participate in biologically important processes, such as receptors for hormones, enzymes, coenzyme inhibitors, proteins for cell adhesion, and ribosome-binding proteins. All proteins containing LRR domains are thought to be involved in protein-protein interactions.

Lysozyme c superfamily consists of conventional lysozymes c, calcium-binding lysozymes c, and α -lactalbumin (Prager, I.M. and Jolles, P. (1996) *EXS* 75: 9-31). The proteins in this superfamily have 35-40% sequence homology and share a common three dimensional fold, but can have different functions. Lysozymes bind and cleave the glycosidic bond linkage in sugars (Iyer, L.K. and Qasba, P.K. (1999) *Protein Eng.* 12(2): 129-139). Lysozymes c are ubiquitous in a variety of tissues and secretions and can lyse the cell walls of certain bacteria (McKenzie, H.A. (1996) *EXS* 75: 365-409).

The glyoxylase system consists of glyoxalase I, which catalyzes the formation of S-D-lactoylglutathione from methylglyoxal, a side product of triose-phosphate energy metabolism, and glyoxylase II, which hydrolyzes S-D-lactoylglutathione to D-lactic acid and reduced glutathione. Methylglyoxal levels are elevated during hyperglycemia, likely due to increased triose-phosphate energy metabolism. Elevated levels of glyoxylase II activity have been found in human and in a rat model of non-insulin-dependent diabetes mellitus. The glyoxylase system has been implicated in the detoxification of bacterial toxins, and in the control of cell proliferation and microtubule assembly. Elevated levels of S-D-lactoylglutathione, the substrate of glyoxylase II, induced growth arrest and toxicity in HL60 cells. Thus, the glyoxylase system, and glyoxylase II in particular, may be associated with cell proliferation and autoimmune disorders such as diabetes.

WO 01/64896

PCT/US01/06806

The alpha/beta hydrolase fold is a protein fold that is common to several hydrolases of diverse phylogenetic origin and catalytic functions. Enzymes with the alpha/beta hydrolase fold have a common core structure consisting of eight beta-sheets connected by alpha-helices. The best-conserved structural feature of this fold is the loops of the nucleophile-histidine-acid catalytic triad.

5 The histidine in the catalytic triad is completely conserved, while the nucleophile and acid loops accommodate more than one type of amino acid (Ollis, D.L., et al. (1992) *Protein Eng.* 5:197-211).

Sulfatases are members of a highly conserved gene family that share extensive sequence homology and a high degree of structural similarity. Sulfatases catalyze the cleavage of sulfate esters. To perform this function, sulfatases undergo a unique posttranslational modification in the endoplasmic reticulum that involves the oxidation of a conserved cysteine residue. A human disorder called multiple sulfatase deficiency is due to a defect in this posttranslational modification step, leading to inactive sulfatases (Recksieck, M., et al. (1998) *J. Biol. Chem.* 273(11): 6096-6103).

Phosphohydrolases are enzymes that hydrolyze phosphate esters. Some phosphohydrolases contain a MutT domain signature sequence. MutT is a protein involved in the GO system responsible for removing an oxidatively damaged form of guanine from DNA.

15 Glycosidases catalyze the cleavage of hemiacetyl bonds of glycosides, which are compounds that contain one or more sugar. Mammalian beta-galactosidase removes the terminal galactose from gangliosides, glycoproteins, and glycosaminoglycans. Beta-galactosidases belong to family 35 in the classification of glycosyl hydrolases.

20 Serine hydrolases are a functional class of hydrolytic enzymes that contain a serine residue in their active site. This class of enzymes contains proteinases, esterases, and lipases which hydrolyze a variety of substrates and, therefore, have different biological roles. Proteins in this superfamily can be further grouped into subfamilies based on substrate specificity or amino acid similarities (Puente, X.S. and Lopez-Ont, C. (1995) *J. Biol. Chem.* 270(21): 12926-12932).

25

Lyases

Lyases are a class of enzymes that catalyze the cleavage of C-C, C-O, C-N, C-S, C-(halide), P-O or other bonds without hydrolysis or oxidation to form two molecules, at least one of which contains a double bond (Stryer, L. (1995) *Biochemistry* W.H. Freeman and Co. New York, NY p.620). Lyases are critical components of cellular biochemistry with roles in metabolic energy production including fatty acid metabolism, as well as other diverse enzymatic processes. Further classification of lyases reflects the type of bond cleaved as well as the nature of the cleaved group.

30 The group of C-C lyases include carboxyl-lyases (decarboxylases), aldehyde-lyases (aldolases), oxo-acid-lyases and others. The C-O lyase group includes hydro-lyases, lyases acting on polysaccharides and other lyases. The C-N lyase group includes ammonia-lyases, amidine-lyases,

35

WO 01/64896

PCT/US01/06806

amino-lyases (deaminases) and other lyases.

Proper regulation of lyases is critical to normal physiology. For example, mutation induced deficiencies in the uroporphyrinogen decarboxylase can lead to photosensitive cutaneous lesions in the genetically-linked disorder familial porphyria cutanea tarda (Mendez, M. et al. (1998) Am. J. Genet. 63:1363-1375). It has also been shown that adenosine deaminase (ADA) deficiency stems from genetic mutations in the ADA gene, resulting in the disorder severe combined immunodeficiency disease (SCID) (Hershfield, M.S. (1998) Semin. Hematol. 35:291-298).

Isomerases

10 Isomerases are a class of enzymes that catalyze geometric or structural changes within a molecule to form a single product. This class includes racemases and epimerases, cis-trans-isomerases, intramolecular oxidoreductases, intramolecular transferases (mutases) and intramolecular lyases. Isomerases are critical components of cellular biochemistry with roles in protein folding, phototransduction, and metabolic energy production including glycolysis, as well as other diverse enzymatic processes (Stryer, L. (1995) Biochemistry, W.H. Freeman and Co. New York, NY pp.483-15 507).

Racemases are a subset of isomerases that catalyze inversion of a molecule's configuration around the asymmetric carbon atom in a substrate having a single center of asymmetry, thereby interconverting two racemers. Epimerases are another subset of isomerases that catalyze inversion of configuration around an asymmetric carbon atom in a substrate with more than one center of symmetry, thereby interconverting two epimers. Racemases and epimerases can act on amino acids and derivatives, hydroxy acids and derivatives, as well as carbohydrates and derivatives. The interconversion of UDP-galactose and UDP-glucose is catalyzed by UDP-galactose-4'-epimerase. Proper regulation and function of this epimerase is essential to the synthesis of glycoproteins and glycolipids. Elevated blood galactose levels have been correlated with UDP-galactose-4'-epimerase deficiency in screening programs of infants (Gitzelmann, R. (1972) Helv. Paediat. Acta 27:125-130). 20 25

The peptidyl prolyl *cis-trans* isomerases (PPIases) are a class of folding enzymes that isomerize certain proline imidic bonds in what is considered to be a rate limiting step in protein maturation and export. PPIases catalyze the *cis to trans* isomerization of certain proline imidic bonds 30 in proteins. There are three evolutionarily unrelated families of PPIases: the cyclophilins, the FK506 binding proteins, and the newly characterized parvulin family (Rahfeld, J.U. et al. (1994) FEBS Lett. 352: 180-184).

The cyclophilins (Cyp) were originally identified as major receptors for the immunosuppressive drug cyclosporin A (CsA), an inhibitor of T-cell activation (Handschumacher, R.E. et al. (1984) Science 226: 544-547; Harding, M.W. et al. (1986) J. Biol. Chem. 261: 8547-8555). 35

WO 01/64896

PCT/US01/06806

Thus, the peptidyl-prolyl isomerase activity of Cyp may be part of the signaling pathway that leads to T-cell activation. Subsequent work demonstrated that Cyp's isomerase activity is essential for correct protein folding and/or protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein activity. For example, in *Drosophila*, the Cyp NinaA is required for correct localization of rhodopsins, while a mammalian Cyp (Cyp40) is part of the Hsp90/Hsp70 complex that binds steroid receptors. The mammalian Cyp (CypA) has been shown to bind the gag protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is reversed by cyclosporin. This effect implicates Cyp in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) J. Biol. Chem. 266:23204-23214; Hunter, T. (1998) Cell 92: 141-143; and Leverson, J.D. and Ness, S.A. (1998) Mol. Cell. 1:203-211).

Another class of folding enzymes are the protein disulfide isomerases. One of the major rate limiting steps in protein folding is the thiol:disulfide exchange that is necessary for correct protein assembly. Although incubation of reduced, unfolded proteins in buffers with defined ratios of oxidized and reduced thiols can lead to native conformation, the rate of folding is slow and the attainment of native conformation decreases proportionately with the size and number of cysteines in the protein. Certain cellular compartments such as the endoplasmic reticulum of eukaryotes and the periplasmic space of prokaryotes are maintained in a more oxidized state than the surrounding cytosol. Correct disulfide formation can occur in these compartments, but at a rate that is insufficient for normal cell processes and inadequate for synthesizing secreted proteins. The protein disulfide isomerases, thioredoxins and glutaredoxins are able to catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges (Lofgren, H. (1995) J. Biol. Chem. 270:26178-26183).

Each of these proteins has somewhat different functions, but all belong to a group of disulfide-containing redox proteins that contain a conserved active-site sequence and are ubiquitously distributed in eukaryotes and prokaryotes. Protein disulfide isomerases are found in the endoplasmic reticulum of eukaryotes and in the periplasmic space of prokaryotes. They function by exchanging their own disulfide for a thiol in a folding peptide chain. In contrast, the reduced thioredoxins and glutaredoxins are generally found in the cytoplasm and function by directly reducing disulfides in the substrate proteins.

These catalytic molecules not only facilitate disulfide formation, but also regulate and participate in a wide variety of physiological processes. The thioredoxin system serves, for example, as a hydrogen donor for ribonucleotide reductase and as a regulator of enzymes by redox control. It

WO 01/64896

PCT/US01/06806

also modulates the activity of transcription factors such as NF- κ B, AP-1, and steroid receptors. More recently, several cytokines or secreted cytokine-like factors such as adult T-cell leukemia-derived factor, 3B6-interleukin-1, T-hybridoma-derived (MP-6) B cell stimulatory factor, and early pregnancy factor have been reported to be identical to thioredoxin (Holmgren, A. (1985) *Annu. Rev. Biochem.* 54:237-271; Abate, C. et al., (1990) *Science* 249:1157-1161; Tagaya, Y. et al. (1989) *EMBO J.* 8:757-764; Wakasugi, H. (1987) *Proc. Natl. Acad. Sci. USA* 84:804-808; Rosen, A. et al. (1995) *Int. Immunol.* 7:625-633). Thioredoxin has also been shown to have many extracellular activities including a role as a regulator of cell growth and a mediator in the immune system (Miranda-Vizuete, A. et al. (1996) *J. Biol. Chem.* 271:19099-19103; Yamauchi, A. et al (1992) *Mol. Immunol.* 29:263-270).

Oxidoreductases can be isomerases as well. Oxidoreductases catalyze the reversible transfer of electrons from a substrate that becomes oxidized to a substrate that becomes reduced. This class of enzymes includes dehydrogenases, hydroxylases, oxidases, oxygenases, peroxidases, and reductases. Proper maintenance of oxidoreductase levels is physiologically important. The pentose phosphate pathway for example, utilizes enzymes which are responsible for generating the reducing agent NADPH, while at the same time oxidizing glucose-6-phosphate to ribose-5-phosphate. NADPH serves as the fuel for reactions undergoing reductive biosynthesis. Ribose-5-phosphate and its derivatives become part of critical biological molecules such as ATP, Coenzyme A, NAD⁺, FAD, RNA, and DNA. The pentose phosphate pathway has both oxidative and non-oxidative branches. The oxidative branch steps, which are catalyzed by the enzymes glucose-6-phosphate dehydrogenase, lactonase, and 6-phosphogluconate dehydrogenase, convert glucose-6-phosphate and NADP⁺ to ribulose-6-phosphate and NADPH. The non-oxidative branch steps, which are catalyzed by the enzymes phosphopentose isomerase, phosphopentose epimerase, transketolase, and transaldolase, allow the interconversion of three-, four-, five-, six-, and seven-carbon sugars.

Another subgroup of isomerases are the transferases (or mutases). Transferases transfer a chemical group from one compound (the donor) to another compound (the acceptor). The types of groups transferred by these enzymes include acyl groups, amino groups, phosphate groups (phosphotransferases or phosphomutases), and others. The transferase carnitine palmitoyltransferase is an important component of fatty acid metabolism. Genetically-linked deficiencies in this transferase can lead to myopathy (Scriver C.R. et al. (1995) *The Metabolic and Molecular Basis of Inherited Disease*, McGraw-Hill New York NY pp.1501-1533).

Yet another subgroup of isomerases are the topoisomerases. Topoisomerases are enzymes that affect the topological state of DNA. For example, defects in topoisomerases or their regulation can affect normal physiology. Reduced levels of topoisomerase II have been correlated with some of the DNA processing defects associated with the disorder ataxia-telangiectasia (Singh, S.P. et al.

WO 01/64896

PCT/US01/06806

(1988) Nucleic Acids Res. 16:3919-3929).

Ligases

Ligases catalyze the formation of a bond between two substrate molecules. The process
5 involves the hydrolysis of a pyrophosphate bond in ATP or a similar energy donor. Ligases are
classified based on the nature of the type of bond they form, which can include carbon-oxygen,
carbon-sulfur, carbon-nitrogen, carbon-carbon and phosphoric ester bonds.

Ligases forming carbon-oxygen bonds include the aminoacyl-transfer RNA (tRNA)
synthetases which are important RNA-associated enzymes with roles in translation. Protein
10 biosynthesis depends on each amino acid forming a linkage with the appropriate tRNA. The
aminoacyl-tRNA synthetases are responsible for the activation and correct attachment of an amino
acid with its cognate tRNA. The 20 aminoacyl-tRNA synthetase enzymes can be divided into two
structural classes, and each class is characterized by a distinctive topology of the catalytic domain.
Class I enzymes contain a catalytic domain based on the nucleotide-binding Rossmann 'fold'. Class II
15 enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel β -sheet
motif, as well as N- and C-terminal regulatory domains. Class II enzymes are separated into two
groups based on the heterodimeric or homodimeric structure of the enzymes; the latter group is further
subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and Cusack,
S. (1995) J. Mol. Evol. 40:519-530). Autoantibodies against aminoacyl-tRNAs are generated by
20 patients with dermatomyositis and polymyositis, and correlate strongly with complicating interstitial
lung disease (ILD). These antibodies appear to be generated in response to viral infection, and
coxsackie virus has been used to induce experimental viral myositis in animals.

Ligases forming carbon-sulfur bonds (Acid-thiol ligases) mediate a large number of cellular
biosynthetic intermediary metabolism processes involve intermolecular transfer of carbon
25 atom-containing substrates (carbon substrates). Examples of such reactions include the tricarboxylic
acid cycle, synthesis of fatty acids and long-chain phospholipids, synthesis of alcohols and aldehydes,
synthesis of intermediary metabolites, and reactions involved in the amino acid degradation
pathways. Some of these reactions require input of energy, usually in the form of conversion of ATP
to either ADP or AMP and pyrophosphate.

In many cases, a carbon substrate is derived from a small molecule containing at least two
carbon atoms. The carbon substrate is often covalently bound to a larger molecule which acts as a
carbon substrate carrier molecule within the cell. In the biosynthetic mechanisms described above,
the carrier molecule is coenzyme A. Coenzyme A (CoA) is structurally related to derivatives of the
nucleotide ADP and consists of 4'-phosphopantetheine linked via a phosphodiester bond to the alpha
35 phosphate group of adenosine 3',5'-bisphosphate. The terminal thiol group of 4'-phosphopantetheine

WO 01/64896

PCT/US01/06806

acts as the site for carbon substrate bond formation. The predominant carbon substrates which utilize CoA as a carrier molecule during biosynthesis and intermediary metabolism in the cell are acetyl, succinyl, and propionyl moieties, collectively referred to as acyl groups. Other carbon substrates include enoyl lipid, which acts as a fatty acid oxidation intermediate, and carnitine, which acts as an acetyl-CoA flux regulator/ mitochondrial acyl group transfer protein. Acyl-CoA and acetyl-CoA are synthesized in the cell by acyl-CoA synthetase and acetyl-CoA synthetase, respectively.

Activation of fatty acids is mediated by at least three forms of acyl-CoA synthetase activity:

- i) acetyl-CoA synthetase, which activates acetate and several other low molecular weight carboxylic acids and is found in muscle mitochondria and the cytosol of other tissues;
 - ii) medium-chain acyl-CoA synthetase, which activates fatty acids containing between four and eleven carbon atoms (predominantly from dietary sources), and is present only in liver mitochondria; and
 - iii) acyl CoA synthetase, which is specific for long chain fatty acids with between six and twenty carbon atoms, and is found in microsomes and the mitochondria. Proteins associated with acyl-CoA synthetase activity have been identified from many sources including bacteria, yeast, plants, mouse, and man.
- The activity of acyl-CoA synthetase may be modulated by phosphorylation of the enzyme by cAMP-dependent protein kinase.

Ligases forming carbon-nitrogen bonds include amide synthetases such as glutamine synthetase (glutamate-ammonia ligase) that catalyzes the amination of glutamic acid to glutamine by ammonia using the energy of ATP hydrolysis. Glutamine is the primary source for the amino group in various amide transfer reactions involved in *de novo* pyrimidine nucleotide synthesis and in purine and pyrimidine ribonucleotide interconversions. Overexpression of glutamine synthetase has been observed in primary liver cancer (Christa, L. et al. (1994) *Gastroent.* 106:1312-1320).

Acid-amino acid ligases (peptide synthetases) are represented by the ubiquitin proteases which are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. In the UCS pathway, proteins targeted for degradation are conjugated to a ubiquitin (Ub), a small heat stable protein. Ub is first activated by a ubiquitin-activating enzyme (E1), and then transferred to one of several Ub-conjugating enzymes (E2). E2 then links the Ub molecule through its C-terminal glycine to an internal lysine (acceptor lysine) of a target protein. The ubiquitinated protein is then recognized and degraded by proteasome, a large, multisubunit proteolytic enzyme complex, and ubiquitin is released for reutilization by ubiquitin protease. The UCS is implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins

WO 01/64896

PCT/US01/06806

(Cicchiover, A. (1994) Cell 79:13-21). A murine proto-oncogene, Unp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells, and the human homolog of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) Oncogene 10:2179-2183).

5 Cyclo-ligases and other carbon-nitrogen ligases comprise various enzymes and enzyme complexes that participate in the de novo pathways to purine and pyrimidine biosynthesis. Because these pathways are critical to the synthesis of nucleotides for replication of both RNA and DNA, many of these enzymes have been the targets of clinical agents for the treatment of cell proliferative disorders such as cancer and infectious diseases.

10 Purine biosynthesis occurs de novo from the amino acids glycine and glutamine, and other small molecules. Three of the key reactions in this process are catalyzed by a trifunctional enzyme composed of glycinamide-ribonucleotide synthetase (GARS), aminoimidazole ribonucleotide synthetase (AIRS), and glycinamide ribonucleotide transformylase (GART). Together these three enzymes combine ribosylamine phosphate with glycine to yield phosphoribosyl aminoimidazole, a precursor to both adenylate and guanylate nucleotides. This trifunctional protein has been implicated in the pathology of Down's syndrome (Aimi, J. et al. (1990) Nucleic Acid Res. 18:6665-6672). Adenylosuccinate synthetase catalyzes a later step in purine biosynthesis that converts inosinic acid to adenylosuccinate, a key step on the path to ATP synthesis. This enzyme is also similar to another carbon-nitrogen ligase, argininosuccinate synthetase, that catalyzes a similar reaction in the urea cycle (Powell, S.M. et al. (1992) FEBS Lett. 303:4-10).

Like the de novo biosynthesis of purines, de novo synthesis of the pyrimidine nucleotides uridylylate and cytidylylate also arises from a common precursor, in this instance the nucleotide orotidylylate derived from orotate and phosphoribosyl pyrophosphate (PRPP). Again a trifunctional enzyme comprising three carbon-nitrogen ligases plays a key role in the process. In this case the enzymes aspartate transcarbamylase (ATCase), carbamyl phosphate synthetase II, and dihydroorotase (DHOase) are encoded by a single gene called CAD. Together these three enzymes combine the initial reactants in pyrimidine biosynthesis, glutamine, CO₂ and ATP to form dihydroorotate, the precursor to orotate and orotidylylate (Iwahana, H. et al. (1996) Biochem. Biophys. Res. Commun. 219:249-255). Further steps then lead to the synthesis of uridine nucleotides from orotidylylate.

30 Cytidine nucleotides are derived from uridine-5'-triphosphate (UTP) by the amidation of UTP using glutamine as the amino donor and the enzyme CTP synthetase. Regulatory mutations in the human CTP synthetase are believed to confer multi-drug resistance to agents widely used in cancer therapy (Yamauchi, M. et al. (1990) EMBO J. 9:2095-2099).

Ligases forming carbon-carbon bonds include the carboxylases acetyl-CoA carboxylase and pyruvate carboxylase. Acetyl-CoA carboxylase catalyzes the carboxylation of Acetyl-CoA from CO₂

WO 01/64896

PCT/US01/06806

and H₂O using the energy of ATP hydrolysis. Acetyl-CoA carboxylase is the rate-limiting step in the biogenesis of long-chain fatty acids. Two isoforms of Acetyl-CoA carboxylase, types I and types II, are expressed in human in a tissue-specific manner (Ha, J. et al. (1994) *Eur. J. Biochem.* 219:297-306). Pyruvate carboxylase is a nuclear-encoded mitochondrial enzyme that catalyzes the conversion of pyruvate to oxaloacetate, a key intermediate in the citric acid cycle.

Ligases forming phosphoric ester bonds include the DNA ligases involved in both DNA replication and repair. DNA ligases seal phosphodiester bonds between two adjacent nucleotides in a DNA chain using the energy from ATP hydrolysis to first activate the free 5'-phosphate of one nucleotide and then react it with the 3'-OH group of the adjacent nucleotide. This resealing reaction is used in both DNA replication to join small DNA fragments called "Okazaki" fragments that are transiently formed in the process of replicating new DNA, and in DNA repair. DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Bloom's syndrome is an inherited human disease in which individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts, B. et al. (1994) *The Molecular Biology of the Cell*, Garland Publishing Inc., New York, NY, p. 247).

Cofactor Biosynthetic Enzymes

Cofactors, including coenzymes and prosthetic groups, are small molecular weight inorganic or organic compounds that are required for the action of an enzyme. One particular cofactor, the molybdenum-containing cofactor molybdopterin, is an ubiquitous molecule which is required for the activity of a variety of enzymes, including nitrate reductase, sulfite oxidase, xanthine dehydrogenase, and aldehyde oxidase.

Molybdopterin biosynthesis is performed via a two step reaction pathway. First, a guanosine derivative is converted to an intermediate called precursor Z. Precursor Z is then converted to molybdopterin. In humans, the MOCS1 transcript encodes two enzymes, MOCS1 A and MOCS1 B, which are involved in the first step. MOCS2 is also a single transcript which encodes the two subunits of molybdopterin synthase, the enzyme catalyzing the second step, in two overlapping reading frames (Reiss, J. et al. (1999) *Am. J. Hum. Genet.* 64:706-11). In *Aspergillus nidulans*, CnxABC catalyzes the first step in the molybdopterin biosynthesis pathway. The second step in *A. nidulans* is catalyzed by molybdopterin synthase, as it is in humans. In addition, a converting factor, CnxF, has also been discovered in *A. nidulans* (Appleyard, M. et al., (1998) *J. Biol. Chem.* 273:14869-14876). CnxF is similar to *E. coli* MoeB, an enzyme which transfers sulfur atoms to the synthase and makes it capable of adding the dithiolene group to precursor Z. CnxF is thought to mediate the same reaction in *A. nidulans*. CnxF is also similar to ThiF, an enzyme required for

WO 01/64896

PCT/US01/06806

thiamin biosynthesis; HesA, which is involved in heterocyst formation; and the eukaryotic ubiquitin-activating protein E1. However, no obvious physiological relationship exists between CnxI⁺ and these latter proteins.

A deficiency in molybdopterin biosynthesis will result in the loss of molybdopterin-dependent enzyme activity. Deficiencies in molybdopterin-dependent enzymes cause neonatal seizures, mental retardation and lens dislocation. Other diseases caused by defects in cofactor metabolism include pernicious anemia and methylmalonic aciduria.

The discovery of new human enzyme molecules 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 autoimmune/inflammation disorders, genetic disorders, neurological disorders, and cell proliferative disorders including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human enzyme molecules.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, human enzyme molecules, referred to collectively as "HEM" and individually as "HEM-1," "HEM-2," "HEM-3," "HEM-4," "HEM-5," "HEM-6," "HEM-7," "HEM-8," "HEM-9," "HEM-10," "HEM-11," "HEM-12," "HEM-13," "HEM-14," "HEM-15," "HEM-16," "HEM-17," "HEM-18," "HEM-19," "HEM-20," "HEM-21," "HEM-22," "HEM-23," "HEM-24," "HEM-25," and "HEM-26." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-26.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-26. In another alternative, the

WO 01/64896

PCT/US01/06806

polynucleotide is selected from the group consisting of SEQ ID NO:27-52.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

WO 01/64896

PCT/US01/06806

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional HEM, 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

WO 01/64896

PCT/US01/06806

agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. 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 HEM, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. 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 HEM, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. 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.

WO 01/64896

PCT/US01/06806

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. 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 sequence selected from the group consisting of SEQ ID NO:27-52, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the

WO 01/64896

PCT/US01/06806

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.

5

BRIEF DESCRIPTION OF THE TABLES

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

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

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

Table 4 lists the cDNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

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

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

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

DESCRIPTION OF THE INVENTION

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

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

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

WO 01/64896

PCT/US01/06806

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

DEFINITIONS

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

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

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

"Altered" nucleic acid sequences encoding HEM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as HEM or a polypeptide with at least one functional characteristic of HEM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding HEM, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HEM. 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 HEM. 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 HEM 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

WO 01/64896

PCT/US01/06806

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

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

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

WO 01/64896

PCT/US01/06806

modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic HEM, 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 HEM or fragments of HEM 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.

WO 01/64896

PCT/US01/06806

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
5	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
10	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
15	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
20	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.

A "fragment" is a unique portion of HEM or the polynucleotide encoding HEM which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 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

WO 01/64896

PCT/US01/06806

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
5 embodiments.

A fragment of SEQ ID NO:27-52 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:27-52, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:27-52 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish
10 SEQ ID NO:27-52 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:27-52 and the region of SEQ ID NO:27-52 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:1-26 is encoded by a fragment of SEQ ID NO:27-52. A fragment of SEQ ID NO:1-26 comprises a region of unique amino acid sequence that specifically identifies
15 SEQ ID NO:1-26. For example, a fragment of SEQ ID NO:1-26 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-26. The precise length of a fragment of SEQ ID NO:1-26 and the region of SEQ ID NO:1-26 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

20 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.

25 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.

30 Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS
35 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as

WO 01/64896

PCT/US01/06806

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

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) *J. Mol. Biol.* 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

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

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

WO 01/64896

PCT/US01/06806

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 used to describe a length over which percentage identity may be measured.

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

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely

WO 01/64896

PCT/US01/06806

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.

5 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
10 in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

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

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

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A
35 hybridization complex may be formed in solution (e.g., C_{3t} or R_{3t} analysis) or formed between one

WO 01/64896

PCT/US01/06806

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

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

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

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

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

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

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

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

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

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

WO 01/64896

PCT/US01/06806

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

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

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

Methods for preparing and using probes and primers are described in the references, for
20 example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that
25 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
30 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3
35 primer selection program (available to the public from the Whitehead Institute/MIT Center for

WO 01/64896

PCT/US01/06806

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

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

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

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

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

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the

WO 01/64896

PCT/US01/06806

nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

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

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

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

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

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

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

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

A "transgenic organism," as used herein, is any organism, including but not limited to

WO 01/64896

PCT/US01/06806

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

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

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

WO 01/64896

PCT/US01/06806

94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

5 The invention is based on the discovery of new human enzyme molecules (HEM), the polynucleotides encoding HEM, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammation disorders, genetic disorders, neurological disorders, and cell proliferative disorders including cancer.

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

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

20 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO.) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the 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, 30 searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are human enzyme molecules. For example, SEQ ID NO:5 is 84% identical, from residue M1 to residue E384, to *Rattus norvegicus* beta-alanine 35

WO 01/64896

PCT/US01/06806

synthase (GenBank ID g203106) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a carbon-nitrogen hydrolase domain as determined by searching for statistically significant matches in the hidden

5 Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLMPs and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:5 is a hydrolase. SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, and SEQ ID NO:26 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-26 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO.) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:27-52 or that distinguish between SEQ ID NO:27-52 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 3540084H1 is the identification number of an Incyte cDNA sequence, and SEMVNOT04 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., SXAD90083V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2795888) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the

WO 01/64896

PCT/US01/06806

identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

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

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

The invention also encompasses polynucleotides which encode HEM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:27-52, which encodes HEM. The polynucleotide sequences of SEQ ID NO:27-52, 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 HEM. 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 HEM. 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:27-52 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:27-52. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of HEM.

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 HEM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide

WO 01/64896

PCT/US01/06806

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

5 Although nucleotide sequences which encode HEM and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring HEM under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HEM 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
10 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 HEM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

15 The invention also encompasses production of DNA sequences which encode HEM and HEM 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 HEM or any fragment thereof.

20 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:27-52 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in
25 "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or
30 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA
35 sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system

WO 01/64896

PCT/US01/06806

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

5 The nucleic acid sequences encoding HliM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

10 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme

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

20 (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

25 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

30 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

35 omitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate

WO 01/64896

PCT/US01/06806

software (e.g., GENOTYPHER 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 HEM may be cloned in recombinant DNA molecules that direct expression of HEM, 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 HEM.

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

15 sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or

20 improve the biological properties of HEM, 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 HEM may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.)

35

WO 01/64896

PCT/US01/06806

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

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

In order to express a biologically active HEM, the nucleotide sequences encoding HEM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding HEM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding HEM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding HEM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

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

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

WO 01/64896

PCT/US01/06806

encoding HEM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.F. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Sonnia (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 HEM. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding HEM can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding HEM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of HEM are needed, e.g. for the production of antibodies, vectors which direct high level expression of HEM 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 HEM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994)

WO 01/64896

PCT/US01/06806

Bio/Technology 12:181-184.)

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

5 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology
10 (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 HEM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential H1 or H3 region of the viral genome may be used to obtain
15 infective virus which expresses HEM in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

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

For long term production of recombinant proteins in mammalian systems, stable expression
25 of HEM in cell lines is preferred. For example, sequences encoding HEM 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
30 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
35 phosphoribosyltransferase genes, for use in *tk*- and *aprt* cells, respectively. (See, e.g., Wigler, M. et

WO 01/64896

PCT/US01/06806

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

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding HEM is inserted within a marker gene sequence, transformed cells containing sequences encoding HEM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding HEM 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 HEM and that express HEM 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 HEM 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 HEM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and

WO 01/64896

PCT/US01/06806

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

Host cells transformed with nucleotide sequences encoding HEM 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 HEM may be designed to contain signal sequences which direct secretion of HEM 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 WB38) 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 HEM 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 HEM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of HEM 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

WO 01/64896

PCT/US01/06806

cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, c-myc, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the HEM encoding sequence and the heterologous protein sequence, so that HEM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

10 In a further embodiment of the invention, synthesis of radiolabeled HEM 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.

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

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

30 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 HEM, either in solution or affixed to a solid support, and detecting the binding of HEM to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical

WO 01/64896

PCT/US01/06806

libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

HEM of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of HEM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for HEM activity, wherein HEM is combined with at least one test compound, and the activity of HEM in the presence of a test compound is compared with the activity of HEM in the absence of the test compound. A change in the activity of HEM in the presence of the test compound is indicative of a compound that modulates the activity of HEM. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising HEM under conditions suitable for HEM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of HEM 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 HEM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capeczuk, 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 HEM 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 HEM can also be used to create "knockin" humanized animals

WO 01/64896

PCT/US01/06806

(pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding HEM 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 HEM, e.g., by secreting HEM 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 HEM and human enzyme molecules. In addition, the expression of HEM is closely associated with pancreatic, urogenital and cancerous tissues. Therefore, HEM appears to play a role in autoimmune/inflammation disorders, genetic disorders, neurological disorders, and cell proliferative disorders including cancer. In the treatment of disorders associated with increased HEM expression or activity, it is desirable to decrease the expression or activity of HEM. In the treatment of disorders associated with decreased HEM expression or activity, it is desirable to increase the expression or activity of HEM.

Therefore, in one embodiment, HEM 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 HEM. Examples of such disorders include, but are not limited to, an autoimmune/inflammation disorder such as acquired immunodeficiency syndrome (AIDS), actinic keratosis, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, 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, paroxysmal nocturnal hemoglobinuria, hepatitis, hyper eosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma and hematopoietic cancer including lymphoma, leukemia, and myeloma; a genetic disorder

WO 01/64896

PCT/US01/06806

such as adrenoleukodystrophy, Alport's syndrome, ocheroideremia, Duchenne and Becker muscular dystrophy, Down's 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, Werner syndrome, von

5 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, mitochondrial electron transport

10 flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; a neurological disorder such as neonatal seizures, epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and

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

20 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,

25 dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a

30 cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall

35 bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid,

WO 01/64896

PCT/US01/06806

penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

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

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

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

In a further embodiment, an antagonist of HEM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of HEM. Examples of such disorders include, but are not limited to, those autoimmune/inflammation disorders, genetic disorders, 15 neurological disorders, and cell proliferative disorders including cancer described above. In one aspect, an antibody which specifically binds HEM 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 HEM.

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

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

An antagonist of HEM may be produced using methods which are generally known in the art. 30 In particular, purified HEM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind HEM. Antibodies to HEM 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 35 dimer formation) are generally preferred for therapeutic use.

WO 01/64896

PCT/US01/06806

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with HEM 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-Guérin*) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to HEM 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 HEM 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 HEM 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 B₂BV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Közbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 6:109-120.)

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

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

Antibody fragments which contain specific binding sites for HEM 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

WO 01/64896

PCT/US01/06806

the F(ab)₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

5 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 HEM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering HEM 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 HEM. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of HEM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. 15 The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple HEM epitopes, represents the average affinity, or avidity, of the antibodies for HEM. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular HEM epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^6 to 10^{12} L/mole are preferred for use in immunoassays in which the 20 HEM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^4 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of HEM, 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, 25 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 30 of HEM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligen et al. *supra*.)

In another embodiment of the invention, the polynucleotides encoding HEM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene 35 expression can be achieved by designing complementary sequences or antisense molecules (DNA,

WO 01/64896

PCT/US01/06806

RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding HEM. 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 HEM. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

5 In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.B. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 10 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(5):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding HEM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. 30 (1988) *Nature* 335:395-396; Poeschia, 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 HEM expression or regulation causes disease, the expression of HEM from an appropriate population of transduced cells may alleviate the clinical manifestations 35 caused by the genetic deficiency.

WO 01/64896

PCT/US01/06806

In a further embodiment of the invention, diseases or disorders caused by deficiencies in HEM are treated by constructing mammalian expression vectors encoding HEM and introducing these vectors by mechanical means into HEM-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) 5 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. Récapion (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of HEM include, but are not 10 limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRH2-LUC, PTK-HYG (Clontech, Palo Alto CA). HEM 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 15 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 dexamethasone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. 20 and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding HEM 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 25 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with 30 respect to HEM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding HEM 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 35 commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc.*

WO 01/64896

PCT/US01/06806

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; Bonder, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

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

25 In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding HEM to target cells which have one or more genetic abnormalities with respect to the expression of HEM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing HEM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type I-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV δ 92 which consists of a genome containing at least one exogenous gene to be transferred to a

30
35

WO 01/64896

PCT/US01/06806

cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned
5 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
10 deliver polynucleotides encoding HEM 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,
15 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 HEM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of HEM-coding RNAs and the synthesis of high levels of HEM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent
20 infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of HEM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating
25 infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

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

WO 01/64896

PCT/US01/06806

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

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

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques
15 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 HEM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines,
20 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
25 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
30 compound which is effective in altering expression of a polynucleotide encoding HEM. 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
35 compounds may alter polynucleotide expression by acting as either inhibitors or promoters of

WO 01/64896

PCT/US01/06806

polynucleotide expression. Thus, in the treatment of disorders associated with increased HEM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding HEM may be therapeutically useful, and in the treatment of disorders associated with decreased HEM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding HEM 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 HEM 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 HEM 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 HEM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:515) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

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

WO 01/64896

PCT/US01/06806

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.

5 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
10 consist of HEM, antibodies to HEM, and mimetics, agonists, antagonists, or inhibitors of HEM.

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.

15 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
20 lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination
25 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 HEM or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, HEM or a fragment thereof may be joined to a short cationic N-
30 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,
35 monkeys, or pigs. An animal model may also be used to determine the appropriate concentration

WO 01/64896

PCT/US01/06806

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

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 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 HEM may be used for the diagnosis of disorders characterized by expression of HEM, or in assays to monitor patients being treated with HEM or agonists, antagonists, or inhibitors of HEM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for HEM include methods which utilize the antibody and a label to detect HEM 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

WO 01/64896

PCT/US01/06806

molecules, several of which are described above, are known in the art and may be used.

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

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

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding HEM or closely related molecules may be used to identify nucleic acid sequences which encode HEM. The specificity of the probe, whether it is made
20 from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding HEM, 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 HEM encoding sequences. The hybridization probes of the subject
25 invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:27-52 or from genomic sequences including promoters, enhancers, and introns of the HEM gene.

Means for producing specific hybridization probes for DNAs encoding HEM include the cloning of polynucleotide sequences encoding HEM or HEM derivatives into vectors for the
30 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 radioisotopes 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 HEM may be used for the diagnosis of disorders
35 associated with expression of HEM. Examples of such disorders include, but are not limited to, an

WO 01/64896

PCT/US01/06806

autoimmune/inflammation disorder such as acquired immunodeficiency syndrome (AIDS), actinic keratosis, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis,

5 cholecystitis, cirrhosis, 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, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue

10 disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and

15 extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma and hematopoietic cancer including lymphoma, leukemia, and myeloma; a genetic disorder such as adrenoleukodystrophy, Alport's syndrome, choroideremia, Duchenne and Becker muscular dystrophy, Down's syndrome, cystic fibrosis, chronic granulomatous disease, Gaucher's disease, Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic dystrophy, pyrenodysostosis,

20 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 very-long-chain acyl-CoA

25 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 neurological disorder such as neonatal seizures, epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease,

30 dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob

35 disease, and Gerstmann-Strausler-Scheinker syndrome, fatal familial insomnia, nutritional and

WO 01/64896

PCT/US01/06806

metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebellar/retinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer 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 HEM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered HEM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding HEM may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding HEM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding HEM 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 HEM, 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 HEM, under conditions suitable for hybridization or amplification.

WO 01/64896

PCT/US01/06806

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
5 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
10 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
15 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 HEM 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
20 encoding HEM, or a fragment of a polynucleotide complementary to the polynucleotide encoding HEM, 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
25 encoding HEM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding HEM are used to
30 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-
35 throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis

WO 01/64896

PCT/US01/06806

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).

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

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

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of

WO 01/64896

PCT/US01/06806

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.

5 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
10 pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaisir, 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
15 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
20 are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of
25 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
30 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.

35 Another particular embodiment relates to the use of the polypeptide sequences of the present

WO 01/64896

PCT/US01/06806

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

A proteomic profile may also be generated using antibodies specific for HEM to quantify the levels of HEM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendez, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound to 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. Sellhammer (1997) *Electrophoresis* 18:533-537), so proteomic toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic 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

WO 01/64896

PCT/US01/06806

cases.

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

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

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

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

WO 01/64896

PCT/US01/06806

inheritance of a particular chromosomal region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

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

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

In another embodiment of the invention, HEM, 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 HEM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with HEM, or fragments thereof, and washed. Bound HEM is then detected by methods well known in the art. Purified HEM 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 HEM specifically compete with a test compound for binding HEM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with HEM.

WO 01/64896

PCT/US01/06806

In additional embodiments, the nucleotide sequences which encode HEM 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.

5 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.

10 The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/186,307, U.S. Ser. No. 60/193,578, and U.S. Ser. No. 60/192,532, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

15 Invertebrate cDNAs were derived from cDNA libraries described in the LITESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or
20 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,
25 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
30 vector system (Stratagene) or SUPERSCRIP^T plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-5.6.) 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-
35 1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column

WO 01/64896

PCT/US01/06806

chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5a, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

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

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

III. Sequencing and Analysis

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

WO 01/64896

PCT/US01/06806

disclosed in Example VIII.

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

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of 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

WO 01/64896

PCT/US01/06806

and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:27-52. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

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

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequences. Sequence intervals in which the entire length of the interval was present on

WO 01/64896

PCT/US01/06806

more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbprl 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.

VI. Chromosomal Mapping of HEM Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:27-52 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:27-52 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 Génethon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

WO 01/64896

PCT/US01/06806

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

In this manner, SEQ ID NO:32 was mapped to chromosome 20 within the interval from 11.0 to 20.9 centiMorgans. SEQ ID NO:35 was mapped to chromosome 2 within the interval from 175.0 to 180.6 centiMorgans and within the interval from 190.8 to 196.3 centiMorgans. SEQ ID NO:41 was mapped to chromosome 1 within the interval from 235.8 to 237.2 centiMorgans and within the interval from 243.3 to 245.2 centiMorgans. SEQ ID NO:47 was mapped to chromosome 2 within the interval from 118.0 to 127.4 centiMorgans. More than one map location is reported for SEQ ID NO:35 and SEQ ID NO:41, indicating that sequences having different map locations were assembled into a single cluster. This situation occurs, for example, when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

20 VII. Analysis of Polynucleotide Expression

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

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

30

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

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is

35

WO 01/64896

PCT/US01/06806

calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding HEM are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hematologic 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 HEM. cDNA sequences and cDNA library/tissue information are found in the LIFSEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of HEM Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

WO 01/64896

PCT/US01/06806

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), HLONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TB and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA

WO 01/64896

PCT/US01/06806

recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

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

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:27-52 are employed to screen cDNAs,
10 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 -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase
15 (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases:
Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

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

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*),
30 mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schemm (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to
35 those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g.,

WO 01/64896

PCT/US01/06806

Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

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

Microarray Preparation

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

WO 01/64896

PCT/US01/06806

Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR

- 5 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

- Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

- Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

- Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

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

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

WO 01/64896

PCT/US01/06806

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

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

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

35

WO 01/64896

PCT/US01/06806

XII. Expression of HEM

Expression and purification of HEM is achieved using bacterial or virus-based expression systems. For expression of HEM 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 HEM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of HEM 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 HEM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (SF9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, HEM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from HEM 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 (1995, *supra*, ch. 10 and 16). Purified HEM obtained by these methods can be used directly in the assays shown in Examples XVI and XVII, where applicable.

XIII. Functional Assays

HEM function is assessed by expressing the sequences encoding HEM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which

WO 01/64896

PCT/US01/06806

contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) *Flow Cytometry*, Oxford, New York NY.

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

XIV. Production of HEM Specific Antibodies

HEM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

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

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A

WO 01/64896

PCT/US01/06806

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

XV. Purification of Naturally Occurring HEM Using Specific Antibodies

Naturally occurring or recombinant HEM is substantially purified by immunoaffinity chromatography using antibodies specific for HEM. An immunoaffinity column is constructed by covalently coupling anti-HEM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing HEM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of HEM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HEM 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 HEM is collected.

XVI. Identification of Molecules Which Interact with HEM

HEM, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled HEM, washed, and any wells with labeled HEM complex are assayed. Data obtained using different concentrations of HEM are used to calculate values for the number, affinity, and association of HEM with the candidate molecules.

Alternatively, molecules interacting with HEM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

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

XVII. Demonstration of HEM Activity

Cofactor Biosynthetic Enzyme Activity

The molybdopterin synthase sulfurylase activity of HEM is measured, for example, by

WO 01/64896

PCT/US01/06806

transforming an *Aspergillus nidulans* *coxF* mutant with a suitable expression vector containing the polynucleotide sequence of HEM and observing whether the *A. nidulans* cells are able to grow on nitrate or hypoxanthine-containing media (see Appleyard, M. et al., *supra*). Since molybdopterin synthase sulfurylase activity is absent in *A. nidulans* *coxF* mutants, and this activity is necessary for utilization of nitrate or hypoxanthine, the rescue of these mutants with an expression vector containing the polynucleotide sequence of HEM indicates that HEM has molybdopterin synthase sulfurylase activity. Also, the levels of molybdopterin and intermediates in the biosynthetic pathway are measured using HPLC analysis (see Appleyard, M. et al., *supra*).

Isomerase Activity

10 Isomerase activity of HEM is demonstrated through a variety of specific enzyme assays, some of which are outlined below.

Peptidyl Prolyl *cis-trans* Isomerase Activity

HEM peptidyl prolyl *cis-trans* isomerase activity can be assayed as described (Rahfeld, J.U. et al. (1994) FEBS Lett. 352:180-184). The assay is performed at 10°C in 35 mM HEPES buffer, pH 7.8, containing chymotrypsin (0.5 mg/ml) and HEM at a variety of concentrations. In this assay, 15 the substrate is a peptide containing four hydrophobic residues. The peptide contains a succinate group at the N-terminus and a nitroanilide group at the C-terminus. The substrate is in equilibrium with respect to the prolyl bond, with 80-95% in *trans* and 5-20% in *cis* conformation. An aliquot (2 µl) of the substrate dissolved in dimethyl sulfoxide (10 mg/ml) is added to the reaction mixture 20 described above. Only the *cis* isomer of the substrate is a substrate for cleavage by chymotrypsin. Thus, as the substrate is isomerized by HEM, the product is cleaved by chymotrypsin to produce 4-nitroanilide, which is detected by its absorbance at 390 nm. 4-Nitroanilide appears in a time-dependent and an HEM concentration-dependent manner.

Alternatively, peptidyl prolyl *cis-trans* isomerase activity of HEM can be assayed using a 25 chromogenic peptide in a coupled assay with chymotrypsin (Fischer, G. et al. (1984) Biomed. Biochim. Acta 43:1101-1111).

Thioredoxin Activity

HEM thioredoxin activity is assayed as described (Luthman, M. (1982) Biochemistry 21:6628-6633). Thioredoxins catalyze the formation of disulfide bonds and regulate the redox 30 environment in cells to enable the necessary thiol:disulfide exchanges. One way to measure the thiol:disulfide exchange is by measuring the reduction of insulin in a mixture containing 0.1M potassium phosphate, pH 7.0, 2 mM EDTA, 0.16 µM insulin, 0.33 mM DTT, and 0.48 mM NADPH. Different concentrations of HEM are added to the mixture, and the reaction rate is followed by monitoring the oxidation of NADPH at 340 nM.

35 Transferase Activity

WO 01/64896

PCT/US01/06806

HEM transferase activity is measured through a methyl transferase assay in which the transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate is measured (Bokar, J.A. et al. (1994) J. Biol. Chem. 269:17697-17704). Reaction mixtures (50 μ l final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM dithiothreitol, 3% polyvinylalcohol, donor substrate (1.5 μ Ci [*methyl-³H*]AdoMet (0.375 μ M AdoMet) (DuPont-NEN)), 0.6 μ g HEM, and acceptor substrate (0.4 μ g [*3⁵S*]RNA or 6-mercaptapurine (6-MP) to 1 mM final concentration). Reaction mixtures are incubated at 30°C for 50 minutes, then 65°C for 5 minutes. The products are separated by chromatography or electrophoresis and the level of methyl transferase activity is determined by quantification of *methyl-³H*-RNA or *methyl-³H*-6-MP recovery.

10 Hydrolase Activity

For purposes of example, an assay measuring the β -glucosidase activity of an HEM molecule is described. Varying amounts of HEM are incubated with 1 mM 4-nitrophenyl β -D-glycopyranoside (a substrate) in 50 mM sodium acetate buffer, pH 5.0, for various times (typically 1-5 minutes) at 37°C. The reaction is terminated by heating to 100°C for 2 minutes. The absorbance is measured spectrophotometrically at 410 nm, and the change in absorbance is proportional to the activity of HEM in the sample. (Hirmova, M. et al. (1998) J. Biol. Chem. 273:11134-11143.)

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

WO 01/64896

PCT/US01/06806

Table 1

Project ID	Incyte Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Incyte Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
2783442	1	2783442CDL	27	2783442CH1
2116380	2	2116380CDL	28	2116380CH1
5785224	3	5785224CDL	29	5785224CH1
1870996	4	1870996CDL	30	1870996CH1
138841	5	138841CDL	31	138841CB1
1485405	6	1485405CDL	32	1485405CH1
2024617	7	2024617CDL	33	2024617CH1
4721827	8	4721827CDL	34	4721827CH1
5406614	9	5406614CDL	35	5406614CH1
1252792	10	1252792CDL	36	1252792CH1
1299305	11	1299305CDL	37	1299305CH1
1385190	12	1385190CDL	38	1385190CH1
2133152	13	2133152CDL	39	2133152CH1
254557	14	254557CDL	40	254557CB1
3903488	15	3903488CDL	41	3903488CH1
5391816	16	5391816CDL	42	5391816CH1
5865989	17	5865989CDL	43	5865989CH1
723432	18	723432CDL	44	723432CH1
1332963	19	1332963CDL	45	1332963CH1
1559410	20	1559410CDL	46	1559410CH1
1752587	21	1752587CDL	47	1752587CH1
1917509	22	1917509CDL	48	1917509CH1
2210170	23	2210170CDL	49	2210170CH1
286544	24	286544CDL	50	286544CB1
4103559	25	4103559CDL	51	4103559CH1
4103559	26	4103559CDL	52	4103559CH1

WO 01/64896

PCT/US01/06806

Table 2

Polypeptide SEQ ID NO:	Enzyme Polyptide ID	Genbank ID NO:	Probability score	Genbank Homolog
1	278344CD1	G132506	1.00E-130	similar to molybdopterin biosynthesis MOBE proteins [Caenorhabditis elegans] Wilson, R. et al. (1994) Nature 368:32-38
2	3116390CD1	G2984222	1.2e-12	Fuculose-1-phosphate aldolase [Aquifex aeolicus] Deckert, G. et al. (1998) Nature 392:553-558
		G10172247	1.00E-55	contains similarity to enolase-phosphatase-gene_id:K19917.1 [Arabidopsis thaliana] Nakamura, Y. et al. (1997) DNA Res. 4:401-414
3	5785224CD1	G1799782	2e-61	5'ile acid-inducible operon Protein F [Escherichia coli] Yamamoto, Y. et al. (1997) DNA Res. 4:91-113
4	1870996CD1	G1788542	1.00E-19	purative sugar nucleoside epimerase [Escherichia coli K12] Bhattar, P.R. et al. (1997) Science 277:1453-1474
5	138841CD1	G203106	0	Beta-alanine synthase (Rattus norvegicus) [Walnes-Krick, K.L. et al. (1993) J. Biol. Chem. 268:5685-5693]
6	1485435CD1	G19137	6.4e-33	Aparaginase [Lupinus arborescens] Lough, T.J. et al. (1992) Plant Mol Biol 19:392-399
7	2024417CD1	G1196914	2.70E-278	L-glutamine amidohydrolase [Rattus norvegicus] [Wang, S.M. et al. (1990) J. Biol. Chem. 265:10651-10656]
8	4721827CD1	G11846091	1.00E-115	alpha-D-glucosyltransferase (similar to galactinase-sulfatases and KRAL077) (Homo sapiens)

WO 01/64896

PCT/US01/06806

Table 2 (cont.)

Polypeptide SEQ. ID NO.	Incyte Polypeptide ID	GenBank ID NO.	Probabilty score	GenBank Homology
9	5406614CD1	G5524927	2.30E-161	deoxyribonuclease III (DNase III) [Homo sapiens] Ross, R. et al. (1999) EMBO J. 19:2869-2875
10	1252792CD1	G7576450	0	ALEX1 [Homo sapiens] Kurochkin, I.V. et al. (2001) Biochem. Biophys. Res. Commun. 280:340-347
11	1299305CD1	G11545670	0	serina beta lactamase-like protein LactY-1 [Mus musculus]
15	3903468CD1	G4103133	6.00E-66	putative transcription factor Pet5 [Schizosaccharomyces pombe] Szpakowski, G.V. et al. (1997) Biocorg. Chim. 23:234-237
16	539-816CD1	G3822036	1.7E-86	Furine nucleotide binding protein fat5 [Schizosaccharomyces pombe]
18	723432CD1	G6459613	1.40E-143	"endo-1,3-1,4-beta-D-glucanase [Zea mays]" Thomas, B.R. et al. (1998) Plant Physiol. 117:1525
19	1332963CD1	G40-73877	2.20E-61	ArgE/DppE/AsyI Family aminocyclase protease (Bainococcus radiodurans) White, O. et al. (1999) Science 286:1571-1577
			3.00E-82	[5-methylaminomethyl-2-thiouridyate)-methyl transferase [Bacillus halodurans] Takami, H. et al. (1999) Extremophiles 3:23-28
			2.70E-69	PF-100p superfamily kinase [Chlaevdia trichomatis] Stephens, R.S. et al. (1998) Science 282:154-159

WO 01/64896

PCT/US01/06806

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
20	1559410CD1	G194367	8.00E-57	weak similarity to APases [Caenorhabditis elegans] the C. elegans genome Sequencing Consortium (1998) Science 282:2042- 2046
21	1752587CD1	G2225906	2.20E-68	FIG-1 deacetylase [Rattus norvegicus] Makamura, N. et al. (1997) J. Biol. Chem. 272:15834-15840
22	1911509CD1	G4825778	1.10E-57	similar to acetyltransferases [Caenorhabditis elegans] Wilson, R. et al. (1994) Nature 368:32-38
23	2210170CD1	G2855691	5.00E-58	sup binding protein-like [Arabidopsis thaliana]
24	234684CD1	G1101009	5.00E-26	endothelin converting enzyme-2 (Bos taurus) Amoto, N. et al. (1995) J. Biol. Chem. 270:15262-15268
25	2884114CD1	G2118131	3.00E-90	histone deacetylase (Arabidopsis thaliana)
		G6132878	0	histone deacetylase 8; HDAC8 [Homo sapiens]
26	4103559CD1	G3028603	1.30E-53	katamin p60 subunit [Strongylocentrotus purpuratus]

WO 01/64896

PCT/US01/06806

Table 3

SBG Inocyte ID Polypeptide (NCI ID)	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1 2783442CDL	404	S86 T152 S180 T201 S350 S381 T386 S389 T134 T243 S296	M150 A356	Th1F Family repeated domain in bacterial MceB/MucH/HesA family and ubiquitin-activating enzyme EI: R72-R221 Molybdopterin biosynthesis (MOBS) proteins: P003566; G222-L369 Opaquin-activating enzyme: T000412/A49890 1-238; Y53-F232 Class II Aldolase and Alducin N-terminal domain Aldolase II: R25-V242 Class II Aldolase and Alducin N-terminal domain P00356; T44-L50, T56-D77, K93-S118, V190-A207 PROTEIN ISOMERASE 21HC PROBABLE SUGAR ADJACENT CYTOSKELETON MEMBRANE CALMODULIN-BINDING A-TERMINATIVE L390: E30-A207 PROTEIN LACONASE BILE ACID-INDUCIBLE OPERON P ISOMERASE TRANSFERASE ALPHANETHYLACTYL COA L-CARBITINE PD022950: P6-T127 TRANSFERASE OF SECTION PROTEIN FORMYL COA PD150345: A282-E330 PROTEIN LACONASE ISOMERASE BILE ACID-INDUCIBLE OPERON P TRANSFERASE ALPHANETHYLACTYL COA L-CARBITINE PD035472: G128-C261 Signal Peptide M1-A39	EMBL-PPM BLAST-PRODB BLAST-DBO EMBL-PPM BLAST-PRODB
2 2116390CDL	242	T57 Y216 S13 S49 S89 T105 S223 T148 S2 T136 T167 T209			BLAST-PRODB
3 5785224CDL	404	T155 S136 T192 S293 T301		PROTEIN CELL DIVISION OF INHIBITOR SECTION PFB SUGA PD011289: X1-V125 CELL DIVISION INHIBITOR PD035984: A66-L10 1-256; E5-V125	BLAST-PRODB BLAST-DBO
4 1870986CDL	126	T57 S113			SPSCM

WO 01/64896

PCT/US01/06806

Table 3 (cont.)

SEQ ID NO	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	1385190CDL	530	S510 S180 T215 T491 S117 S193 T228 S254 S277 T398 I299 S355 S435	N30 N451	ATP/GTP binding site motif (P Loop): G211-K217	MOTIFS
13	2133162CDL	697	S257 T632 S2 T242 T252 S256 T383 T402 T422 S431 S527 T547 S595 S608 T622 T641 S657 S40 T94 S117 S232 S234 S236 S501 S463 S488 S519 S552 S585	N48 N173 N283	ATP/GTP binding site motif (P Loop): K615-K621	MOTIFS
14	354567CDL	190	S187 S198 T66	M168	ATP/GTP binding site motif (P Loop): C9-T16 C9X/C9Y SUPERFAMILY SIGNATURE: H4-L19 ATP-BINDING PROTEIN: V5-G186 ATP-BINDING PROTEIN: V5-G186	MOTIFS BLASTS-PRODOM MOTIFS
15	390348CDL	284	Y196 S14 S68 S148 S180 S208 T215 T174 S197 S202 S236 S272	N31 N47	ATP/GTP binding site motif (P Loop): G10-S17 Signal cleavage: M1-A29 ATP-BINDING PROTEIN: I7-E269 ATP WE BIND: M1-E268	SPSCAN BLASTS-PRODOM BLAST-DOMO
16	5391816CDL	245	T191 Y155 S17 S223		Epoxide hydrolase signature: I128-L141	BLASTS-PRODOM
17	588698CDL	331	S12 T99 T120 T166 T185 T87 T194 S149		ATP/GTP binding site motif (P Loop): G8-K14 Signal cleavage: M1-G13	MOTIFS SPSCAN
					Uridine Kinase signature: M1-I18	BLASTS-PRODOM
					ATP-BINDING 1'REGION: F2-G180	BLAST-PRODOM
					ATP WE BIND: M1-G180	BLAST-DOMO

WO 01/64896

PCT/US01/06806

Table 3 (cont.)

SEQ Incyte ID Polypeptide No. ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18 723423CDL	475	S117 S130 T150 S168 S223 S235 S249 T270 S370 T423 S439 S446 Y377 S223 S399 S341 T362 T428 M183		INVERSEMIC REGION CONTAINING TRPASP REPEATS: N157-V436 A Beta-transducin family ttp-asp repeats: A3-Q472	BLAST-PRODOM BLAST-DOGO
19 1332863CDL	421	S41 T131 S132 S169 S248 S406 S410 S169 T255 S284 T107 S380 Y58 S218 S178 S20	M153 M249	Signal peptide: M1-A20 PROTEIN TRANSFERASE: R130-K215	SIGPEPPI BLAST-PRODOM
20 1559410CDL	310		N300	ATP/GTP binding site motif (P Loop): G16-D23 Leucine zipper: L81-I102 Adenylate kinase signature: A13-C26 ATP-BINDING RM41-PE22 INVERSEMIC REGION: I15-L182 ATP NP BIND: G11-K269 Signal peptide: Y1-W30 Signal cleavage: X1-G17 Transmembrane domain: M1-D23	MOTIFS BLIAMS-FRIMPS BLAST-PRODOM BLAST-DOGO SIGPEPPI SFGCAN THAP
21 1752887CDL	194	S35 T87			
22 1911509CDL	349	S122 S134 S238 T244 S298 T49 W55 S73 T100 S115 S163 S310 S217	N303 M71 M765	Beta hydrolyase: R102-C345 Alpha/Beta hydrolyase fold Alpha/beta hydrolyase fold signature: M11-S116, L149-L62, S163-V176 HYDROLASE: K52-F181 Hydrolyase: W70-P182 ATP/GTP binding site motif (P Loop): G75-K82 Signal peptide: M1-A23	BLIAMS-FRIMPS BLAST-PRODOM BLAST-DOGO MOTIFS SIGPEPPI
23 2210170CDL	245	S14 T44 S74 S97 M85 T238 S153 T209		ATP (ATP-binding) family proteins: V96-L92, E101-V113, V154-V195 HYPOPHOSPHATE ATP-BINDING PROTEIN: I185-V234 R1PH/PMXC FAMILY: Y66-M224	BLIAMS-BLOCKS BLAST-PRODOM BLAST-DOGO

WO 01/64896

PCT/US01/06806

Table 3 (cont.)

SEQ ID No.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
24	234664CD1	235	S47 S194 T85 S444	N83	Hydrogen uptake protein: I62-F74 ERDOTHIELIN-CONVERTING HYDROLASE: M97-ELF0	BLIMPS-FINDMS BLAST-PRODOM
25	288414CD1	377	S26 S43 S63 Y174 T192 T311 S351 S39	K136	Histone deacetylase family: V17-G324 Histone deacetylase family: S65-L79, D101-G117, G139-I162, R171-M196, F208-V240, I243-I284, I300-R313 HISTONE DEACETYLASE: R3/G324	HOMER-PFAM BLIMPS-PFAM
26	410359CD1	403	T7 T106 S215 T297 S330 S339 S172 S93 S94 S155 S171 S365 S443 S174	N113 S316	RPD3 histone deacetylase/ACUC acetooin utilization: F35-V321 AcP/GTF binding site motif (P loop): G293-E298 AAA: ATPases associated with various cellular activities: G288-W603 ENDOPETIDASE LA (LONG S): G293-Y312 ATP-dependent CIP protease NTP-binding subunit: L288-A307 AAA-protein family protein: W252-P272, W186-A307, A119-R361 ATP-BINDING PROTEASE SUBUNIT HOMOLOG: T300-E402 AAA-PROTEIN FAMILY: F248-W403	BLAST-DOMO MOTIFS HOMER-PFAM BLIMPS-FINDMS BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO

WO 01/64896

PCT/US01/06806

Table 4

Polynucleotide SEQ ID No.	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
27	2783442CB1	1478	1-537	3540084HL (SERRRTG2)	1	290
				1619664HL (BRALTTU13)	807	1478
				2084638RS (PANCNO704)	255	882
				897793R1 (BRSTNOR05)	531	1142
28	2116390CB1	1383	1-62	190870F1 (SXTORAB01)	404	1183
				776806S1 (COLMOR05)	1	565
29	5785224CB1	1572	1-656	644802R6 (BRSTT002)	401	835
				5785224HL (FTRRTXS07)	699	994
				2506827R6 (CONUTU01)	1	516
				12123357R6 (BRSTT001)	929	1572
30	1870986CB1	1535	1-170, 255-312	1795181R6 (PROSFU05)	1311	1335
				1398049F1 (BRALTTU08)	802	1335
				882090083V1	1	463
				2182959H1 (SJMINGT01)	328	770
				555900511V1	651	1133
31	136841CB1	1341	1-46	3507433HL (COMCNGT01)	464	780
				138841R5 (LTYRNGT01)	1	480
				161919R5 (LTYRNGT01)	820	842
				SCPAG5578V1	515	1094
32	1485405CB1	755		SCPAG6393D1	392	1061
				831524R1 (PROSTU04)	538	795
				1971451F6 (COMCL5F01)	370	747
				3075932HL (BONDEF0F1)	1	232
				1970651X3D1L (UCMCL2T01)	213	684
33	2024617CB1	2390	1-1688	1444713T6 (TYRNGT03)	1785	2168
				52719388R	1619	2175
				4144157T6 (BRSTT001)	981	1524
				5872924	1948	2390
				2026172X3D1L (SERRNGR02)	705	1218
				2024617X304B1 (SERRNGR02)	1794	2372
				2024617X302D1 (SERRNGR02)	48	567
				2026172X3D1L (SERRNGR02)	1	481
				2024617R6 (SERRNGR02)	1355	1770
				1444713T2 (TYRNGT03)	1746	2164
				93431607	1942	2378

WO 01/64896

PCT/US01/06806

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Enzytic Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
34	4721327CBI	2340	1-265, 856-2340	173004775 (BRSTTT08)	1747	2273
				502259076 (S20NN0708)	200	835
				173004776 (BRSTTT08)	777	1222
				493543041 (BRSTTT02)	942	1231
				82749381 (PROSN0706)	574	846
				210721486 (BRATTT03)	2858	2940
				181905456 (PROSN0702)	1238	1703
				284460776 (DRGLK0701)	1686	2268
				153893871 (SINTTT01)	2257	2916
				551563741 (LVVDF01)	1	257
35	5406614CBI	1417	1-840	284460776 (DRGLK0701)	1174	1686
				163613876 (UPRSN0706)	930	1417
				129184471 (PGANN0703)	265	863
				87571871 (LUNGAST01)	557	1173
36	1257792CBI	2133	1-863, 2111-2133	SV2A00333VA	*	455
				125279276 (LUNGEP03)	1599	2133
				6180396 (PGANN0701)	854	1401
				125653672 (HEMTU03)	210	769
				157122576 (UPRSN0705)	1017	2624
				256611071 (BRATTT02)	1	243
				268850771 (LUNGRC023)	718	961
				129930576 (BRSTN0707)	783	1306
37	1299305CBI	1829	1-750	SV2A00951VA	249	772
				1299305349V1 (BRSTN0707)	1397	1829
				22873071 (S20NN0701)	1856	1866
				12993053152V1 (BRSTN0707)	1309	1829
				SV2A02187VA	524	1046
				SV2A01827VA	298	798
369226171 (HEAN0701)	1	301				

WO 01/64896

PCT/US01/06806

Table 4 (cont.)

Poly nucleotide SEQ ID NO:	Incyte Polymucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
38	1385180CB1	2323	1-1607	4560082HL (REARVUT01)	505	847
				1461459KL (PANGCOT04)	771	1300
				866833KL (BRAITUT03)	1746	2323
				817831HL (OVANRUT01)	492	737
				4621649HL (ENDCNGT01)	1421	1691
				1436421FL (PANGCOT08)	1280	1565
				140593FL (TLEHSDR01)	1608	2310
				4384773HL (SEEVOT02)	1	347
				4761218G (MMLR2P01)	107	614
				21244437G (ESVINGT07)	1660	2308
39	2133162CP1	2592	1-107, 901-1515, 2398-2441	262173HL (KFRANGT02)	2267	2543
				3605342HL (LONGCOT3)	727	1042
				0402038G (VHLXOT01)	2018	2480
				15793127G (LNDNGT03)	839	1282
				1829765F (REELT01)	2448	2681
				960712HL (ERETVUT03)	2751	2994
				2539058F (DONTUT01)	1	766
				2174569F (ENDCNGT03)	1216	1736
				SA8A0952FL	1378	1897
				15732127G (LNDNGT03)	1733	2260
40	254567CB1	850		378345HL (SPJNNT01)	1979	3268
				3142957HL (HRTAZS07)	1	312
				2192041HL (REUTUT03)	595	850
				1879586CT (BEALUT08)	405	847
				18521537G (PROSNOT19)	290	835
				3240016HL (COLADCT01)	536	1472
				13530417G (LATETUT02)	490	1019
				1216572R1 (BNSVTUT01)	1	586
				2466017G (BEALNOR01)	675	1575
				1211478L (PROSNOT06)	988	1561
41	3903486CB1	1472	1-58, 1387-1450	148307G (COLNNT06)	904	1522
				118023FL (MUSCNOT1)	416	1342
				26085758E (BONNCT01)	1	516
				148307G (COLNNT06)	904	1522
				118023FL (MUSCNOT1)	416	1342
				26085758E (BONNCT01)	1	516
				148307G (COLNNT06)	904	1522
				118023FL (MUSCNOT1)	416	1342
				26085758E (BONNCT01)	1	516
				148307G (COLNNT06)	904	1522
42	5391816CB1	1561		148307G (COLNNT06)	904	1522
				118023FL (MUSCNOT1)	416	1342
				26085758E (BONNCT01)	1	516
				148307G (COLNNT06)	904	1522
				118023FL (MUSCNOT1)	416	1342
				26085758E (BONNCT01)	1	516
				148307G (COLNNT06)	904	1522
				118023FL (MUSCNOT1)	416	1342
				26085758E (BONNCT01)	1	516
				148307G (COLNNT06)	904	1522

WO 01/64896

PCT/US01/06806

Table 4 (cont.)

Polynucleotide SEQ ID NO.	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
43	588698CBI	1949	566-754, 1-243	1563312R6 (BMS70704)	1319	1749
				150418DF1 (BRA3TU07)	541	1107
				911123H1 (CERV2001)	1714	1949
				SRA0308CD1	635	1287
44	733432CBI	2221	28-117	333520P6 (BRA3FT01)	1	646
				870083R6 (LUNG85T01)	1128	1726
				1983045R6 (LUNG85T01)	1629	2221
				72842F1 (SINCONT01)	848	1394
				563718PH (CRR2HR01)	1	262
				1426405X13F1 (SIFHS901)	275	842
				2072529X1AC1 (SIFHS901)	142	687
				SAFC01647FL	1344	1917
				SAPC01486F1	885	1419
				SBA02284F1	235	924
45	133296VCBI	1619	1-24, 715-1025	859640R1 (BRA1TU03)	555	1133
				295602F8 (KIDNFT01)	1	518
				1576265T1 (LUNG0T03)	981	1619
				1614269F6 (PBCFT012)	617	1147
				4637615H1 (Y7EFT01)	500	740
46	1559410CBI	1448	1-227	789274H1 (COLINERT01)	1333	1448
				1798950P6 (COLN0737)	950	1447
				3520828H1 (LUNG0N03)	33	364
				4062190H1 (BRA1N0T2)	358	529
				3593710H1 (FBEN0T01)	1	207
				SCM05260V1	870	1450
47	1752587CBI	2225	1-44, 563-1515	084895H1 (PTNUN03L)	1	234
				1500773P6 (SERPBT01)	1665	2225
				SCMA05704V1	237	719
				SCMA00926V1	1317	1803
				4864824H6 (PHOSTF09)	1575	2203
				SCMA03952V1	45	497
SCMA04987V1	607	1211				

WO 01/64896

PCT/US01/06806

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Inxyle Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
48	1911595CBI	1400	1-28	244890F6 (DRNGPTL8)	448	919
				1404974F6 (LATRGTU02)	1137	1400
				1563768F6 (BNSNGT04)	726	1384
				2149013H1 (BRANOT09)	1131	1393
				4897260H1 (LIVRVTU12)	580	639
				3911509F6 (CORRVTU1)	1	445
49	2210170CBI	1361	739-1103	SMAD1233D1	342	840
				2210170F6 (STFFPT03)	1	372
				SMAD454D1	500	593
				2210170F6 (SENFPT03)	595	1361
				SKCA00325V1	1	497
				SCA00576V1	611	593
50	23466ACBI	993	1-34	SCA00327V1	525	584
				2172996F6 (BMDGN003)	433	497
				2047322F6 (THPTU1)	166	652
				4288169X302D1 (LIVRDR01)	1	331
				1446241F6 (PLAENG02)	979	1357
				5816107H1 (THWNG02)	949	1258
51	2884114CBI	1387	774-1367	4794583H1 (LIVRVT09)	611	870
				4288169H1 (LIVRDR01)	800	1108
				91894065	356	735
				3356420H1 (PROSTT16)	1	270
				3356420F6 (PROSTT16)	582	1374
				82784609	868	1317
52	4103555CBI	1317	1-32, 76-355	3356420F6 (PROSTT16)	2	555

WO 01/64896

PC/T/US01/06806

Table 5

Polynucleotide SEQ ID NO:	Incyte Env. seq ID	Representative library
27	273442CBL	PARVUT02
28	211530CBL	BLADZF04
29	275224CBL	CONRUF01
30	187094CBL	PROSTUF05
31	138841CBL	LEVEROT01
32	149540CBL	UCKLST01
33	2024617CBL	LEVERUT04
34	472127CBL	SSRFUF08
35	540654CBL	LUNGAST01
36	123272CBL	BRALUT01
37	129930CBL	BRALIF01
38	138519CBL	PANCAST01
39	212165CBL	TLYMCOT06
40	254567CBL	PROSKOT18
41	390348CBL	LUNGROT9
42	539181CBL	LUNGPR03
43	588699CBL	CORFNG02
44	721432CBL	BRALUT02
45	133296CBL	BRALUT03
46	155941CBL	PROSTOT12
47	1752487CBL	STIMBSOT1
48	191150CBL	SLWMCOT2
49	221017CBL	SLWMCOT3
50	234664CBL	ATDMACOT13
51	289416CBL	EMPL7T01
52	410359CBL	PROSKOT16

WO 01/64896

PCT/US01/06806

Table 6

Library	Vector	Library Description
BGALTUT04	PSF0RT1 pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology indicated grade 3 transitional cell carcinoma in the left bladder wall. Carcinoma in-situ was identified in the dome and trigone. Patient history included tobacco use. Family history included Type I diabetes, malignant neoplasia of the stomach, atherosclerotic coronary artery disease, and acute myocardial infarction.
BGALTUT01	PSF0RT1	Library was constructed using RNA isolated from brain tumor tissue removed from a 50-year-old Caucasian female during a frontal lobectomy. Pathology indicated recurrent grade 3 oligodendroglioma with focal necrosis and extensive calcification. Patient history included a speech disturbance and epilepsy. The patient's brain had also been irradiated with a total dose of 5,042 cGy (fraction 8). Family history included a brain tumor.
BGALTUT02	PSF0RT1	Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningioma lesion. Pathology indicated a grade 2 metastatic meningioma. Patient history included a grade 2 renal cell carcinoma, insulinoma, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
BGALTUT03	PSF0RT1	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 47-year-old Caucasian female during excision of a cerebral meningioma lesion. Pathology indicated a grade 4 fibrillary giant and small-cell astrocytoma. Family history included benign hypertension and cerebrovascular disease.
BGALTUT08	PSF0RT1 pINCY	Library was constructed using RNA isolated from breast tumor tissue removed from a 49-year-old Caucasian female during unilateral extended simple mastectomy. Pathology indicated invasive nuclear grade 3 adenocarcinoma, ductal type, with 3 of 23 lymph nodes positive for metastatic disease. Greater than 50% of the tumor volume was in situ, both comedo and non-comedo types. Immunostains were positive for estrogen/progesterone receptors, and uninvolved tissue showed proliferative changes. The patient concurrently underwent a total abdominal hysterectomy. Patient history included a valvulopathy of mitral valve without replacement, rheumatic mitral regurgitation, and rheumatic heart disease. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and Type II diabetes.

WO 01/64896

PCT/US01/06806

Table 6 (cont.)

Library	Vector	Library Description
CONUTU01	P1MVC	Library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed müllerian tumor present in the sigmoid mesentery at two sites.
CORNO002	P1MVC	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
KIDNOU013	P1MVC	Library was constructed using RNA isolated from kidney tumor tissue removed from a 51-year-old Caucasian female during a nephroureterectomy. Pathology indicated a grade 3 renal cell carcinoma. Patient history included depressive disorder, hypoglycemia, and uterine endometriosis. Family history included calculus of the kidney, colon cancer, and type II diabetes.
LIVRNO001	PHLOESCRIFT	Library was constructed at Stratagene, using RNA isolated from the liver tissue of a 43-year-old male.
LIVRNU004	P1MVC	Library was constructed using RNA isolated from liver tumor tissue removed from a 50-year-old Caucasian male during a partial hepatectomy. Pathology indicated a grade 3-4 hepatoma, forming a mass. Patient history included benign hypertension and hepatitis. Hepatitis B core antigen and hepatitis B surface antigen was present in the patient.
LONGAST001	PSFORP1	Library was constructed using RNA isolated from the lung tissue of a 17-year-old Caucasian male, who died from head trauma. Patient history included asthma.
LONGHE003	P1MVC	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
LONGMO009	P1MVC	Library was constructed using RNA isolated from the lung tissue of a 23-week-old Caucasian male fetus. The pregnancy was terminated following a diagnosis by ultrasound of bilateral polycystic kidney disease.
MGCRI001	PSFORP1	Library was constructed using RNA isolated from adherent mononuclear cells, which came from a pool of male and female donors. The cells were cultured for 24 hours following Ficoll Hypaque centrifugation.
PANCTU002	P1MVC	Library was constructed using RNA isolated from pancreatic tumor tissue removed from a 45-year-old Caucasian female during radical pancreatectomy. Pathology indicated a grade 4 anaplastic carcinoma. Family history included benign hypertension, hyperlipidemia and atherosclerotic coronary artery disease.

WO 01/64896

PCT/US01/06806

Table 6 (cont.)

Library	Vector	Library Description
PROST019	pRNCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrectomy. Pathology indicated adenofibromatous hyperplasia; this tissue was associated with a grade 3 transitional cell carcinoma. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
PROST005	PSPOR11	Library was constructed using RNA isolated from prostate tumor tissue removed from a 69-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenocarcinoma (Gleason grade 3+4). Adenofibromatous hyperplasia was also present. Family history included congestive heart failure, multiple myeloma, hyperlipidemia, and rheumatoid arthritis.
PROST012	pRNCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 69-year-old Caucasian male during a radical prostatectomy. Pathology indicated an adenocarcinoma (Gleason grade 2+2). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA).
PROST016	pRNCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 59-year-old Caucasian male. Pathology indicated adenocarcinoma, Gleason grade 3+4. Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Patient history included calculus of the kidney. Family history included lung cancer and breast cancer.
SLN00701	pRNCY	Library was constructed using RNA isolated from ileum tissue obtained from an 18-year-old Caucasian female during bowel anastomosis. Pathology indicated Crohn's disease of the ileum, involving 15 cm of the small bowel. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.
SLN00703	pRNCY	Library was constructed using RNA isolated from small intestine tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
SLN00702	VEL00SCR1PT	Library was constructed using RNA isolated from the small intestine of a 55-year-old Caucasian female, who died from a subarachnoid hemorrhage. Serologies were positive for cytomegalovirus (CMV). Previous surgeries included a hysterectomy.

WO 01/64896

PCT/US01/06806

Table 7

Program	Description	References	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACCEL 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 <5%
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) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	<p> <i>ESTs</i>: Probability value= 1.0E-8 or less <i>Fold Length sequences</i>: Probability value= 1.0E-10 or less <i>ESTs</i>: fasta E value= 0.0E-6 <i>Aligned/ESTs</i>: fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value= 1.0E-8 or less <i>Fold Length sequences</i>: fastx score= 100 or greater Probability value= 1.0E-3 or less </p>
PASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. PASTA comprises at least five functions: fastp, tfastn, fastx, fasty, and search.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	
BLIMPS	A BLIMPS 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.	Hunkler, S. and J.G. Henikoff (1991) <i>Nucleic Acids Res.</i> 19:6965-6972; Henikoff, J.G. and S. Henikoff (1996) <i>Methods Enzymol.</i> 256:88-105; and Atwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-3 or less
BLIMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1998) <i>Nucleic Acids Res.</i> 26:3203-3222; Durbin, R. et al. (1998) <i>Our World View</i> , in a Kunshel, Cambridge Univ. Press, pp. 1-350.	<p> <i>PFAM hits</i>: Probability value= 1.0E-3 or less <i>Signal peptide hits</i>: Score= 0 or greater </p>

Table 7 (cont.)

Program	Description	Reference	Parameter/Threshold
ProfilesScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Grishov, M. et al. (1988) CABIOS 4(6): 66-66; Grishov, M. et al. (1989) Methods Enzymol. 183: 146-159; Barroch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality scores;GCC-specificity "HGCF" value for that particular Prosite motif Generally, scores=1,4=1.
Pirced	A base-calling algorithm that examines automated sequence traces with high sensitivity and probability.	Ewing, R. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Pirap	A Phil's Revised Assembly Program including SWAT and CrossMatch programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1983) J. Mol. Biol. 147:155-197; and Green, P., University of Washington, Seattle, WA.	Scores=120 or greater; Match lengths=56 or greater
Conseq	A graphical tool for viewing and editing Pirap assemblies.	Cordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Chavakis, J.M. and S. Audie (1997) CABIOS 12:431-439.	Scores=3.5 or greater
TM/AP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Reason, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Pearson, B. and P. Argos (1996) Protein Sci. 5:565-571.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Searhammer, J.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Chicago et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Medis	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Barroch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page ME1-39; Genetics Computer Group, Madison, WI.	

WO 01/64896

PCT/US01/06806

WO 01/64896

PCT/US01/06806

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
- 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-26,
b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26,
c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and
10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-26.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:27-52.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
35 b) recovering the polypeptide so expressed.

WO 01/64896

PCT/US01/06806

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
- 5 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52,
b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52,
c) a polynucleotide sequence complementary to a),
d) a polynucleotide sequence complementary to b), and
10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
- 15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization
20 complex is formed between said probe and said target polynucleotide or fragments thereof, and
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
- 25 15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.
- 35

WO 01/64896

PCT/US01/06806

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

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

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

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

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

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

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

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

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

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

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

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
35 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

WO 01/64896

PCT/US01/06806

compound that specifically binds to the polypeptide of claim 1.

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

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

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

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

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

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

WO 01/64896

PCT/US01/06806

<110> INCYTH GENOMICS, INC.
 TANG, Y. Tom
 LU, Dyrung Aina M.
 SANDMAN, Olga
 YUE, Henry
 AZIMZAI, Valda
 TAL, Praseki
 BURFORD, Neil
 BRUGHN, Mariah R.

<120> HUMAN ENZYME MOLECULES

<130> EF-0763 PCT

<140> To Be Assigned
 <141> Herewith

<150> 60/186,307; 60/192,532; 60/193,578
 <151> 2000-03-01; 2000-03-28; 2000-03-30

<160> 52
 <170> PERL Program

<210> 1
 <211> 404
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> incyte ID No: 2783442CD1

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

WO 01/64896

PCT/US01/06806

```

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

```

<210> 2

<211> 242

<212> PFT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 2116390CD1

<400> 2

```

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

```

WO 01/64896

PCT/US01/06806

<210> 3
<211> 404
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5765224CD1

```

<400> 3
Met Asn Asn Ile Lys Pro Leu Glu Gly Val Lys Ile Leu Asp Leu
1      5      10      15
Thr Arg Val Leu Ala Gly Pro Phe Ala Thr Met Asn Leu Gly Asp
20     25     30
Leu Gly Ala Glu Val Ile Lys Val Glu Arg Pro Gly Ala Gly Asp
35     40     45
Asp Thr Arg Thr Trp Gly Pro Pro Phe Val Gly Thr Glu Ser Thr
50     55     60
Tyr Tyr Leu Ser Val Asn Arg Asn Lys Lys Ser Ile Ala Val Asp
65     70     75
Ile Lys Asp Pro Lys Gly Val Lys Ile Ile Lys Glu Leu Ala Ala
80     85     90
Val Cys Asp Val Phe Val Glu Asn Tyr Val Pro Gly Lys Leu Ser
95     100    105
Ala Met Gly Leu Gly Tyr Glu Asp Ile Asp Glu Ile Ala Pro His
110    115    120
Ile Ile Tyr Cys Ser Ile Thr Gly Tyr Gly Gln Thr Gly Pro Ile
125    130    135
Ser Gln Arg Ala Gly Tyr Asp Ala Val Ala Ser Ala Val Ser Gly
140    145    150
Leu Met His Ile Thr Gly Pro Glu Asn Gly Asp Pro Val Arg Pro
155    160    165
Gly Val Ala Met Thr Asp Leu Ala Thr Gly Leu Tyr Ala Tyr Gly
170    175    180
Ala Ile Met Ala Gly Leu Ile Gln Lys Tyr Lys Thr Gly Lys Gly
185    190    195
Leu Phe Ile Asp Cys Asn Leu Leu Ser Ser Gln Val Ala Cys Leu
200    205    210
Ser His Ile Ala Ala Asn Tyr Leu Ile Gly Gln Lys Glu Ala Lys
215    220    225
Arg Trp Gly Thr Ala His Gly Ser Ile Val Pro Tyr Gln Ala Phe
230    235    240
Lys Thr Lys Asp Gly Tyr Ile Val Val Gly Ala Gly Asn Asn Gln
245    250    255
Gln Phe Ala Thr Val Cys Lys Ile Leu Asp Leu Pro Glu Leu Ile
260    265    270
Asp Asn Ser Lys Tyr Lys Thr Asn His Leu Arg Val His Asn Arg
275    280    285
Lys Glu Leu Ile Lys Ile Leu Ser Glu Arg Phe Glu Glu Glu Leu
290    295    300
Thr Ser Lys Trp Leu Tyr Leu Phe Glu Gly Ser Gly Val Pro Tyr
305    310    315
Gly Pro Ile Asn Asn Met Lys Asn Val Phe Ala Glu Pro Gln Val
320    325    330
Leu His Asn Gly Leu Val Met Glu Met Glu His Pro Thr Val Gly
335    340    345
Lys Ile Ser Val Pro Gly Pro Ala Val Arg Tyr Ser Lys Phe Lys
350    355    360
Met Ser Glu Ala Arg Pro Pro Pro Leu Leu Gly Gln His Thr Thr
365    370    375
His Ile Leu Lys Glu Val Leu Arg Tyr Asp Asp Arg Ala Ile Gly
380    385    390
Glu Leu Leu Ser Ala Gly Val Val Asp Gln His Glu Thr His
395    400

```

<210> 4
<211> 126

WO 01/64896

PCT/US01/06806

<212> PFT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1870996CD1

```

<400> 4
Met Gly His Met Leu Leu Pro Phe Arg Leu Gly Leu Gly Gly Pro
  1      5      10      15
Ile Gly Ser Gly His Gln Phe Phe Pro Trp Ile His Ile Gly Asp
  20      25      30
Leu Ala Gly Ile Leu Thr His Ala Leu Glu Ala Asn His Val His
  35      40      45
Gly Val Leu Asn Gly Val Ala Pro Ser Ser Ala Thr Asn Ala Glu
  50      55      60
Phe Ala Gln Thr Phe Gly Ala Ala Leu Gly Arg Arg Ala Phe Ile
  65      70      75
Pro Leu Pro Ser Ala Val Val Gln Ala Val Phe Gly Arg Gln Arg
  80      85      90
Ala Ile Met Leu Leu Glu Gly Gln Lys Val Ile Pro Arg Arg Thr
  95     100     105
Leu Ala Thr Gly Tyr Gln Tyr Ser Phe Pro Glu Leu Gly Ala Ala
 110     115     120
Leu Lys Glu Ile Val Ala
 125

```

<210> 5
<211> 384
<212> PFT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 138841CD1

```

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

```

WO 01/64896

PCT/US01/06806

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

<210> 6

<211> 164

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1485405CD1

<400> 6

Met Asn Pro Ile Val Val Val His Gly Gly Ala Gly Pro Ile
 5 10 15
 Ser Lys Asp Arg Lys Glu Arg Val His Cln Gly Met Val Arg Ala
 20 25 30
 Ala Thr Val Gly Tyr Gly Ile Leu Arg Glu Gly Gly Ser Ala Val
 35 40 45
 Asp Ala Val Glu Gly Ala Val Val Ala Leu Glu Asp Asp Pro Glu
 50 55 60
 Phe Asn Ala Gly Cys Gly Ser Val Leu Asn Thr Asn Gly Glu Val
 65 70 75
 Glu Met Asp Ala Ser Ile Met Asp Gly Lys Asp Leu Ser Ala Gly
 80 85 90
 Ala Val ser Ala Val Gln Cys Ile Ala Asn Pro Ile Lys Leu Ala
 95 100 105
 Arg Leu Val Met Glu Lys Thr Pro His Cys Phe Leu Thr Asp Gln
 110 115 120
 Gly Ala Ala Gln Phe Ala Ala Ala Met Gly Val Pro Glu Ile Pro
 125 130 135
 Gly Glu Lys Leu Val Thr Glu Arg Asn Lys Lys Arg Leu Glu Lys
 140 145 150
 Glu Lys His Glu Lys Gly Ala Glu Lys Thr Asp Cys Gln Lys
 155 160

<210> 7

<211> 602

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2024617CD1

<400> 7

Met Arg Ser Met Lys Ala Leu Gln Lys Ala Leu Ser Arg Ala Gly

WO 01/64896

PCT/US01/06806

1 5 10 15
Ser His Cys Gly Arg Gly Gly Trp Gly His Pro Ser Arg Ser Pro
20 25 30
Leu Leu Gly Gly Gly Val Arg His His Leu Ser Glu Ala Ala Ala
35 40 45
Gln Gly Arg Glu Thr Pro Arg Ser His Gln Pro Gln His Gln Asp
50 55 60
His Asp Ser Ser Gly Ser Gly Met Leu Ser Arg Leu Gly Asp Neu
65 70 75
Leu Phe Tyr Thr Ile Ala Glu Gly Gln Glu Arg Ile Pro Ile His
80 85 90
Lys Phe Thr Thr Ala Leu Lys Ala Thr Gly Leu Gln Thr Ser Asp
95 100 105
Pro Arg Leu Arg Asp Cys Met Ser Glu Met His Arg Val Val Gln
110 115 120
Glu Ser Ser Ser Gly Gly Leu Leu Asp Arg Asp Leu Phe Arg Lys
125 130 135
Cys Val Ser Ser Asn Ile Val Leu Leu Thr Gln Ala Phe Arg Lys
140 145 150
Lys Phe Val Ile Pro Asp Phe Glu Glu Phe Thr Gly His Val Asp
155 160 165
Arg Ile Phe Glu Asp Val Lys Glu Leu Thr Gly Gly Lys Val Ala
170 175 180
Ala Tyr Ile Pro Gln Leu Ala Lys Ser Asn Pro Asp Leu Trp Gly
185 190 195
Val Ser Leu Cys Thr Val Asp Gly Gln Arg His Ser Val Gly His
200 205 210
Thr Lys Ile Pro Phe Cys Leu Gln Ser Cys Val Lys Pro Leu Thr
215 220 225
Tyr Ala Ile Ser Ile Ser Thr Leu Gly Thr Asp Tyr Val His Lys
230 235 240
Phe Val Gly Lys Glu Pro Ser Gly Leu Arg Tyr Asn Lys Leu Ser
245 250 255
Leu Asn Glu Glu Gly Ile Pro His Asn Pro Met Val Asn Ala Gly
260 265 270
Ala Ile Val Val Ser Ser Leu Ile Lys Met Asp Cys Asn Lys Ala
275 280 285
Glu Lys Phe Asp Phe Val Leu Gln Tyr Leu Asn Lys Met Ala Gly
290 295 300
Asn Glu Tyr Met Gly Phe Ser Asn Ala Thr Phe Gln Ser Glu Lys
305 310 315
Glu Thr Gly Asp Arg Asn Tyr Ala Ile Gly Tyr Tyr Leu Lys Glu
320 325 330
Lys Lys Cys Phe Pro Lys Gly Val Asp Met Met Ala Ala Leu Asp
335 340 345
Leu Tyr Phe Gln Leu Cys Ser Val Glu Val Thr Cys Glu Ser Gly
350 355 360
Ser Val Met Ala Ala Thr Leu Ala Asn Gly Gly Ile Cys Pro Lys
365 370 375
Thr Gly Glu Ser Val Leu Ser Ala Glu Ala Val Arg Asn Thr Leu
380 385 390
Ser Leu Met His Ser Cys Gly Met Tyr Asp Phe Ser Gly Gln Phe
395 400 405
Ala Phe His Val Gly Leu Pro Ala Lys Ser Ala Val Ser Gly Ala
410 415 420
Ile Leu Leu Val Val Pro Asn Val Met Gly Met Met Cys Leu Ser
425 430 435
Pro Pro Leu Asp Lys Leu Gly Asn Ser His Arg Gly Thr Ser Phe
440 445 450
Cys Gln Lys Leu Val Ser Leu Phe Asp Phe His Asn Tyr Asp Asn
455 460 465
Leu Arg His Cys Ala Arg Lys Leu Asp Pro Arg Arg Glu Gly Ala
470 475 480
Glu Ile Arg Asn Lys Thr Val Val Asn Leu Leu Phe Ala Ala Tyr
485 490 495
Ser Gly Asp Val Ser Ala Leu Arg Arg Phe Ala Leu Ser Ala Met
500 505 510

WO 01/64896

PCT/US01/06806

```

Asp Met Glu Gln Lys Asp Tyr Asp Ser Arg Thr Ala Leu His Val
515 520 525
Ala Ala Ala Glu Gly His Ile Glu Val Val Lys Phe Leu Ile Glu
530 535 540
Ala Cys Lys Val Asn Pro Phe Ala Lys Asp Arg Trp Gly Asn Ile
545 550 555
Pro Leu Asp Asp Ala Val Gln Phe Asn His Leu Glu Val Val Lys
560 565 570
Leu Leu Gln Asp Tyr Gln Asp Ser Tyr Thr Leu Ser Glu Thr Gln
575 580 585
Ala Glu Ala Ala Ala Glu Ala Leu Ser Lys Glu Asn Leu Glu Ser
590 595 600
Met Val

```

```

<710> 8
<211> 434
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 472182/CD1

```

```

<400> 8
Met Asn Gly Ser Lys Asn Tyr Ala Ser Arg Pro Gly Thr Arg Gln
1 5 10 15
Pro Val Asn Asn Arg Gly Arg Ser Gly Asn Ala Leu Gln Asp Thr
20 25 30
Ser Gly Lys Leu Arg Ile His Lys Cys Lys Gly Pro Ser Asp Leu
35 40 45
Leu Thr Val Arg Gln Ser Thr Arg Asn Leu Tyr Ala Arg Gly Phe
50 55 60
His Asp Lys Asp Lys Glu Cys Ser Cys Arg Glu Ser Gly Tyr Arg
65 70 75
Ala Ser Arg Ser Gln Arg Lys Ser Gln Arg Gln Phe Leu Arg Asn
80 85 90
Gln Gly Thr Pro Lys Tyr Lys Pro Arg Phe Val His Thr Arg Gln
95 100 105
Thr Arg Ser Leu Ser Val Glu Phe Glu Gly Glu Ile Tyr Asp Ile
110 115 120
Asn Leu Glu Glu Glu Glu Glu Leu Gln Val Leu Gln Pro Arg Asn
125 130 135
Ile Ala Lys Arg His Asp Glu Gly His Lys Gly Pro Arg Asp Leu
140 145 150
Gln Ala Ser Ser Gly Gly Asn Arg Gly Arg Met Leu Ala Asp Ser
155 160 165
Ser Asn Ala Val Gly Pro Pro Thr Thr Val Arg Val Thr His Lys
170 175 180
Cys Phe Ile Leu Pro Asn Asp Ser Ile His Cys Glu Arg Glu Leu
185 190 195
Tyr Gln Ser Ala Arg Ala Trp Lys Asp His Lys Ala Tyr Ile Asp
200 205 210
Lys Glu Ile Glu Ala Leu Gln Asp Lys Ile Lys Asn Leu Arg Glu
215 220 225
Val Arg Gly His Leu Lys Arg Arg Lys Pro Glu Glu Cys Ser Cys
230 235 240
Ser Lys Gln Ser Tyr Tyr Asn Lys Glu Lys Gly Val Lys Lys Gln
245 250 255
Glu Lys Leu Lys Ser His Leu His Pro Phe Lys Glu Ala Ala Gln
260 265 270
Glu Val Asp Ser Lys Leu Gln Leu Phe Lys Glu Asn Asn Arg Arg
275 280 285
Arg Lys Lys Glu Arg Lys Glu Lys Arg Arg Gln Arg Lys Gly Glu
290 295 300
Glu Cys Ser Leu Pro Gly Leu Thr Cys Phe Thr His Asp Asn Asn
305 310 315

```

WO 01/64896

PCT/US01/06806

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

<210> 9
 <211> 369
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incomplete ID No: 5406614CD1

<400> 9
 Met Gly Pro Gly Ala Arg Arg Gln Gly Arg Ile Val Gln Gly Arg
 1 5 10 15
 Pro Glu Met Cys Phe Cys Pro Pro Pro Thr Pro Leu Pro Pro Leu
 20 25 30
 Arg Ile Leu Thr Leu Gly Thr His Thr Pro Thr Pro Cys Ser Ser
 35 40 45
 Pro Gly Ser Ala Ala Gly Thr Tyr Pro Thr Met Gly Ser Gln Ala
 50 55 60
 Leu Pro Pro Gly Pro Met Gln Thr Leu Ile Phe Phe Asp Met Glu
 65 70 75
 Ala Thr Gly Leu Pro Phe Ser Gln Pro Lys Val Thr Glu Leu Cys
 80 85 90
 Leu Leu Ala Val His Arg Cys Ala Leu Glu Ser Pro Pro Thr Ser
 95 100 105
 Gln Gly Pro Pro Pro Thr Val Pro Pro Pro Pro Arg Val Val Asp
 110 115 120
 Lys Leu Ser Leu Cys Val Ala Pro Gly Lys Ala Cys Ser Pro Ala
 125 130 135
 Ala Ser Glu Ile Thr Gly Leu Ser Thr Ala Val Leu Ala Ala His
 140 145 150
 Gly Arg Gln Cys Phe Asp Asp Asn Leu Ala Asn Leu Leu Leu Ala
 155 160 165
 Phe Leu Arg Arg Gln Pro Gln Pro Trp Cys Leu Val Ala His Asn
 170 175 180
 Gly Asp Arg Tyr Asp Phe Pro Leu Leu Gln Ala Glu Leu Ala Met
 185 190 195
 Leu Gly Leu Thr Ser Ala Leu Asp Gly Ala Phe Cys Val Asp Ser
 200 205 210
 Ile Thr Ala Leu Lys Ala Leu Glu Arg Ala Ser Ser Pro Ser Gln
 215 220 225
 His Gly Pro Arg Lys Ser Tyr Ser Leu Gly Ser Ile Tyr Thr Arg
 230 235 240
 Leu Tyr Gly Gln Ser Pro Pro Asp Ser His Thr Ala Glu Gly Asp
 245 250 255
 Val Leu Ala Leu Leu Ser Ile Cys Gln Trp Arg Pro Gln Ala Leu
 260 265 270
 Leu Arg Trp Val Asp Ala His Ala Arg Pro Phe Gly Thr Ile Arg
 275 280 285
 Pro Met Tyr Gly Val Thr Ala Ser Ala Arg Thr Lys Pro Arg Pro
 290 295 300

WO 01/64896

PCT/US01/06806

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

<210> 10
 <211> 453
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc feature
 <223> Incyte ID No: 1252792CD1

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

WO 01/64896

PCT/US01/06806

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

<210> 11
 <211> 511
 <212> FRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1299305CD1

<400> 11
 Met Tyr Arg Leu Met Ser Ala Val Thr Ala Arg Ala Ala Ala Pro
 1 5 10 15
 Gly Gly Leu Ala Ser Ser Cys Gly Arg Arg Gly Val His Gln Arg
 20 25 30
 Ala Gly Leu Pro Pro Pro Asp Pro Glu Ala Ser Pro Leu Ala Glu
 35 40 45
 Pro Pro Gln Glu Gln Ser Leu Ala Pro Trp Ser Pro Gln Thr Pro
 50 55 60
 Ala Pro Pro Cys Ser Arg Cys Phe Ala Arg Ala Ile Glu Ser Ser
 65 70 75
 Arg Asp Leu Leu His Arg Ile Glu Asp Glu Val Gly Ala Pro Gly
 80 85 90
 Ile Val Val Gly Val Ser Val Asp Gly Lys Glu Val Trp Ser Glu
 95 100 105
 Gly Leu Gly Tyr Ala Asp Val Glu Asn Arg Val Pro Cys Lys Pro
 110 115 120
 Gln Thr Val Met Arg Ile Ala Ser Ile Ser Lys Ser Leu Thr Met
 125 130 135
 Val Ala Leu Ala Lys Leu Trp Glu Ala Gly Lys Leu Asp Leu Asp
 140 145 150
 Ile Pro Val Gln His Tyr Val Pro Glu Phe Pro Glu Lys Glu Tyr
 155 160 165
 Glu Gly Glu Lys Val Ser Val Thr Thr Arg Leu Leu Ile Ser His
 170 175 180
 Leu Ser Gly Ile Arg His Tyr Glu Lys Asp Ile Lys Lys Val Lys
 185 190 195
 Glu Glu Lys Ala Tyr Lys Ala Leu Lys Met Met Lys Glu Asn Val
 200 205 210
 Ala Phe Glu Gln Glu Lys Glu Gly Lys Ser Asn Glu Lys Asn Asp
 215 220 225
 Phe Thr Lys Phe Lys Thr Glu Gln Glu Asn Glu Ala Lys Cys Arg
 230 235 240
 Asn Ser Lys Pro Gly Lys Lys Lys Asn Asp Phe Glu Gln Gly Glu
 245 250 255
 Leu Tyr Leu Arg Glu Lys Phe Glu Asn Ser Ile Glu Ser Leu Arg
 260 265 270
 Leu Phe Lys Asn Asp Pro Leu Phe Phe Lys Pro Gly Ser Gln Phe
 275 280 285
 Leu Tyr Ser Thr Phe Gly Tyr Thr Leu Leu Ala Ala Ile Val Glu
 290 295 300

WO 01/64896

PCT/US01/06806

```

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

```

```

<210> 12
<211> 520
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 1385190CD1

```

```

<400> 12
Met Ser Glu Ser Gly His Ser Gln Pro Gly Leu Tyr Gly Ile Glu
1 5 10 15
Arg Arg Arg Arg Thr Lys Glu Pro Gly Ser Gly Gly Pro Gln Asn
20 25 30
Leu Ser Gly Pro Gly Gly Arg Glu Arg Asp Tyr Ile Ala Pro Trp
35 40 45
Glu Arg Glu Arg Arg Asp Ala Ser Glu Glu Thr Ser Thr Ser Val
50 55 60
Met Gln Lys Thr Pro Ile Ile Leu Ser Lys Pro Pro Ala Glu Arg
65 70 75
Ser Lys Gln Pro Pro Pro Thr Ala Pro Ala Ala Pro Pro Ala
80 85 90
Pro Ala Pro Leu Glu Lys Pro Ile Val Leu Met Lys Pro Arg Glu
95 100 105
Glu Gly Lys Gly Pro Val Ala Val Thr Gly Ala Ser Thr Pro Gln
110 115 120
Gly Thr Asp Pro Pro Pro Pro Ala Ala Pro Ala Pro Pro Lys Gly
125 130 135
Glu Lys Glu Gly Gln Arg Pro Thr Gln Pro Val Tyr Gln Ile Gln
140 145 150
Asn Arg Gly Met Gly Thr Ala Ala Pro Ala Ala Met Asp Pro Val
155 160 165
Val Gly Gln Ala Lys Leu Leu Pro Pro Glu Arg Met Lys His Ser
170 175 180
Ile Lys Leu Val Asp Asp Gln Met Asn Trp Cys Asp Ser Ala Ile
185 190 195

```

WO 01/64896

PCT/US01/06806

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

<210> 13
 <211> 687
 <212> PRT
 <213> Homc saplens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2133162CD1

<400> 13
 Met Ser Asn Arg Asn Asn Asn Lys Leu Pro Ser Asn Leu Pro Gln
 1 5 10 15
 Leu Gln Asn Leu Ile Lys Arg Asp Pro Pro Ala Tyr Ile Glu Glu
 20 25 30
 Phe Leu Gln Gln Tyr Asn His Tyr Lys Ser Asn Val Gln Ile Phe
 35 40 45
 Lys Leu Gln Pro Asn Lys Pro Ser Lys Glu Leu Ala Glu Leu Val
 50 55 60
 Met Phe Met Ala Gln Ile Ser His Cys Tyr Pro Glu Tyr Leu Ser
 65 70 75
 Asn Phe Pro Gln Glu Val Lys Asp Leu Leu Ser Cys Asn His Thr
 80 85 90

WO 01/64896

PCT/US01/06806

Val Leu Asp Pro Asp Leu Arg Met Thr Phe Cys Lys Ala Leu Ile
95 100 105
Leu Leu Arg Asn Lys Asn Leu Ile Asn Pro Ser Ser Leu Leu Glu
110 115 120
Leu Phe Phe Glu Leu Phe Arg Cys His Asp Lys Leu Leu Arg Lys
125 130 135
Thr Leu Tyr Thr Thr His Ile Val Thr Asp Ile Lys Asn Ile Asn Ala
140 145 150
Lys His Lys Asn Asn Lys Val Asn Val Val Leu Gln Asn Phe Met
155 160 165
Tyr Thr Met Leu Arg Asp Ser Asn Ala Thr Ala Ala Lys Met Ser
170 175 180
Leu Asp Val Met Ile Glu Leu Tyr Arg Arg Asn Ile Trp Asn Asp
185 190 195
Ala Lys Thr Val Asn Val Ile Thr Thr Ala Cys Phe Ser Lys Val
200 205 210
Thr Lys Ile Leu Val Ala Ala Leu Thr Phe Phe Leu Gly Lys Asp
215 220 225
Glu Asp Glu Lys Gln Asp Ser Asp Ser Glu Ser Glu Asp Asp Gly
230 235 240
Pro Thr Ala Arg Asp Leu Leu Val Gln Tyr Ala Thr Gly Lys Lys
245 250 255
Ser Ser Lys Asn Lys Lys Lys Leu Glu Lys Ala Met Lys Val Leu
260 265 270
Lys Lys Gln Lys Lys Lys Lys Lys Pro Glu Val Phe Asn Phe Ser
275 280 285
Ala Ile His Leu Ile His Asp Pro Gln Asp Phe Ala Glu Lys Leu
290 295 300
Leu Lys Gln Leu Glu Cys Cys Lys Glu Arg Phe Glu Val Lys Met
305 310 315
Met Leu Met Asn Leu Ile Ser Arg Leu Val Gly Ile His Glu Leu
320 325 330
Phe Leu Phe Asn Phe Tyr Pro Phe Leu Gln Arg Phe Leu Gln Pro
335 340 345
His Gln Arg Glu Val Thr Lys Ile Leu Leu Phe Ala Ala Gln Ala
350 355 360
Ser His His Leu Val Pro Pro Glu Ile Ile Gln Ser Leu Leu Met
365 370 375
Thr Val Ala Asn Asn Phe Val Thr Asp Lys Asn Ser Gly Glu Val
380 385 390
Met Thr Val Gly Ile Asn Ala Ile Lys Glu Ile Thr Ala Arg Cys
395 400 405
Pro Leu Ala Met Thr Glu Glu Leu Leu Gln Asp Leu Ala Gln Tyr
410 415 420
Lys Thr His Lys Asp Lys Asn Val Met Met Ser Ala Arg Thr Leu
425 430 435
Ile His Leu Phe Arg Thr Leu Asn Pro Gln Met Leu Gln Lys Lys
440 445 450
Phe Arg Gly Lys Pro Thr Glu Ala Ser Ile Glu Ala Arg Val Gln
455 460 465
Glu Tyr Gly Glu Leu Asp Ala Lys Asp Tyr Ile Pro Gly Ala Glu
470 475 480
Val Leu Glu Val Glu Lys Glu Glu Asn Ala Glu Asn Arg Glu Asp
485 490 495
Gly Trp Glu Ser Thr Ser Leu Ser Glu Glu Glu Asp Ala Asp Gly
500 505 510
Glu Trp Ile Asp Val Gln His Ser Ser Asp Glu Glu Gln Gln Glu
515 520 525
Ile Ser Lys Lys Leu Asn Ser Met Pro Met Glu Glu Arg Lys Ala
530 535 540
Lys Ala Ala Ala Ile Ser Thr Ser Arg Val Leu Thr Gln Glu Asp
545 550 555
Phe Gln Lys Ile Arg Met Ala Gln Met Arg Lys Glu Leu Asp Ala
560 565 570
Ala Pro Gly Lys Cys Gln Lys Arg Lys Tyr Ile Glu Ile Asp Ser
575 580 585
Asp Glu Glu Pro Arg Gly Glu Leu Leu Ser Leu Arg Asp Ile Glu

WO 01/64896

PCT/US01/06806

```

590                               595                               600
Arg Leu His Lys Lys Pro Lys Ser Asp Lys Glu Thr Arg Leu Ala
605                               610                               615
Thr Ala Met Ala Gly Lys Thr Asp Arg Lys Glu Phe Val Arg Lys
620                               625                               630
Lys Thr Lys Thr Asn Pro Phe Ser Ser Ser Thr Asn Lys Glu Lys
635                               640                               645
Lys Lys Gln Lys Asn Phe Met Met Met Arg Tyr Ser Gln Asn Val
650                               655                               660
Arg Ser Lys Asn Lys Arg Ser Phe Arg Glu Lys Gln Leu Ala Leu
665                               670                               675
Arg Asp Ala Leu Leu Lys Lys Arg Lys Arg Met Lys
680                               685

```

```

<210> 14
<211> 199
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 254567CD1

```

```

<400> 14
Met Ala Arg His Val Phe Leu Thr Gly Pro Pro Gly Val Gly Lys
1 5 10 15
Thr Thr Leu Ile His Lys Ala Ser Glu Val Leu Lys Ser Ser Gly
20 25 30
Val Pro Val Asp Gly Phe Tyr Thr Glu Glu Val Arg Gln Gly Gly
35 40 45
Arg Arg Ile Gly Phe Asp Val Val Thr Leu Ser Gly Thr Arg Gly
50 55 60
Pro Leu Ser Arg Val Gly Leu Glu Pro Pro Gly Lys Arg Glu
65 70 75
Cys Arg Val Gly Gln Tyr Val Val Asp Leu Thr Ser Phe Glu Gln
80 85 90
Leu Ala Leu Pro Val Leu Arg Asn Ala Asp Cys Ser Ser Gly Pro
95 100 105
Gly Gln Arg Val Cys Val Ile Asp Glu Ile Gly Lys Met Glu Leu
110 115 120
Phe Ser Gln Leu Phe Ile Gln Ala Val Arg Gln Thr Leu Ser Thr
125 130 135
Pro Gly Thr Ile Ile Leu Gly Thr Ile Pro Val Pro Lys Gly Lys
140 145 150
Pro Leu Ala Leu Val Glu Glu Ile Arg Asn Arg Lys Asp Val Lys
155 160 165
Val Phe Asn Val Thr Lys Glu Asn Arg Asn His Leu Leu Pro Asp
170 175 180
Ile Val Thr Cys Val Gln Ser Ser Arg Lys
185 190

```

```

<210> 15
<211> 284
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 3903488CD1

```

```

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

```

WO 01/64896

PCT/US01/06806

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

<210> 16

<211> 245

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5391816CD1

<400> 16

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

WO 01/64896

PCT/US01/06806

Val Ile Pro Leu Lys Asp Val Ser Leu Leu Thr Gln Lys Leu Lys
 185 190 195
 Glu His Cys Lys Val Glu Tyr Gln Ile Lys Thr Phe Ser Gly Gln
 200 205 210
 Thr His Gly Phe Val His Arg Lys Arg Glu Asp Cys Ser Pro Ala
 215 220 225
 Asp Lys Pro Tyr Ile Asp Glu Ala Arg Arg Asn Leu Ile Glu Trp
 230 235 240
 Leu Asn Lys Tyr Met
 245

<210> 17
 <211> 231
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5886989CD1

<400> 17
 Met Phe Leu Val Gly Leu Thr Gly Gly Ile Ala Ser Gly Lys Ser
 1 5 10 15
 Ser Val Ile Gln Val Phe Gln Gln Leu Gly Cys Ala Val Ile Asp
 20 25 30
 Val Asp Val Met Ala Arg His Val Val Gln Pro Gly Tyr Pro Ala
 35 40 45
 His Arg Arg Ile Val Glu Val Phe Gly Thr Glu Val Leu Leu Glu
 50 55 60
 Asn Gly Asp Ile Asn Arg Lys Val Leu Gly Asp Leu Ile Phe Asn
 65 70 75
 Gln Pro Asp Arg Arg Gln Leu Leu Asn Ala Ile Thr His Pro Glu
 80 85 90
 Ile Arg Lys Glu Met Met Lys Glu Thr Phe Lys Tyr Phe Leu Arg
 95 100 105
 Gly Tyr Arg Tyr Val Ile Leu Asp Ile Pro Leu Leu Phe Glu Thr
 110 115 120
 Lys Lys Leu Leu Lys Tyr Met Lys His Thr Val Val Val Tyr Cys
 125 130 135
 Asp Arg Asp Thr Gln Leu Ala Arg Leu Met Arg Arg Asn Ser Leu
 140 145 150
 Asn Arg Lys Asp Ala Glu Ala Arg Ile Asn Ala Gln Leu Pro Leu
 155 160 165
 Thr Asp Lys Ala Arg Met Ala Arg His Val Leu Asp Asn Ser Gly
 170 175 180
 Glu Trp Ser Val Thr Lys Arg Gln Val Ile Leu Leu His Thr Glu
 185 190 195
 Leu Glu Arg Ser Leu Glu Tyr Leu Pro Leu Arg Phe Gly Val Leu
 200 205 210
 Thr Gly Leu Ala Ala Ile Ala Ser Leu Leu Tyr Leu Leu Thr His
 215 220 225
 Tyr Leu Leu Pro Tyr Ala
 230

<210> 18
 <211> 475
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 723432CD1

<400> 18
 Met Ala Ala Leu Thr Thr Leu Phe Lys Tyr Ile Asp Glu Asn Gln
 1 5 10 15
 Asp Arg Tyr Ile Lys Lys Leu Ala Lys Trp Val Ala Ile Gln Ser

WO 01/64896

PCT/US01/06806

```

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

```

```

<210> 19
<211> 471
<212> PRT
<213> Homo sapiens

```

WO 01/64896

PCT/US01/06806

<220>
<221> misc feature
<223> Incyte ID No: 1332963CD1

```

<400> 19
Met Gln Ala Leu Arg His Val Val Cys Ala Leu Ser Gly Gly Val
1      5      10      15
Asp Ser Ala Val Ala Ala Ileu Leu Leu Arg Arg Arg Gly Tyr Gln
20     25     30
Val Thr Gly Val Phe Met Lys Asn Trp Asp Ser Leu Asp Glu His
35     40     45
Gly Val Cys Thr Ala Asp Lys Asp Cys Glu Asp Ala Tyr Arg Val
50     55     60
Cys Gln Ile Leu Asp Ile Pro Phe His Gln Val Ser Tyr Val Lys
65     70     75
Glu Tyr Trp Asn Asp Val Phe Ser Asp Phe Leu Asn Glu Tyr Glu
80     85     90
Lys Gly Arg Thr Pro Asn Pro Asp Ile Val Cys Asn Lys His Ile
95     100    105
Lys Phe Ser Cys Phe Phe His Tyr Ala Val Asp Asn Leu Gly Ala
110    115    120
Asp Ala Ile Ala Thr Gly His Tyr Ala Arg Thr Ser Leu Glu Asp
125    130    135
Glu Glu Val Phe Glu Gln Lys His Val Lys Pro Glu Gly Leu
140    145    150
Phe Arg Asn Arg Phe Glu Val Arg Asn Ala Val Lys Leu Leu Gln
155    160    165
Ala Ala Asp Ser Phe Lys Asp Gln Thr Phe Phe Leu Ser Gln Val
170    175    180
Ser Gln Asp Ala Leu Arg Arg Thr Ile Phe Pro Leu Gly Gly Leu
185    190    195
Thr Lys Glu Phe Val Lys Lys Ile Ala Ala Glu Asn Arg Leu His
200    205    210
His Val Leu Gln Lys Lys Glu Ser Met Gly Met Cys Phe Ile Gly
215    220    225
Lys Arg Asn Phe Glu His Phe Leu Leu Gln Tyr Leu Gln Pro Arg
230    235    240
Pro Gly His Phe Ile Ser Ile Glu Asp Asn Lys Val Leu Gly Thr
245    250    255
His Lys Gly Trp Phe Leu Tyr Thr Leu Gly Gln Arg Ala Asn Ile
260    265    270
Gly Gly Leu Arg Glu Pro Trp Tyr Val Val Glu Lys Asp Ser Val
275    280    285
Lys Gly Asp Val Phe Val Ala Pro Arg Thr Asp His Pro Ala Leu
290    295    300
Tyr Arg Asp Leu Leu Arg Thr Ser Arg Val His Trp Ile Ala Gln
305    310    315
Glu Pro Pro Ala Ala Leu Val Arg Asp Lys Met Met Glu Cys His
320    325    330
Phe Arg Phe Arg His Gln Met Ala Leu Val Pro Cys Val Leu Thr
335    340    345
Leu Asn Gln Asp Gly Thr Val Trp Val Thr Ala Val Gln Ala Val
350    355    360
Arg Ala Leu Ala Thr Gly Gln Phe Ala Val Phe Tyr Lys Gly Asp
365    370    375
Glu Cys Leu Gly Ser Gly Lys Ile Leu Arg Leu Gly Pro Ser Ala
380    385    390
Tyr Thr Leu Gln Lys Gly Gln Arg Arg Ala Gly Met Ala Thr Glu
395    400    405
Ser Pro Ser Asp Ser Pro Glu Asp Gly Pro Gly Leu Ser Pro Leu
410    415    420
Leu

```

<210> 20
<211> 310
<212> PRT

WO 01/64896

PCT/US01/06806

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1559410CD1

<400> 20

```

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

```

<210> 21

<211> 194

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1752587CD1

<400> 21

```

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

```

WO 01/64896

PCT/US01/06806

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

<210> 22

<211> 349

<213> PRI

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1911509CD1

<400> 22

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

WO 01/64896

PCT/US01/06806

```

375          280          285
Pro Asp Ile Pro Val Ser Val Ile Phe Gly Ala Arg Ser Cys Ile
290          295          300
Asp Gly Asn Ser Gly Thr Ser Ile Gln Ser Leu Arg Pro His Ser
305          310          315
Tyr Val Lys Thr Ile Ala Ile Leu Gly Ala Gly His Tyr Val Tyr
320          325          330
Ala Asp Gln Pro Glu Glu Phe Asn Gln Lys Val Lys Glu Ile Cys
335          340          345
asp Thr Val Asp

```

```

<210> 23
<211> 245
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2210170CD1

```

```

<400> 23
Met Gly Ile Trp Gln Arg Leu Leu Leu Phe Gly Gly Val Ser Leu
1          5          10          15
Arg Ala Gly Gly Gly Ala Thr Ala Pro Leu Gly Gly Ser Arg Ala
20          25          30
Met Val Cys Gly Arg Gln Leu Ser Gly Ala Gly Ser Glu Thr Leu
35          40          45
Lys Gln Arg Arg Thr Gln Ile Met Ser Arg Gly Leu Pro Lys Gln
50          55          60
Lys Pro Ile Glu Gly Val Lys Gln Val Ile Val Val Ala Ser Gly
65          70          75
Lys Gly Gly Val Gly Lys Ser Thr Thr Ala Val Asn Leu Ala Leu
80          85          90
Ala Leu Ala Ala Asn Asp Ser Ser Lys Ala Ile Gly Leu Leu Asp
95          100          105
Val Asp Val Tyr Gly Pro Ser Val Pro Lys Met Met Asn Leu Lys
110          115          120
Gly Asn Pro Glu Leu Ser Gln Ser Asn Leu Met Arg Pro Leu Leu
125          130          135
Asn Tyr Gly Ile Ala Cys Met Ser Met Gly Phe Leu Val Glu Glu
140          145          150
Ser Glu Pro Val Val Trp Arg Gly Leu Met Val Met Ser Ala Ile
155          160          165
Glu Lys Leu Leu Arg Gln Val Asp Trp Gly Gln Leu Asp Tyr Leu
170          175          180
Val Val Asp Met Pro Pro Gly Thr Gly Asp Val Gln Leu Ser Val
185          190          195
Ser Gln Asn Ile Pro Ile Thr Gly Ala Val Ile Val Ser Thr Pro
200          205          210
Gln Asp Ile Ala Leu Met Asp Ala His Lys Gly Ala Glu Met Phe
215          220          225
Arg Arg Val His Val Pro Val Ser Val Tyr Ser Phe Thr Val Lys
230          235          240
Asn Ile Lys Leu Phe
245

```

```

<210> 24
<211> 255
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 234664CD1

```

```

<400> 24

```

WO 01/64896

PCT/US01/06806

```

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

```

<210> 25

<211> 377

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2884114CD1

<400> 25

```

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

```

WO 01/64896

PCT/US01/06806

```

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

```

```

<210> 26
<211> 403
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Eocyte ID No: 4103559CD1

```

```

<400> 26
Met Glu Leu Ser Tyr Gln Thr Leu Lys Phe Thr His Gln Ala Arg
1      5      10      15
Glu Ala Cys Glu Met Arg Thr Glu Ala Arg Arg Lys Asn Leu Leu
20     25     30     35
Ile Leu Ile Ser His Tyr Leu Thr Gln Glu Gly Tyr Ile Asp Thr
40     45     50     55
Ala Asn Ala Leu Glu Gln Thr Lys Leu Gly Leu Arg Arg Phe
60     65     70     75
Glu Val Cys Asp Asn Ile Asp Leu Glu Thr Ile Leu Met Glu Tyr
80     85     90     95
Glu Ser Tyr Tyr Phe Val Lys Phe Gln Lys Tyr Pro Lys Ile Val
100    105    110    115
Lys Lys Ser Ser Asp Thr Glu Asn Asn Leu Pro Gln Arg Cys Arg
120    125    130    135
Gly Lys Thr Arg Arg Met Met Asn Asp Ser Cys Gln Asn Leu Pro
140    145    150    155
Lys Ile Asn Gln Gln Arg Pro Arg Ser Lys Thr Thr Ala Gly Lys
160    165    170    175
Thr Gly Asp Thr Lys Ser Leu Asn Lys Glu His Pro Asn Gln Glu
180    185    190    195
Val Val Asp Asn Thr Arg Leu Glu Ser Ala Asn Phe Gly Leu His
200    205    210    215
Ile Ser Arg Ile Arg Lys Asp Ser Gly Glu Glu Asn Ala His Pro
220    225    230    235
Arg Arg Gly Gln Ile Ile Asp Phe Gln Gly Leu Leu Thr Asp Ala

```

WO 01/64896

PCT/US01/06806

Ile	Lys	Gly	Ala	Thr	Ser	Glu	Leu	Ala	Leu	Asn	Thr	Phe	Asp	His	185
															190
Asn	Pro	Asp	Pro	Ser	Glu	Arg	Leu	Leu	Iys	Pro	Leu	Ser	Ala	Phe	195
															200
Ile	Gly	Met	Asn	Ser	Glu	Met	Arg	Glu	Leu	Ala	Ala	Val	Val	Ser	205
															210
Arg	Asp	Ile	Tyr	Leu	His	Asn	Pro	Asn	Ile	Lys	Trp	Asn	Asp	Ile	215
															220
Ile	Gly	Leu	Asp	Ala	Ala	Lys	Gln	Leu	Val	Lys	Glu	Ala	Val	Val	225
															230
Tyr	Pro	Ile	Arg	Tyr	Pro	Gln	Leu	Phe	Thr	Gly	Ile	Leu	Ser	Pro	235
															240
Trp	Lys	Gly	Leu	Leu	Leu	Tyr	Gly	Pro	Pro	Gly	Thr	Gly	Lys	Thr	245
															250
Leu	Leu	Ala	Lys	Ala	Val	Ala	Thr	Glu	Cys	Lys	Thr	Thr	Phe	Phe	255
															260
Asn	Ile	Ser	Ala	Ser	Thr	Ile	Val	Ser	Iys	Trp	Arg	Gly	Asp	Ser	265
															270
Glu	Lys	Leu	Val	Arg	Val	Leu	Phe	Glu	Leu	Ala	Arg	Tyr	His	Ala	275
															280
Pro	Ser	Thr	Ile	Phe	Leu	Asp	Glu	Leu	Glu	Ser	Val	Met	Ser	Gln	285
															290
Arg	Gly	Thr	Ala	Ser	Gly	Gly	Gln	His	Glu	Gly	Ser	Leu	Arg	Met	295
															300
Iys	Thr	Glu	Leu	Leu	Val	Gln	Met	Asp	Gly	Leu	Ala	Arg	Ser	Glu	305
															310
Asp	Leu	Val	Phe	Val	Leu	Ala	Ala	Ser	Asn	Leu	Pro	Trp			315
															320
															325
															330
															335
															340
															345
															350
															355
															360
															365
															370
															375
															380
															385
															390
															395
															400

<210> 27
 <211> 1478
 <212> RNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> InCyte ID No: 2783442CB1

<400> 27
 ctccagagaa ggccctgtggg agtctcggag acgtgtctgt ctgtgaggcg ctgpytgcac 60
 gtccccaggg ctctgggcta ggaagcagc ggcgaggtgc ctcccacgt acccctggcg 120
 gcccaccccg agcaacatgg ggcgaagggc ggcgcgaagg cccgggctgg gagcgttggc 180
 ggccggagtc ccagccatgg cggagctctgt ggagcgcctg cagcagcggg tccuggagct 240
 gggcggggaa cttgccacgg agaggagctct gcaggtcccg aggagcggcg accgaggggg 300
 cggccggggtc cgcctcggag agatgagctc agaggtgggt gattcgaac cctacagccg 360
 ctgtatggca tgaaacgaa tgggaattgt aagcactat gagaatccc gtacccttgc 420
 cgtagcaata gttaggttg gtggcctgg tagtctact gctgaatgc tgaacagatg 480
 tggcattggt aggttcttc tottgnatg tgaacaggtg ganttagcca atatgaatg 540
 attttcttc caacctcacc aagcaggatt aagtaagtt caagcagcag aacatactct 600
 gaggaacatt aatccctgat ttctttttga agtaaccac tataatataa ccacagtggg 660
 aaactttcaa catttcctgg atagaataag taatggtggg ttagaagaag gaaacacctg 720
 tgatctagtt cttagctgtg tggacaatt tgaagctcga atgacaataa atacagcttg 780
 taatgaactt gacaaaatc ggtgggactc tgggtcagt gaaaatgcag ttccagggca 840
 tatacagctt ataattctg gagaatctgc ttgttttggc tgtgtccac caactgtagt 900
 tgcagaatc atgatgaaa agcctcggaa ccgagaaggt gtttggcag ccagcttcc 960
 taacctatg ggttggttg ctggatctt agcaaaaac gtgtlaaagt ttctgtaaa 1020
 ttctgtaob gttagttttt acctggata caatgcaatg caggattttt ttctactat 1080
 gtcacatgag ccaactctcc agtctgatga cagaatctgc aggaagcagc aggaggaata 1140
 taagaaaaag gtagcagcac tgcctaaca agaggttata caagaagagg agagagcaat 1200
 ccatgaagat aatgaatggg gtattgaact ggtatctgag gtttcagaag aggaactgaa 1260
 aaatthtca ggtccagtcc cagacttacc tgaaggaatt acagtggcat acacaattcc 1320
 aaaaagcaag gaagattctg toactgagtt aacagtgaaa gattctggtg aaagcttggg 1380
 agacctcagc gcaaaaatga agaatatgta gataatggac tgggatacat tgcattctc 1440
 atgtlaaac cctctccctt gaaattaaaa aagaatta 1478

<210> 28

WO 01/64896

PCT/US01/06806

<211> 1183
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2116396CB1

<400> 28
ccacagcgtc cgcaccagcg tccgatccca ggcataagcgc cgcgcgcaaa gcgctgggga 60
gatcggaggg cgcgggggct cctggtccgg gccatgtctg gctgtagatc tccgggggga 120
gactgttggt ccggggagtg cgcgcgcag gacaaggagc atccaagata cctgatccca 180
gcactttgca aacagtttta ccattbhggc tgggtcactg ggcctggagg aggaattagc 240
ttgaagcagc gggatgaaat ctacattgct ccttcaggag tgcacaagga agcaattcag 300
cclgaagaca tgtttgtttg tgatataaat gaaaaggaca taagtggacc ttcgccatcg 360
aggagctna aaaaagcna tggtaactct cttttcaaga atgcttcaac aatgagagga 420
ggcggcagc tgaatcctac ccaactctaa gctgctgaga tggcaacct cctcttcca 480
ggcaggagc ttaaaattac acatcaagag atatacaag gataaagaa atgtaacttc 540
ggaggctatt atagatata tgaatgtta tggtaacca ttattgaaa tacacctgag 600
gagaagaac tcaagatag aatgctcat gcaztgaat aatcccca cctcctgca 660
gtactgta cagcicatg aglalatlg tgggggaaa catgggaga gcccaaac 720
atgctgagc gttatgacta tllalattgat attgacct caatgaaga agtagactt 780
gactctccac agctccagc tggagaaat ggaattgct agcnaaaag caagtctaat 840
tatacaaga gataagcna aagcaata ttattcaat gaagctatt ttttaaa 900
aatgaaat tttatgat ctactaall gccactaaat actgcaast glcactctg 960
atctctcag acattggag ttatttgcct atattctat aatttaaat gagggcaag 1020
tgaatgaa atttalact ctatgttct gttttttt aaacctta cgcnaaata 1080
ttgtctta atttcttt tatactact ctgaggaat tctctaat ttttaagat 1140
gctggtata aaaaaatca ttagaaatt taaaaaaa aaa 1183

<210> 28
<211> 1573
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5785224CB1

<400> 28
ggcaccgctd ggcagggctg cagctctgcs cagaacctgc ctctctccg gccgggggg 60
cgggagggcg ctgtggactg gccgcgcga gtccgatatg acaaatata agcaattgga 120
agggataaaa atctggatc taacagagc cctggcggga ccttttcta ctatgaattt 180
aggagatcct ggagcagaag ttataaaagt ggagagaca gtagctggt atgataccag 240
aacctggggg ccacctttg ttgggacaga aagtacatb talctcagtg ttaaccgaaa 300
taaaaaaagt attgctgta atataagga tccaaaagg tgaaaatca tcaaaagct 360
tgcagctgtt tggatgtgt tggggaaaa ctatgtctct gccaacctg ctgcaatggg 420
cctggatata gaagatag agagattgc tctcaalc atctatggt ccaacaagg 480
gtatgctcag acaggtccaa ttctcaggc agcgggtat galctgttg cctcgcctg 540
ttctggtctg atgcaatca cagggctcga gaatggagc caagtctcc cuggagtagc 600
tatgactgal ctggcaactg gctgtatgc atatggagc alctggctg gattgatata 660
aaatatacaa actgggaag gactgtleal tgaatgtaac clactgtcat cccaggtggc 720
gtgtttgtct caaactatct caaattatct tattggtcaa aaggaagca aagctgggg 780
taacgtctat gcaagtatg ttectacca ggcctttcaa accaaggatg gctatattgt 840
agctggagca ggaataacc agcagtttgc caccgtctgc agatcttgg atttgcclga 900
gttgattgat aatccaagt ataaaacta ccaactccg gtaaccaata gaaagagct 960
tataaata tatarcgaac ggtttgaagc agaacctgac agcaatggt tatatcttt 1020
tgaagcagc ggactccctg atggcccaat caacaactg aagaatgtat ttgcagacc 1080
tcaggtatta ccaatggcc tonttatgga gatggagcat ccaactgtgg ggaagatctc 1140
cgtcccaggc ccagctgta gatacagtaa gttcaagat tcaagggca ggcgccccc 1200
gctgctggg cagcacacaa cgcacatct gaaggaggtc ctgagatacg atgacagggc 1260
cactggggag ctgctcagc ctggagggg gnaaccaat gaactcact gacaagaa 1320
aagctctctt cclcataacc tggatccgaa taccctggca aagcaaac tttgcttga 1380
cctctctccc cagtctgat accactaaa agaaatlla ggttaact agatcttca 1440
cctgctatct caagatgca tctggtata atgaatag tgcctttta atgatacca 1500
cgtttgttc ccaactctb ttttttccag atgagattt cattatgat tbtgggatt 1560
tttaaaata aa 1572

WO 01/64896

PCT/US01/06806

<210> 30
<211> 1335
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Inyte ID No: 1670994CB1

<400> 30
ggacgggaagc ggcgccacc ccaactccac gggaaalatac acaalllglc ltttccgac 60
aggtgggggg acaggcttca ttgggacagc cctaacccag ctgctgaatg ccagaggcca 120
cgaaagtacc ttggtctccc gaaagccccc gccccggccg atcaagtggg atgagctcgc 180
tgcctcgggg ctgcccagct gcgatgccgc cgtcaacctg gccggagaga acatcctcaa 240
ccctctccga aggtccagcc cgggcccata agctgatacc cactagagca caggagggaac 300
agtgcccac tgaigagaa cgtgageta hgtatggat gaaccttcc aaaaagagt 360
tctcggcagc cgtctagaga ccaaccaatt gctgcttasa gccatcccca aagcccaca 420
accocccasg gactgggtct tagtccasg tctagcttac taccagccca gctctgactc 480
gggtatgat gaagacagcc caggaggaga ctttgaattt ttctccacc tcttaaccda 540
atggaaact ccagccagc tctctggaga tctacacgc cagtgglgg tgcctcagg 600
ggttgtctg ggcctgggg ggggtccat gggccacat ctgctgcct tccgctggg 660
cctgggggg cccatggct caggccacca attctccccc tggatataca tgggggaact 720
ggcggaaac ctgaccatg ccttgaagc aaaccacgtg cccgggtcc tgaatggagt 780
ggctccacc tccgcacta agctgagtt gccccagcc tccggtctg cccgggccc 840
cggagcctc atccctctcc ccagcctgt ggtccagct gctctgggc gacggctgc 900
calcatctg ctggagggc agaaagctat cccacggcga acactggcca ctggctacca 960
gtattctct ccagagctag gggctgcctt aaagyaalt gtatccaaag tagtctatg 1020
caagggcctg agcctgttc ctccacaggt tccaggttag ccactgtgaa taggtccagc 1080
tctctcagag agctgaagcc atctggttct tagatctcc tccagctct ctttcccaat 1140
gttctgttgc tccaacttat tctctcaagg ccgtaatct atcaggttgy gacatbaac 1200
tttccaactc ctctgaagt ctccgggttt gttttotota catgctctgc agctgacca 1260
cttcccttt agctgtgta gaaatgctc tcaagtttag caataaaaa taatctgtc 1320
cactaaaaa aaaaa

<210> 31
<211> 1341
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Inyte ID No: 138641CB1

<400> 31
gcaagggca ggcggttctg gcgaggacc aagcactggc gacccgtggc catggcggc 60
gctgagtgga agtctgttga ggaatgcttg gagaagccc tgcctctccc gactctgag 120
gaagtgagcc ggtlctcta tggcaggaaa ctccaggacc ttgactctcc cagggaagct 180
ttcgaagctg cctccagaga agatttga actcagggat agcctttga agcagggag 240
gagcagctga gacagcccg cactgtgccc gtggggctgy ttcagaccag aatccccc 300
ccycaaatg cccctgtgpc ayaacagctc tctgcccctc atagagccat aaagctctc 360
gtagaggtg ctgcaatgtg lggagtcac atcatctgtt tccaggaagc atggaatly 420
ccctctgctc tctgtaagag agaaagcct ccttggaccg aatttctgta gtcagagag 480
gatgggcca ccaacagatt ctgtcaagag ctggggaaga accatgacat ggtggtggtg 540
tctcccaacc tggaaacgaga caggagcact gggatgttt tctggaatac agccgtggtg 600
atctccact cggagcagtt cctgggaag accagaaaa accacalccc cagaglyggt 660
gattccag agtcaacta ctactggag gaaacctgy gcaacccgt gttccagag 720
cagttcggaa ggtctgggt gacatcttc taccggggcc acccccct caactggct 780
atgtacagca tcaacggggc tgagatcac ttcaccctcc cggccacgat aggaactc 840
agcagctcc tgtggccat cagggccaga aacggagcca ttgccaalca ctgctcacc 900
tgcgccata atcaggtggg ccaggagcc tcccgnaag agtttaactc gggagatgga 960
aagaaagctc accaggaact tggcacttt batggctcga gctatgtgg agccctcag 1020
agcagccyga ctctgggtt gtcctgagc cgggatggac tctagttc taagctcag 1080
ctcaaccct gccagcaggt gaatgtgct tggaaclta agatgacggg caggtatgag 1140
agthacgac gggagctcgc cgaagctgc agtccact aagmccac cactgtgaa 1200
gaatcccg ctccagtcac tccctggg ttaggaagac acctctgccc cagtgatta 1260
gcaagtgtg cagcttaac atgtccaggt tctcccaat aacatgtcc aggttggtt 1320
taaaattccc agccaaggg a

WO 01/64896

PCT/US01/06806

<210> 32
<211> 755
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Inseyte ID No: 1485405CBI

<400> 32
ggctgagcgg ttctgagccg gctctgggga gggggcgtc cgggcggctg cggggctggc 60
tcgaccagc ttgaggctct ggcgtccgcg tctgcygtg cctggggtc tccggagac 120
cttgaccgc cggcgttcc tggggctggc ttggaccgc gcttctgct tctctgctgc 180
taggatccgc cgcacatgaat cccatcgtag tggctccagg cggcggagcc ggtcccatct 240
ccagagctcg gaaggagcga gtgcaccagc gcatggtcag agccgccacc gggggctcag 300
gcatctctcg gggggcggg agcgcgctgg atgctcaga gggcctgic gtcgctctgg 360
agcgcgalec cggcttcaac gcaggctgca ggtctgtctt gaacacaaat ggttaagctg 420
aatggatgc tagtatcatg gatgaaagc acctgtctgc aggagcagtg tccgcagctc 480
agtgtatagc aatcccatc aaactctctc gctctgtcat ggaagagaca cctcattgct 540
ttctactga ccaagccca gcgcagcttg cagcagctat gggggtcca gacttctctg 600
yagaaaaact ggtgacagag aaaaacaaa agcgcctgga aaaaagaaag catgaaaaag 660
gtgctcagaa aacagatgtt canaagtaag tcttacctg ggtctgcatt atttgggagt 720
tallaaaaa tgaagcttg gcaataaaa aaaaa 755

<210> 33
<211> 2390
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Inseyte ID No: 2024617CBI

<400> 33
ggcgcacctg cagagccgga ggttctggg gcatgcctc catpaaggct ctgcagaagc 60
ccctgagcgg ggcctggcagt cctctggggc gaggaggctg gggctccccc agcggagcc 120
ccctccctcg cggggcgctc cggcaccacc ttagtgaggc agggccagc ggcagagaga 180
cgcaccagcg ccaccagcgc cagcaccagg atcatgatl cccagaagc ggcctgctgl 240
ccggcctggg tgatttctct ttttaccata ttgtctgagg accggaaccg atccctatcc 300
acaagtctac cactgcacta aagggcaatg gactgcagc atcagatctc cggctccctg 360
atgcatgag cagatgcac cgcctggctc aagaatccg tagtggctgc ctctgggacc 420
gagatctctt ccgaagctgt gtgagcagca acattgtctc nctgaccagc gaattccgaa 480
agaattctgt cattctctgt ttgagagatc tccgggcca tctggatcgc atcttctggg 540
atgctcaaga gctcactgga ggcgaagtgg cagcctacat cctcagctg gcccaagcca 600
accagacct gtggggtgtc tccctgtgca ctgtgcatg tcaaccgacc tctgtggccc 660
acacaagat cccctctctc ctgcagctct gctggaagcc cctccacctat gccatctcca 720
taagaccctt agcactgac lacgtgcaca agtctgtggg caaagagcca aglygctctc 780
gctcaacaaa gctctcctc aatgagagag gaatccccc taaccaccg gtaaatgctg 840
gtgctcttct tgaagctcc ctgatcaaga tggactgtan caaagcagag aagtttgatt 900
ttgtgttgcg glatctcaac aaaaaggctg ggaatgaat calgggttcc agcaatacca 960
catctcagtc agagaagga acaggggac ggaattatgc catcgctctat tatctccagg 1020
aaagaagctg ctttccctag ggggtggana tgaatggctg ccttgatctc tacttccagc 1080
tgtgtctctg gtaggtcact tgtgaatcag gcaatgcat ggcagccacc ctgcgcaagc 1140
gtaggactct cccatccaca ggcagagctg tctgagctg tgaagcagtg cgcacacccc 1200
tcagcctcat gcatctctgc ggcctgtatg actctctctg ccagtttccc tcccaagctg 1260
gctctgagc caagctagct gtaacaggg ccatctctct ggggtaccg aagtctatg 1320
gaatgatgta cctgtaccac ccattggaca agctgggaaa cagccatagc gggaccagct 1380
tcgccaagaa gttggtgtct ctcttcaatt tcccaacta tgacaaccty aggcactgtg 1440
ctggagatg agaccacagc ctgtaagggg cagaatctg gaacaagact gttgtccacc 1500
tgttatctgc tgcctatagt ggcgatgtct cagctctctc aaggtttgoc ttgtcagcca 1560
tgatattgga acagaagac tatgactcgc gcaagctctc gcatgttctc gcactgaag 1620
gcaacatcga agttgttaaa tctctgacty aggtttgcaa agtgaatctt ttgcccaagg 1680
accagctggg caacatcccc ctggatgatg ctgtgcagct caaccatcty gaggctgtca 1740
aacctgctca agatccacc gactcctaca cactctctga aactcaagct gaggcagcag 1800
ctgaagctcc gtcacaagag aacttgaaca gcatgttatg accacaagtc atggaagccc 1860
cctgctcaag aaaaagctgt agctggccac acatgtaatc cataaccacc aaaaactc 1920
tggagagcta cactgcthoa gttgggagcca agcagctcat tgggtgactta ggtatgctg 1980

WO 01/64896

PCT/US01/06806

```

ttctatggga gtcaaaatac cccattccct cagcagcag agtaccadaga agggccctcag 2040
aggcaaccctg cagtacagct atccagagag actgggcttc aaggtacacg ctaatygctt 2100
gcccaccaca aaaccatccc agctcttcac ccaggtctcc tcttcctctc cctgaagaaa 2160
ccatcatcag agagatactc tgggtggagg actctagctc ccatgcacat gtacatactc 2220
acagaatag gaaagtggga ctggtctatc acatggcttt agtctctgg agaaatctac 2280
tonccttggc caggacalgc tctctctact gctaacagcc aatllalag acagagaag 2340
tcttttggc tcaatnaac ttttaattacc aatcaaaaa aaaaaaaa 2390

```

```

<210> 34
<211> 2940
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Invt ID No: 4721827cb1

```

```

<400> 34
caaaatttgg cptgatacat tctctagcga aagagcaca tttctacgta agaaggaaga 60
atccagcaag aatattccaa agtcaaatca ctggcccaaa tatgaacggg tcaagaact 120
atgccagcag gccaggctacc agacagcctg tgaacaaccg gggcagaagc gccaatgcat 180
tgcaaggata abctggcaag ctctgaattc acaagtgtaa aggaaccagc gacctgctca 240
cagtccggga gagcagcggg aacctctaac ctccggcttt ccatgcaaaa gacaagagat 300
ggagctctag gtagctctgt taccgtgcca gcagagctca aagaaatggt caacggcaat 360
tcttgagaaa ccaggagact ccaaaagtac agccagattt tctcaatcct cggcagaac 420
gttctctgtc cgtcgaattt gaaggtgaaa tatatgacat aatctggaa gaagaagaa 480
aatctcaagl gttgcaacca agaaccattg ctaagctcca tgatgtaggc ccaaggggc 540
caagagatct ccaggcttcc agtgggtgca acagggcag gatctggca gatagcaac 600
accgctggg cccacttacc actgtccgag tgaacacaa gtglltlat ctcccaatg 660
actctatcca ttgtgggaga gaactgtacc aatggccag agcgtggag gccataaag 720
cacaacttga caaagagall gaagctctgc aagaataat laagaallta agagaagtg 780
gagacatct gaagagaag aagctctgag aatgtagcty cagttaacaa agctalltaca 840
ataagagaaa aggtgtaaaa agcgaagcga aatgaagag cctcttcc acattcagc 900
aggctgctca ggaagttagt agcaaacctc aacttttcaa ggaacacac cgttaggaga 960
agaagggag gaagagaag agcagcaga ggaaggggaa agagtgcag ctgctggcc 1020
tcacltgccl cagcagatgac acaaccaact gccagcagc ccgctctgg aacctggat 1080
cttctctgyc ttgcacagat tctaacata acacttacty gtgttgcgt acagttaatg 1140
agacgcabaa tttctcttct tgtgagtttg ctactggctt ttggagtat tttgatatga 1200
atacagatcc ttatcagctc acaaatcac tgcacagggt agucagggc attttgaatc 1260
agctacagct acactcaaly gactctaga gctgtcaag atataacag tgcacacaa 1320
gaactaaaga tcttgatgtt ggaatcaag atggaggag ctatgacata cacagaggac 1380
agtatagga tggatggaaa gtttaetrag cccctctca ctgcagacat caalggca 1440
ggcctagagg agctacacag tgtgagttaa accctctaty agtcaagaca aaaaacaga 1500
cttagcctgg tggactggac taattacttg aaggttttag atagatatt tgcactgcol 1560
gaagagtca ctagagcaaa ataaaaaaa taagactcaa actgctcaa gtgacgggtt 1620
cttggttgct tctgctgagc acgtctgtc aaaggagatg gctctgctg actcagatga 1680
agaccagaag cattaagttg gaaaaccacc taatttgacc ttgcccagcty accttcaaac 1740
ctgactttg accagcaaa catcaagctc agagagctaa ctgaaatgga aaaaagacat 1800
tccagaagtt aatcatttga attctgaaca ctgagaaaa accgaaaaat ggcctgggca 1860
tgaaagact aatcactctg aaaccgattt cagtggcag gccatgacag agctagagct 1920
cggcccaagc cccaggtctc agccatttg aggcaacag aagaaactc cccagtatg 1980
tggctctgga aagacattt ttgaagata actatctct ctgtgcatc cccagtgaat 2040
tccggttcat cagatgttca ccatgcccac cgcagaacac cgaatattt ccagcatagc 2100
ggggagatg ttgccaaggt tggagaagaa tcaagaaaag gagaagtca cccacctaga 2160
agggagggcc tcccllccc tctctcttga ltagatgaaa ctgttaectt acctaaaca 2220
cagtattct ttttaacttt ttatttga aactaaana gtaactaca gccaccaca 2280
ttccangctn cctctggtag ctcttctgag tgaagctag tggactctg agccagcgt 2340
gtgcaacagg agactcactg ttataattt ctactctcca agagtgaag gaaagctg 2400
ggatatttgg gttggcttgg ttttgaattt ttgctgtttt gtttgtttt taataaaa 2460
gtatattct ttgaatacag lagggacata agtatataca tttatccaa tcaagatggc 2520
tagaatggt ccttctctgag tctcaaaa cttgaccccc tggtaaatc ttaaacac 2580
ttcaactgcc tgcgtaatga agttttgntt cahttttano cantggaatt tttcaatg 2640
gtcaalltca gttagatgat ttgcaacttt gagattaaaa tggcaatgct atttgatag 2700
tcttatttt ttatttcc agctctaca gctcactc tggctgctt tctgcaaa 2760
tcaataaac ccccaaggac gcaaacagf alggatcaaa ttttggttga catlaagct 2820
ttgcaagaaa atgttgcagt ttttttact cgaactgcta aatctgatta gcaaaaagg 2880
atggtctata atgttggtyg tgaataaau laantaagta aacnaaaaa aaaaaaaa 2940

```

WO 01/64896

PCT/US01/06806

```

<210> 35
<211> 1417
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Inyte ID No: 540661CB1

<400> 35
tgctgcacgac gacagccgag gtagagtggt cagccgagtc actactgect gcttgcctgc 60
ctgtccacgtt gactatggcc cccacaatgg gctggcgcag ggcaggaggg caatgggttc 120
ccccaccaca gactaagggtt gcaactagggt aggggcgcag tcatg'gaag agggagacc 180
tctatagacag tggagtgtgc tggccccact aaggaaacca cctcaccctc tccaacttc 240
tgcctgaaaa tgggcccctgg agctgcgaga cagggcagga tggtcgaggg aaggcctgag 300
atgtgclctt gcccacccc taccaccaic cctccccttc ggaactaac actgggcaat 360
caacaacca cccatgtctc ctctccagtc tcaagagcag gtagtacc aaccatgggc 420
tgcaggccc tgccccggg gccatggag accctactct ttttgcact gaggccact 480
ggcttgcctc tctcccagc caaggtcacg gactgtgccc tggctgctgt ccaagatgt 540
gccctggaga gccccccac ctctcaggg ccaaccacca cagttccctc accaccggt 600
gtgttagaca agctctcct gtgtgtggt cgggggaagg cctgcagcc tgcagccgc 660
gagatccacg gctgagcac agctgtgctg gcagcgcag ggcg'caatg tttgatgac 720
aaactggcca acctgtctct agcctctctg cggcpcacc cacapccctg gtgcttgggt 780
gcaacacatg gtraccgcta cgaactccc ctgctccaag cagaactggc talgctgggc 840
ctcaccatg ctctggatg tggctctgt g'gatataga tcactggct gaagggctg 900
ggcggagcaa cagcccctc aaaaacagtc ccaagaaag gctacagcct aggragatc 960
taactctgca tctcgggca gctccctcca gactgcaca cggc'lgagg tgalgtccg 1020
gccctgtcca gcatctgtca gtygagaca caggaacctc tggcttgggt gtagtctcac 1080
gccagccctc tgggaccat cagcccctg talgggtca cagctctgc taggacnaag 1140
caagagcaat ctgtgtccc aaccactgca cactctgcca caaccaggaa cactag'lecc 1200
agcttggag agagcaggg taccagyat ctctctccag tgaaggacc tggagcccta 1260
tccagggag ggtctctggc cccactggat ctctggcca tctgacct ggcagtagcc 1320
acctgtatg gactatcct ggcacacct gggagtagy ccaagaagg aaactctgag 1380
aataaagacc cccgtgccc ctaaaaaaa aaaaaa 1417

```

```

<210> 36
<211> 2133
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Inyte ID No: 125279CB1

<400> 36
cgtttggtcc ggttgcactc tctctatagc ccagagggcg agagggccty tggcctgggg 60
gagagagcac ggggtctctc ctggatccca gcagtggac gctgtgccat tgggaacaa 120
aggaatagtc tgcctggaat ccttgcagat ctggggccg gggcccagtc caaccclgg 180
agcaggaaga aagc'caagt bgtcaagacc caatgcagc tgcctcagag ccggcccga 240
gtagctcagc actc'ccccc cgaagtgtgc gccttctct gggccagagc gactctgtt 300
tgtctctggg taaagagatt tglcccagct ataccatggg ccgacctcgg gaagctgct 360
ggctggccgc tggatgtggt atcggggcty gtgcctgcta ctgtgtatac agactgctt 420
ggggaagaga cagaaacgag aaatctcggg acgaagcga gtagtctacg gacacatcag 480
agat'ggggt t'gagactgty aaaggaacta aaactaacg tggg'caggg tctggggcca 540
aaactcaggg tgattcagag gtcacgcty aggtgagttt gggactcag gattgtcgg 600
gctaaagp gaggcccaat kaaggatccc aagc'ggggy tggcctagay gccaa'gcca 660
aggcctttt caac'ccctg aagaaacag caag'gcaaa ggcagcaaa gggc'baggy 720
tgggtacocat ctctggganc agdacccttg caccgagttt accctgcca gaggcaagg 780
gtggaggctg caa'ccccac agagatggat ctaggcccg gggcagagca agtggaaat 840
caaa'ggaaa gccccgaagt aagagacaca gggctccagc tacaacatgy cctgtcc'gga 900
gagycaggt caacttctc tataaaatg atgatatct gactgctccc gacctccaa 960
aggt'cctca cactctggag cga'caaatg atccttttat tcaagaa'gta gactt'ggtca 1020
ctctgggtaa caatgcagca tatteattta accagatgc catabc'tgaa tgggtgggt 1080
tccaa'tat tgc'aaactg ataaaa'ca aagccccat aat'laggaa aaga'cttaca 1140
atgoc'taa taact'guyt g'gaacgag aaatccggg caagat'aaq agt'acata 1200
gtcaagtgtg tgatgacaac atggtctgtc gcttggactc agctat'gag atg'ctggc 1260
taagactgtt aaccacatg actgtgacta atcatt'aca acattt'gctt tctatt'ctt 1320

```

WO 01/64896

PCT/US01/06806

```

tfoagaact ttttggttg ttatccctg gaaatcaact caccaagata cagnttatga 1380
aacctattat aaactttact gaaantccag ccatgac-aag agggctgctc agttgttaag 1440
taccatacaga attqatttcc ctctttaata aaqaatgggn tagagagact cctoftaata 1500
tccctaccct atttgagaat ataatgaca acataaasaa tgaagggctc geatcaatca 1560
ggaagaantt cagcagaagt tcactttttt tcttatcaa agagcttgya gtttgggta 1620
agaaatcaa agcaatgaca aatcccaatg atctgggtgt gaagtaaaa gtcctgaaag 1680
tattaaocaa acclcaattt ggagctctgc ccaacacaat ttgagatatt tgcagtggc 1740
acgatgtgat tigttaattc ttgttttcc attg-gcgta tatggttaag agacttttc 1800
agctgtatt ttgaaatc gacatataa tatcaataa gggctgtag ttggttgha 1860
tggttgggtt taggatgaac cakttaagc atgcaaatg aatatbagt attgtacc 1920
agaaagantt tattgattg atctttatc ctgattgag attttttat ctttctcta 1980
cctaactga caatgaatg gttatacat atgaataagc taactttta tattgttta 2040
tattgttat tctaagact gtgttccat aataaagttg tgttttaagc agcagaaaa 2100
aaaaaaaaa gggcggccgc tgcgctta gaa 2133

```

```

<210> 37
<211> 1629
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1299305CB1

```

```

<400> 37
ggggccggct gtgcagagc gccatgtacc gctctatgct agcagtgaat gcccgggctg 60
ccgccccggg gggcttggcc taaagtgcgc gacgaacggg ggtccatcag cccgcccggc 120
tgccgcctcc cgaacctggg gctgtgcttc tggccgagcc gccacaggag cagtcctcg 180
cccctggctc ccgcagacc cgggcccgc cctgtctcag gtgcttccgc agagcctag 240
agagcagccg agactgctg cacagatcc aggatgagtt gggcgcacc gcaatagtg 300
ttggagtttc tctagatgga aaagaagtct ggtcagaagc tttaggttat gctgatgtt 360
agaaactgtt accatgtaaa ccagagaccg tctagcgaat tctagcctc agcaaatgc 420
taccactggt tctcttgc aaattgttg agaggggaa actgacttt gatattccag 480
tacaacatta tcttcccga tcccagaaa aagaatttga aggtgaagc gtttctgta 540
caacaagatt actgatttc cattaagtg gaattcgtca ttatgaaaag gacataaaa 600
aggtgaaaga agagaagct tataagcct tgaagttagt gaagagaat gttgacttg 660
agcaagaaaa agaaagcaaa agtaatgaaa agaattgatt tactaaattt acaacagag 720
agggaatga agcaaatgc ccgaatlcga aacctggcaa gaaaaaat gatttgaac 780
aagggaatt ctatttggga gaaagtttg aaattccat tgaatccca agattatta 840
aaatgatic ttgttctc aaactgtgta gtcagtttct gattccact ttgtgtata 900
cctactggc agcaatgta gaggagctt cagatgta alatttggac tatatgaga 960
aaatalcga tgncttggat atgtgagca ctg-gcayga agaaacagc ccagtatit 1020
acaaagagc aagatlttat gtttcaata aaagaagc tctgtcaac aacacttac 1080
tggataactc ctataaatg gctggtggtg gattctgtc tccgtgggt gaccttctg 1140
aatttgggaa tgcactgctt tatggttacc aagttyggct gtttaagca ccaatgaaa 1200
atcttttacc tggatacctc aaaccagaaa caatggttat gatgtgacc ccagtccca 1260
acacagagat gctctgggtt aaagaggta aatctgcaal ggcctggggt gtttgggaa 1320
ggaaacaaac gtatggttcg ttagaagc aacggcatca tgcctcaact actgagggg 1380
cagtggtgac cagtagtgc ctgctgctc tctctganda actggaatac ggcactaaa 1440
atcaaacagt tcccacaaga ggaatcattt ttctctatc atgtaacatg caatcgtgt 1500
gctcaatag cacctgtttg aaatlgccc ttgaattga taagacaga tcaagctgat 1560
aaccttaaca ccataggtgc aatagagtt gtctgaggt tttttgaaa catfaagtt 1620
ccaaaacatg acattttta gaataaattt gaataagat ataattgna tcaagaaat 1680
atgtactct aattgctaa ttttgaatg gcttttatt gtayaaltg tctttatac 1740
tcagggaagt aattatatt ttttacttt ttgaaaaag tlttaactct tgaataaaa 1800
tattctgata aataaaaaa aaaaaaaa

```

```

<210> 38
<211> 2323
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1385190CB1

```

<400> 39

WO 01/64896

PCT/US01/06806

```

gaagatgag ctoalgaacc ttatctccag atttggtggga attcatgagc ttttctcttt 1050
caatttctat ccccttlttg aaaggtttct gaagcccacc caaagagaag taaccaagat 1140
ccltctgttt gctgcaaacg catctcatca cctagtcccc ccagagatta ttcaactatt 1200
gcttbaagat gtggcaaacg atttgtttac cgcacaagac ctgggagaag tcatgacagt 1260
aggaatcaat gctataaagg agataaacagc tctgagtctt ctggccatga ctgaagaact 1320
tctccaaagc ctggtctcagc ataaaaacac caaggataag aatgcaatga tctctgctag 1380
aaccttgaat caccctctcc gaacactgaa tctctcagatg ctgcagaaga aattccgggg 1440
taagcttaca gaggcttcaa tagaagcag agtaaacaga taaggagat tagatgcaa 1500
agatcaactt ccagggycag aagttctgga agttgagaaa gaggagagc ctgaaaaaga 1560
tgaagatgga tgggaagta ccagctcag tggaggagg gatgtgatg tgaatggat 1620
tgbatgcaa cctctctcag atgaagaca gaagaaaac tccaaagagc tgaacagat 1680
gcccataag ggcgggaag ccaaaagctc agccatcagc actagccagc ttttaactca 1740
ggaagcttc cagaanaatc gcatggccc aatgagaaa gaacttgat ctgcccvcg 1800
gaaatgccc aaggggaaat acattgaaat agcaatgat gaagagaca ggggtgaatt 1860
acttctctt cgggacattg aagccttca taaaaagc aagcttgaca aagagacaag 1920
actagcaat gcaalgctg gaagacaga ccgaaagaa tttgtgaga agaaaccaa 1980
aacaaacca lllccagtt cgcacaataa agagaagaaa aaacagaaga actttatgat 2040
gatgctgat agcagactg tccgttcaa aataaagct tcttccag aaanaactt 2100
gpcactcga gatgacttt tgaaaaagag aaaaagatg aagtaactc ctggcaagt 2160
ttccatctc agaagaatc taagtgtgt tctctgctt gaaatgtgt aactcaagca 2220
tgbttgttta cattaaaag tccagacaca ctgtatgtg naactcctg accctgtgc 2280
agcaalltg tgbtttatt ttggagcgg ctatgtgtag gaatgttat gtaaatagt 2340
gtgttaatg aaaaactll calltatcat tcatgcaaaa aaaaatagt attgagccc 2400
tattcaatgt cactgtagt gcaaaagaa tagagtga cccttcaat caaggaatg 2460
acagcttagc tgtggggg gacacacata tgtgtatla cagtccagt caaatagtc 2520
tgbatgagc gtaaaccca lattctgtg aggtcctga ggccttatg actaacctc 2580
tggagtagc agttatata ttttataaga caaaacaaa agggcaatg ttaacagat 2640
aagagcttc tactgtaca taggettcc lgcgaaaac aggcctctg tgcacagat 2700
ttggtagcat aatttagctc ttttagtcaa tccaaagat taaagtgac cccccccc 2760
cgggttttt ttgtttttg tttgttttg aatgcaatg aaagctttt tggttaaagc 2820
ctcaactta aaacggcct aagataaak agtaccttg gaatacatt agttcact 2880
ttgagctgc tcaactcgg ttctctcag ctctctcag ctgttaatc lttcaagcva 2940
ctgtttcaat tgnrcaata aaagtttct aatgcaaat aaaaaaaa aa 2992

```

```

<210> 40
<211> 850
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 254567C31

```

```

<400> 40
gcaacaagc gtgggggggt cctgagtcgc gacccctgct cggactgac ctgaattgag 60
accacaacc ggaactgctcc ctggaaccga accctaccc ccgccacca gtatggccc 120
gcccgtctc cfaacggggc ccccaggagt tggaaaaca acattgacc ataaagccag 180
tgggtttta aactctctg gttgtgctgt tgaatgatt tahaccgaag aagttagca 240
gcaagggaga agaataggat tcatgctcgt cactgtctc gcccccggg gccctttatc 300
gaagttggg tttagacct caactgaaa acgtgaatc cagttgggc agtatgtgt 360
cgacctgact tcttttgagc agttggcact accgtcttg agaatggc actcagcag 420
tggcccagg caaagagctg gctcatcga tagatggg aagatggac tcttcagtc 480
gctttcatt caagctgttc gtcagagct gtcaccca ggcctataa tcttggcar 540
aatccccgt cctaaaggaa agccactggc tctttagca gaaatcaga acagaagga 600
tptcaaggt tttaatgca ccaggyaaa cagaaccac ctctccag atatgtgac 660
ctactcag agcaagca agtgaagca cgtcaatcc tccctccgt gaagagtc 720
ccagttcaag aggaacctg tggcccctg ctgtcaggg ctgtatgct atggggtal 780
ggaacattg ggcctttct agaaaauof caacagctgt tcccaaaa atgtttaaa 840
gatcaaaaa 850

```

```

<210> 41
<211> 1473
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature

```

WO 01/64896

PCT/US01/06806

<223> Incyte ID No: 3903489CB1

```

<400> 41
cgcggcaagg ctaaggagat gtggcgggcg gclccggggg cccacatgcc tgggtatgag 60
cgcgtgtgca tggycocccg ggcgagcggg aagagcaact actgfgccac catggtccag 120
cactgtgaaq ccccaaacng gtcgtgccaa gttgtaaac ccgabcocag agcagaacac 180
ttcaactact ccgtgatggc tcaactccgg gaactgatcg agyggatga tgcactggag 240
gatgatctc tgcgatccgg tcccaacgga ggaattgcat ttgcaatgga gtaacttggc 300
aataattttg actgctggga gaactgtctt ggcactgtag aggagcacta talctctttt 360
gatgtccag gtcagatgga gttgtacacl cacctgcctg tcatgaaaca gctgggtccag 420
cagctccagc agtggggggt ccgagtcctg ggaatttttc ttgttgatc tcaagttcatg 480
gtgggtccat tcaagtttat ttctggcctc ttggcagccc tgagtccat gatctcttca 540
gaaactccgc aagtcaacat catgacaaaa alggatcgc tgagtaaaa agcaaaaaag 600
gaaactgaga aattttttaga tccagacatg taatttttat tggagatto tcaaatgac 660
ttagagaga aaaaattcca gaaactgact aaagctatat gggactgat tcatgactac 720
agcaagphtc gacttttenc ttncpatcag tcaatgaaq aaagctgaa catgtattg 780
cagcatatg attttgcct tcaatgga gaagacctag aattaaaga accaaaggaa 840
cgtgaagatg agtcttctc tatgtttgac gaatatctc aagatgcca ggaatgatga 900
agagtttact caaagttaac atcaaaagag ctgtggcca aaccagcaga aacttctct 960
cttcaagga tcaatagta gaaactact tattttaa gaaaaagta aacttctct 1020
cttatcagc ctcatgctg aatcaaatb tbaattatc tgaactgct gctgtttaa 1080
gtgaaactct ttatattat aacagcaca cttagattt tgaagtcca aattgaaatg 1140
aaagcaca gatttatata taacttaga cctgagctaa ggttaaggt ggtcaaac 1200
tatttctact tttgtatta ttctgagat gaggaaata ctgtaacaa atgtattg 1260
cggagggaa aaagctgcaa ggtatata abgocactg ctatctgta tcttccall 1320
ttctatatt gaaatgata attattata laacttaaa agtaaaaaa actatcttt 1380
gagatgata tgttatata taaagaaa aaaggtctt aatgatctt ggaactgat 1440
aataatata cahttctca tattgaaat gt 1472

```

<210> 42

<211> 1561

<212> DNA

<213> Homo sapiens

<220>

<221> Misc_feature

<223> Incyte ID No: 5391816CB1

```

<400> 42
cgcggcaagg cgggcttggc tctcccagga ggtacgact gtagccactg gtcggcagg 60
atcccccgct cctcgttgcg cgggtccacg tccctctcgc gtcggccgc ggcgcacgc 120
cgcctctctc ggtttggcc cccggcggtt ccggcgagc ctccggccg ggcacacgc 180
cgacttaaat ctctgcaatg gctaacgaag cttaacctg tccyctgac atggccaca 240
gacttgata tggagggcta ggcctgaaq ttcaagtca gcaactcaag gcttatgta 300
ccaaatcccc cgttgatgca ggcacagctg tgattgctat caagatata ttggctggc 360
agttgcccc aaccagatat atagctgana tgatcnaqy aaalgatc acaaccattg 420
ttccagactt ctctgtagg caagagcctt gggaccctc tggcactg tctatctcc 480
ctgactggct gaaacaaaga aatgcccaga gcatcgatag agagatcct gctatcttga 540
agtatctgaa acaaacgtgt catgcccaga aattggcat cgtgggattc tctgggggtg 600
gaactgctg ccatcatttg atgatgaaat actcagaatt cagggcagg gttgctgct 660
atggatctg caagattct gaagacattt acaattaaa gaaccact ttgttcaatt 720
thgctgaaa tgatgtgtg atccactca aggaagtatc ttgtgact cagaagtga 780
aagaacactg caaagtga aatcaatta aaacatttc tggcagact catgggttg 840
tgcctcgaa gagagaagat tgcctaccg cagacaagc ctacattgac gaggccaga 900
gaaatttaat tggatggcty acaagtaca tgtagcaaga atcaaggtca agctctca 960
gaatagctt caccocaaa ttgcttga aatgttga taattgat taatttcc 1020
tttatanaa tzagttag antctaaa ttgltattl catgaaac acaaatcag 1080
taggctgaa tgcagaaat aattlaatt ttgcaatga cctcgaatc taattana 1140
acaagctg accagtgta gtcctcatg ccgtaatto cagcaactg ggggccaaa 1200
gtgggtgat cactgaggt cnggagttg agaccagct gcccacatg gtagacccc 1260
atctctaca aaaaacaaa aattagcctg gctggtgtt gcaacctgl agtccagct 1320
acttgggag ctgaggcaca agaatcaata gaaccagga ggtggagct gaagtggcc 1380
agatgtg cactactgta ctctagctg ggcagcagay tgaacccctg tctcaaaaat 1440
aataagtaa aataaata aataaaa aaacccgct caglnctgg aatattga 1500
tgalagaa tgatgagta atacctaat tallagtaa aactcaggg ccgggtg 1560
t 1561

```

WO 01/64896

PCT/US01/06806

<210> 43
<211> 1949
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Ineyle ID No: 5886989CB1

<400> 43
gcccggccg cggctgctac gcccggctgc caccgcccgc gggcaccnaa ttccgggccc 60
tagtggggcg taaggccctc ttgtgaaagc ctgagttacg atgtattgag cggctcgtat 120
ggcggcagca ctaaggctct tctggcactc ctctgggtgga cgcgccccat cggccacact 180
tgcctgctc tccagtgatt ctgtagctcc tggctccgta gtctcgcgga cagctctctt 240
tccctggaag tatttctggc gggcctgaca ggggcacty cctcaggcaa gagctcagtg 300
atccaggct tcagcagct gggctytcg gtaattgac tggacgtgat ggcggcga 360
gtcggcagc caggatacc tgncccccg cgcacgtag aggtcttgg actgaggtc 420
ttcctgaga acggcagact aaatcgcaag gtccggggg accctgatct taaccagct 480
gcccggggcg agctgctcaa cgcctaccc cccccaga ttcccaagga galyalgaag 540
gagcagttca agtaactct cgggggatac cgtcaagtga ttctggatat cccctccty 600
tttgagacca agaaqttgct caagtaactg aagcacccg tggtagtata ctgcgacgg 660
gacacacag tggcaccgct gatggccgg aacagcctga accgcaagga cgcagagccc 720
cgaatcaatg cccgctgccc cctgacagac aagcccgga tggcccgca tctcctagac 780
aacctggggc agggcaatct caccaaaccg aggtccacc tcttgcacc ttagctggag 840
cgtctctgg agtaactccc gctggaggtt ggggtctca cgggtctgc tgccttggc 900
agcctctctc acctgctcac ccaactcct ctgcctaacg cctagtgggg cactcaagc 960
agggagcccc aggcctccat ctatctctt ggggctgaa gccaggtaac acatccgtt 1020
tctctcagc cccctccaca cacacacaca cacacacaca cacacacaca cacacacaca 1080
aggactctc gaatcagggc tggcctcty tgcaggcca cccctctt ggagtgatc 1140
ctatctagg tcaagaaata gontgctgto ctcccact tccctgagca gccactaac 1200
aacgcaacc gtagtggtgt g'gaaagcaa aagtcagaa ctgcctaca tctctgtgg 1260
cggggtagg gaatactgg cctgtggccc tctcagggc tcttgcagca aggttcaat 1320
gggcttgaat cctctcaag caactcccc attatgctt aaactctc agggaccac 1380
gggaaatcat gttgacaaa ggtttccat gttcaaacag tcaaccaagc cctccctag 1440
gttcagatac ttctgatact gaagggccag gcccccagc atggtctgtt gccagcaagc 1500
agcaaggalt ctctctggg taacgtctga gccagcagc accagctct tctgggagac 1560
cccctcccc cgtlagggtt tgtgtaglyc lcccaatcc tcttattgc ctgccacco 1620
tgcctctgc ctggaactct tggatctcc gttacccat ctttctgtt tcttgcctc 1680
gttccctcaa cctgctctc ctggctggg ataccctta tttcttcta tttctcttc 1740
ctttttttt ttttttttt ttltttaac taacatagc tctcagtca ggcctgtct 1800
aaagcccag gctggttag gttgcttag ataccactc cccctcaaa ttttaactt 1860
atccaaact acagtaata gccaaactt gctggtttt tctactaga ctgtaaacc 1920
tctagggaca gggcaaggt ctattctat 1949

<210> 44
<211> 2221
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Ineyle ID No: 723432CB1

<400> 44
ccgggaccgc tggtaaggaa cggggccgc gcttctgct ggttaacagg ccttgcctag 60
tgggcttcc ttccaggtc gccctcagt ctccactga gacaggaag accagtlyc 120
cttctctccn agaaccttg agatctcgg tctggggtct ggttgaaga tggcggcct 180
cactaccctg ttaagtaaa tagatgaaa tcaaggatcc taactaaga aactgcgaa 240
atgggtggct atccagagtg tctctcctg gccggagaag agggcgcaaa caaggagat 300
gatggaggtt cctyctcag atgttaagca gttgggggc tcttggaac tggtygat 360
cggaaacaaa aagctcctg atgctcga gatcccgct cctctatc tcttggcag 420
gctgggtccc gaccacaga agaaagact gtgcattac gggcaactgg atgtgagcc 480
tgcagcctg gaggacgct gggacagca gccctcacc ctggtggagc gagaaggca 540
gttgnlgyy agagllcga ctgatataa gggcccgct gcccgtgaa taaccgccc 600
ggagcctat cggaaaccg gccggagct tctgtaac gtcgtattt gctcgaag 660
cttgaggag tcaagctct gggcctaga cagctgatt ttgcocgga aagccactt 720
ctttaagaa gtggactacg tctgcattc tgcacttac tggctggaa agaaagacc 780

WO 01/64896

PCT/US01/06806

```

ctgcatcaac taaggctctc ggggcatttg ctacttttc atcgaggtgg agtccagcaa 840
caaaagacct ccttctgggg tttacggggg ctcggtgcot gaggccatga ctgatctcat 900
tttctgatgc ggtcttttgg tggacaagag ggggaaatcc ctgactcccg gcattaaoca 960
ggccctggcc gccctccagg aagaggagca caagctgtcc gacgacatcg actttgacat 1020
agagaggttt gcccaagatg tggggggcca gatccctcct cacagccaca agnaagacat 1080
ctctatgcac cgatggcgtt acccgtctct gtccctccat ggcacccaag gccctctctc 1140
tggctctggg gcccaagacc tgatccccc gaaggtggtt ggcaagttct ccaltcaggct 1200
cgtccagcaa atgactccct agtctgtcgg cagcaggttc acaagctacc taactaaaga 1260
gtttctgtaa ctctcgngcc ccnaagatt caaggtctac atgggccaag gggggaagcc 1320
ctgggtctcc gactctagtc accctcatta cctgctgggg agaagagcca taadacagat 1380
ttctgggtct gagccagact tgaccagga agggggagat attccctgta ctttgacctt 1440
tcaggagctc accggcaaga agtctatgct gctgcccctg ggtccagcgg atgaaggggc 1500
ccactcccaa aatgaaaagc tcaacaggta laactacata gagggaacca agatgtcggc 1560
cggctacctg tatgaggtct cccagctgaa ggaactaggcc aagccctctg tgtgccatct 1620
ccaattgaaa ggaatccctg cctcacctca cccttttcca acttgcccag ggaagtggag 1680
gttcccltll tccctccctc cttgtcaggt cabcactgac tttaggaaac agcccaagct 1740
gcatccagct gtcacaggtt gggagctacc gttggcttca tgggtgacct ggaagcaag 1800
ctgagctccc ctgggttaagt tctccagagt gtccagatgc ctgacactgc agaagatacc 1860
caagctccaa aagcacaagc tctcgggaaa gttccggttg tggctgggc acccaggtctc 1920
accctataa tcgggacctt lgggaggcca agacaggagg atccacttgc gccaggagtc 1980
tgagccagc ctaggcaaca aaacaagact ctgtctctac aaaaagtta agaatgagc 2040
cagaatggt ggtgtatgcc tgtagtctca gccactcaga aggtgagcc aggagatgc 2100
cttggagcca aagattttag cctcggglga gctgtgaaat cacacggcac tcaagcctgg 2160
gcaattgacc aagatccctg ctctacaaga agatttttta aagatgagcc aagtttgggt 2220
g

```

```

<210> 45
<211> 1619
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> IncyTe ID No. 1332963CB1

```

```

<400> 45
tgcagctggc gaaqtbgggc gactggcgga tgcaggcctt gccgcacgtc gttgtcggcc 60
gtcccgccgg cgtgacacgc gccctggccg cgtctctgct gaggccgaaa ggttaccagg 120
tgacaggggt gtttatgaaq aactgggact cactggatga acatggggtc tbtactgcgg 180
accagagctc tyaaslgct tacagagttt gccagatctt agaatccctc tccatcaag 240
tgtccctcgt aaggaglal tggactgagc tglccgtyga ctttttgaal gatttcaaaa 300
aaggcaagca tcccactcct gcaatagttt gccncaaga cctcaaatct agttgtctt 360
ttcattatgc tgtggalaac ctggggcgag algccactgc caccgylcar tatccagaa 420
cttccctgga agatguagaa gtcctttagc agagccctgt taagaagccc gaagggcltt 480
tcagaaatcg gtttgaagtt agaaatccgg taacctctct ccaggccagct gacagcttta 540
aegaccagac cttctttctc agccagcttt ccaggaatgc cctgaggaga accatcttcc 600
ctctgggggg attaacgaaa gactttgtaa agaaatctgc tgcagaaat agacttcttc 660
atgtcttca gangaagag agcctgggca tghgttcaat cgggagagay aabttgaa 720
attctctctc tcaqtatctg cagcctagac ctggtcaatt hcttccata gaacaaata 780
aggtctcggg aacaataaa ggttggttcc tbtatacctt gggccagaga gcaaacatag 840
gtggcctgag agagccctgg tacgtgttgg agaaggaact cgtcaaggtt gacgtgtttg 900
tggccccccg gaacagacc cagcctctgt acagggaact gctgaggacc agccggctgc 960
actgualtgc ggaagagccl ccgcagccc lyytccggga caagatgatg gagtgcact 1020
tcagatcccg ccaccagatg ccactagtc cctgtgtgct gacctcaat caagatgcca 1080
ccgtctgggt gacagctgtg caggctgtgc gtgcccttgc cacaggacag tttgtgtgt 1140
tclacaaggt gacagctgtc ctgggagcc ggaagatctc ccgctgggg ccgtctgct 1200
acacgctcca gaaggccag ccaagagcty ggalggccc tgaagcccc agtgaagcc 1260
cagagatggt tcaagcctg agtcccttgc tctgacagag atggatctgc tagaagyaac 1320
ctgagagacc ggaaccatgg ctggcggctt ggtgagcagt ccaggtgccc aagggccagc 1380
ttctctctgc ccaagcaga ggaagcgggy ctggctgagg gtcgaaaag cctgcagggg 1440
ccgggcccgc ccaagaaaga gctttagctc caggctgggg ctctgctgc tggagcactc 1500
gctggctggt ggggtggccc gagtctccct tcaacgccc cagggaggtt tccccctc 1560
agattacacc gaggggacct gcagaggggg ctgtcgggac agcgtgggaf aaacataag 1619

```

```

<210> 46
<211> 1448
<212> DNA

```

WO 01/64896

PCT/US01/06806

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1559410CBL

```

<<400> 46
gctgcccaga agtcagagtha gtttagllag tateggcctg ttatctctctt ttggcggaca 60
cggctctcagc ttttccgctt gggcggactg acgtgggccc ccaacagagc akacgtactg 120
ggaccctcgc cctcagctct gtctccggcg cggctcaact ccccgtttc cctgtgattt 180
aacctcctcc gggccggcgg ccgccaatgg cagggcggcg tccgaccacg goottcggcg 240
aggccggtgt cggcccgccg ggtctaggga agaccacgla ctgctgggc atgagtgaat 300
tctcggcggc gctggggcgg cggctggcgg tggtagacct ggaaccggcc aacgagggcg 360
tgcctgacga gttgcccgtg gacgtggcgg agtgggtggg gctcggggac gtagtggacg 420
cgttgcctcl gggcccacac gggggcctgc tctactgcac ggagtaacct gaagccaacc 480
tggactgagt ggttccacag ctggaccccc tccggggcca ctacilcctt ttcgactgac 540
cagggccagc gtagctctgc atgcatcagc gggcttggcg cagcatcttc tcccacagg 600
cgcagtgaga cctcaggctg actgcctgac acctcgtgga ttctcactac tgcacagacc 660
ctgcacagct catttcagta ctgtgtacct cctggcccac catgctccac gtggaactgc 720
ccacatcaaa cctccttccc aagatggacc tcaltggaga ttatgggagc ctggccttca 780
acctggacta ctacacagag gttctggacc tctcctacct gcttgaccac ctggctctg 840
acctcttctt ccggcactac cgcagctca atgagaagct agtgcagctc atcgaagact 900
aragcctgtg ctcttttato cctctcaaca tccaggacaa ggagagctc cagcgaactc 960
tgcagctgct ggttaagacc aatggactct gttctgggag ccaagagcag cgaagcttgg 1020
aagccatgat gttctcgcga atggagcgc actctcatit ctctccaca ctgggactcc 1080
aggccagatc cctggcacc ccaaacagct cagtggagca ggaaccctcg cagctgtgac 1140
aaccaagctg acctcgaga gcaggtatgca tntctcagca ctggggaagc tggaggtacc 1200
tgcctcaggg ctccagacca agaacagctc ctcccagcca gactggcggc gctggcaagg 1260
ggatattcag ctctcgaag gactcttggc caaaaagcca gacatgtg ccaagcagac 1320
accctccala ctgtcagttg tgtctgtgag ctctggggcc tgcacccaga aagttagaca 1380
ctgtctctag tcaagctgtg atgaattgt ctacaaatca agagtttatt ttctaaaaaa 1440
aaaaaaa

```

<210> 47

<211> 2225

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1752587CBL

```

<<400> 47
agagagagag agagagaact agtctcagac cactcatgga gcaatgtggc tctgtgtgt 60
ggcttggcg gtcttggcat gggcttctct ctgggtttgg gactctcag aacgaatgaa 120
gagctggagc cagggaaggac gctggggagc cgaagccgg acctgctgg tcatagcga 180
ccctcagcat gaagcctagt ttttctctcc caccgtgcta gcttggccc gctcaaggca 240
ctggctgac ctgcttgcct tctctggcgg aactctctac aatcaggag agactcghaa 300
gaaggactt ttgcagagct gtagtcttll gggatccca ctctccagc taatgatat 360
tgcacaaagg gatttcccag atgaccagc catgcctggc gacaccagc aagtgccag 420
aytccctccl cagcacatag aagtgaatgg cctcaactcg gtagtgaact tcatgacag 480
gggaglaagl ggcacacgca atcactctgc tctgtatgca gctgtgagg cctgcactc 540
agaagggaag ttaactaaag gtaaggcttg ttcttctg caaaggccac aagatactgt 600
cctctcagga aacttatgag ggggaattt gcaaatccac agagagctgc cctcagccga 660
gcagagctag tggttgggga ggcggccaca ctctctcaga ctccactcaa ggtacctggc 720
agttgtatc taacagcag tgcctgcctg cagggccacc tctggatgag tgggagccc 780
aggtttggcc ctcaacctgt ctacttttt ccaaggtgag atgttttaca cctctgtat 840
tctctcagtc tgaactcaat cccctcctcc aaccaattag gcaggttgra aatgaactgt 900
ggctcctlga tgaactcaat tgcctcagct aggaactctt cctaacctcc ataaagtatt 960
tgcacctacc tgcacatlla ctgcalgacc cccatgctta gctctctgg accagtctt 1020
ggtcagctcat acctcaccgg cgcacataga agatcaatg aatgaaaaag taagtattt 1080
gtgccttctc tatgtgggat gactgaaagg agcttaccaa gccatgatg ctttggaaa 1140
tgcactatag ttttctgact ctctatctaa tctctgtgaga cagggcagctc tcaactatc 1200
accctcaatt taccagatag aaactcagaa ctggagctgg ttaagggact ttgccaagtc 1260
tctcaattct tgaactagc agctgggacc aggtctclaa actccacata cgttaattc 1320
tctctctctt actctctctg gcttggaggt agccagatct caactctag cagcccaact 1380
ttttctggtc gggccaccgg caggtngagg gtgaccgcc aggtcttcta tggccctttt 1440

```

WO 01/64896

PCT/US01/06806

```

actcgttcta gctgcaagaa tcatgggca agggcagcgt tagtgccaca gaaacacgtg 1500
gcaatgctgc ggcactgtc ctgccttgc ctgagccaca ggglagaggg aggatgctca 1560
ggggaaate tggctggagg cctcaaggct tateacttca cctgtctcc tctcaatca 1620
gggtgctctg tgcacagct tcaagtctgt aatgtctctc gcaagtacat cctcctctctg 1680
gatctgcctc tctctctgct tcatagccag galgctctct tctgtctcaa cagcaaaagaa 1740
gtggcaaggg caagaagag cagtctctgc caccgagcc agctcctctg gttccgccc 1800
ctctacatta tctctcccg gtacalggg atcaactcac tgagctcct ctgaagcctt 1860
gaaggtttt caggtccag gaacaaagg gaaactagc aaagagatc agggacttg 1920
gctgtgcact ggttattta cctgagctca aggagatccc cctggagga gctctgcaa 1980
aaagggccc atgtaggcca ggggctgtcc aaactccagc ttcttccctc gggaaaaaac 2040
ccaagaaac aaaaaaac caccocaaag atantcatag ctacactcgt agctctcraa 2100
gtctctgtga aaaaacttt acataatgac acagtatgat tggaaacact agccagctgc 2160
ctggcagcgt cctattatc ataaatgaa ataanagtc tctaaaaaca ctcaaaaaca 2220
aaaaa

```

<210> 48

<211> 1400

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1911509CB1

<400> 48

```

gccccagccc gggcggccc agtggcctg tcagcggctc tcagataag tcccggcct 60
tggcggcgg cccclatggc ggcggagag gaggagtg acctgcgca cccggagag 120
aggtcaggat ggttaactgg ttggctccc acatggtgc ctactlclal atccacott 180
aaagagcgt aaggaagat gtaaaatgt gtgccttga catcaaaaa agaaccctgt 240
cgtalalcla atgaaataa aatattgaca ctgaaatct clcataatat ttcaataag 300
actcaactgt tcttcccca tggttttgga gaggctctg gctctgggc actgaatbt 360
ggagatctt gcaacacag accctctat gctttgac tattygttt tggcagaat 420
atlgaccca gtttgacag tgatccagaa gaagtggaga atcagttct gaaatcatt 480
gaagagtga gatgtacct aggattggc aaaaatgat tctctgggca caacctaggt 540
ggatctctg ctgctgctta ctgctgaag taacctcaa ggttaatac tccacttta 600
gtgagcctt ggggtlccc tyaacgaca gacttgcct atcaagacag accaattcca 660
gtttggatca gagccttggg agcagcattg actccctta accctllagc tggcctaagg 720
allgcagac cctttgctt aagtctagtg cagcyltlaa ggcctgattt caaacgaag 780
tattctcaa tcttcaaga cgtactctg acagatcaa ctaccactg taatgpcag 840
actcaagtt gtaacagc ttcaagat atgacttc ottatgagt gcaaaaag 900
ccaatgccc agcgaattg taaaatgac cctgacatc cagttcagt gatcttggc 960
gcccgatcct gcaatagatg caattctgac accagatcc agtcttaag accacatta 1020
tatgtgaga aatagatct tcttgggcca gacattatg tatatgaga tcaaccgaa 1080
gaatcaacc aaaaataaa gtagatctgc gacactgag actgaacaca ctgaagctct 1140
gatgggaaa cctggtgact gatagcttg ttacgcaata attcatgct tglgatgag 1200
agfagtgaat acaaacacac accagggagc ctctttgact atacttga catgtttct 1260
ctaggaatc actccacat taaaccagt cagtgcctc tagaagaatg gcttccctt 1320
ctctacaca aattgaaat atcaagctc ctcaatata tccctttaa taaaagtta 1380
ttgtcccto tyaaaaaaa

```

<210> 49

<211> 1361

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2210170CB1

<400> 49

```

ccggcagcgt ttcccagcag gctccacagc agcgttccc gtcatgggga tbtggcagcg 60
ctcgtctctc ttggctggg tctgctctg gctggtggc gggccactg cccgcttgg 120
gggaagcga gcalgcttc gtggcggca gttctctggc gccggagtg agaccctaaa 180
caaaagaga acccaatca tgtcccagg acttccaaag cagaaacca tagaaggtg 240
taaacagtt atagtttgg cttctgaaa ggtggagtc gcaaacata ctacagcgt 300
gaatcttga cttgactag caggaacga ttctccaaag gccattggt tctatagat 360
ggatctgtat ggccttcaag ttccaaagat gatgaatcg aaagaaatc cyyaatatc 420

```

WO 01/64896

PCT/US01/06806

```

acagagcaac otaatgagcc ctctcttggc ttatggatct gollgatgt ctatgypct 486
tcctgtttaa gaagtgaac cagtagtttg gagggcoct atggtaatg cggccattga 540
gaattgttg aggcagtag atggggta acggactac ttatgttag acatgcaacc 600
aggaactgga gatgtgcagt tctcagtcct acagaatatt cctataacag gtgctgtgat 660
tctctccag ccccaggaa tgcgattgat gpatgcacac aaggytctg agatgtttcg 720
cagagtcacc gtgcccgtaa gcttttacc cttcaccttg aaaaatctaa aactctctc 780
ccfgagcaag caagctatg aaaaatacag ttgagtttt aabttagtc tgtgcagaag 840
tccttclgaa agtctagtt tctcattctg gtgacttaa aaccacact ataggtacca 900
acatgtctat acatgatatt aatccatag aatgtttaa cccctctag gtnaggttt 960
ctctatgagc tactacttct ctctactctg cagctatcag tgcacaaaaa aaattacaag 1020
agtgctttag gaattattt agatagatag gohgtaatg atcaactat ggaagcaga 1080
tgaaaaaaag attacttttt tctctctgpc tctacagtt caactcatt atagctaca 1140
ttctctag agacacttca gcatctctc cctaatttg tggaaatgaa gtaaaaaa 1200
agataaaat agacaaaata attaaaatta gatgggtta gaagtgaat cagttaacaa 1260
attgcctgct attagaagac taaatacat gctaggggtg taccgtgact ttctatgag 1320
ccaaagaaa agaaaaatag taagcaatat aaaaactgc c

```

```

<210> 50
<211> 993
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No. 234654CB1

```

```

<400> 50
gaagggccga cggctctggc tgcocggcgg ttgagagcat ggcctotoca ggggcaggtta 60
gggcoccccc gtagtctacc gagcgggaact gggglaccg cgaagtcag tactgggac 120
agcgtaccca aggcagagcc gattctgccc cctacagatg gttcggggac ttctctctc 180
tctgttccct cctagagccg gactctggcc ccagggaccg tctctctgct ctaggttgcg 240
ggaaacgtgc cctagatcac gagctgttcc tggaggtct cctaaatgc accagtggg 300
actactcacc agtctgtggt gctgcactgc aggtctgct tgcacatgct cgtgagctgc 360
gcttggagac cctgtgatgt cggagctgct actccccg tctctctttt gatgtgtgc 420
tcgagcaggg caogctgga gcccgtctgt ctggggaaag agatccctgg accgtctct 480
ctgaaggtgt ccacactgct gaccaggtgt ttagtgaggt gagcccgct ctgttccctg 540
gagcgggllt latctcaatg acttctgctg cccccactc tggacccaga cactatgcc 600
aagcctatta tggctggtcc ctgagccatg ctacctatg cagcggtttc cacttccalc 660
tctactctat gcaaaaaggg gggagctca gtttggccca gctggctctg gggcccaaa 720
tctctcaac ccccagact ccaactcac ttctctctc tccagctca gatcatgag 780
acttcttag tgcattcag ctctgagccc agagcatggt cctcawctc tctgcaatt 840
ctgcccctgg ctctcaggt agttggaatt ctgacttag gacttgggt tgggtccaa 900
gtcttcaact cccagggccc tcatgctcaa gatagaggt gggagcgaac cccatgaac 960
caatcacgcc cagctccaaa tagaaaaaa aaa
993

```

```

<210> 51
<211> 1367
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No. 288414CB1

```

```

<400> 51
ctggaaacct ggcaggtcc gaaaaagcc agatctgaaa gllggtgag gacggcttt 60
aagcagaagc tggagggacc gggagaccg gggagcagtg ggcagtcgt ggtcccggt 120
tatactata gtcocagta tctcagtat tctgactccc tggccaagat ccccaagcg 180
gcaagtatg tgcattcttt gattgaagca tatgactctg ataacagat gaggatggt 240
aagctcaagc tggcctccat gaggagatg gccacctccc acactgatg llatctgca 300
cactccaga aggtcagcca agagggcgal galgactate cggactccat aqaatctgg 360
ctaggttatg actgcccagc cactgaaggy atatttgaat alycagcagc talaggggg 420
gctacagatca cagctgccc aatgctgatt gaggaatgt gcaagtagc aabtaactgg 480
ctggagggll ggcctatgc aaaaaagat gaagcactg gttctgtta tctcaatggt 540
gctgtccagg gaatttaag atggagcgg aaattctgag gattctctc cgtgatttg 600
gctctgcaac atggagatgd ttagaagac gactccagt tcaactcaa agtcatgac 660
gtgtcccgc acaaatcttc ccnagattt ttcccaggaa caggtgaggt gttctgatt 720

```

WO 01/64896

PCT/US01/06806

```

ggcttaggga agggacggta ctacagtgtt aatgtgcca tttaggatgg catacaagat 780
gaaaatatt ccacgactgt tgaagagcta ctaaaggagc tatcccaagc ctttaatecc 840
aaagcagctg tcttacagct gggagctgac acaatagctg gggatccat gtgctcatt 900
aacatgactc cagtgggaat tggcaagctt cLtaagtaac tectcaatg cagatlggca 960
acactcattt tggaggagg aggtataaac ctgtccaca cggctcgatg ctggacatac 1020
ttgacggggg tcatctagg gaaaacacta tectctgaga tccagatca tggatltttc 1080
acagcatatg gtctgatta tgtgtgga acaacgcaa gctgcccgc agaccgcaat 1140
gggcccacc gaatccaaa aatctcaac tacatcaag ggaatctgaa gcalgtggtc 1200
tatttgacag aagagataa gytttccag gctgaggat gctgctata atgaaacag 1260
cgttttatg caagcagttt gtggaattg tgaetccag gaaatbtga aggaattac 1320
ttctgaaaa ttccaaggg gcatcaagt gtagctgct tcttggg 1380

```

```

<210> 52
<211> 1317
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 4103539C81

```

```

<400> 52
actagggccc tagcacagtg totgatggag ctttctacc agaccctgaa attcaccgat 60
cagcgcggcg aagcgtgca gatgaggaca gaagcacyac gaaaaaatct tctcatlltg 120
atttgcatt atttaacaa agaagggat atgatacag caatgcttt gggcaagaa 180
actaaactcg cgttccagcg gtttgagtt tglgcaaca tggatcttga acfatlltg 240
atgaaatag agagtatta tttgtaaaa tttaagaaat acccaaaat tqtcaaaaag 300
tcatacagaa cagaaataa tttaacgaa agatgtagag ggaagacag aagatgatg 360
aacgacagtt gtcaaaatct tccaagatc aatcagcaga ggcocggtc caaaaccaca 420
gggggaaag caggggacac caaalgcctc aataaggagc atcctaatac gtaggtagtt 480
gataacactc gctgggaag tgcacaactc ggcatacata tatcaagaat cgttaagac 540
agtgaagcgg aaaaagccca ccccgaaag ggcacaata tgcactcca aggtctgtc 600
acagatgca tcaaggagc aacagatga ctgacttga acacttga caaaatca 660
gaacctcag aacgactgct gaaactctg agtgcttca ttggctgaa cagttagatg 720
cagcaatag cagcctggt ggcggggac attaatctc ataactcaa cataaagtg 780
aatgacatta ttgacttga tgcagcaag cagttagtca aagaagctg tgtgatact 840
ataagratc cacagctart tacaggaatt ctttctcct ggaaggact actgctgac 900
ggcctccag gtacagaaa gacttactg gccaaagctg tggcoactga atgtaaaaa 960
acnttctta acattctgc atccaccatt gtacgcaaat ggaggggga ttcaaaaaa 1020
ctgtctggg tttatttga gcttgcggc taccaagccc catcaagat cttctggac 1080
ggctggag' cgtgatgag tcaagaggc acagctctg ggggaaaaa tgaagaaag 1140
ctgggatga gaaagagtt actggtcag atggtggyr lggtaagctc agaagatctc 1200
gtatttctc tagcacttc taacctgag tggtaagaga ccaagagat aaattttgaa 1260
tcaatttca ggaatcacta agtcaaaaa aaaaalltal attgactact tcaaaaa 1317

```

【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
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64896 A3

(51) International Patent Classification: C12N 15/52, 900, 57J0, A01K 67/027, C12P 21/00, C07K 16/40, C12Q 1/68, C01N 3/57, C12Q 1/00
Avanue, Mountain View, CA 94043 (US), YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US); AZIMZAI, Yalid [US/US]; 5318 Boulder Canyon Drive, Castro Valley, CA 94532 (US); LAL, Preeti [IN/US]; P.O. Box 5742, Santa Clara, CA 95056 (US); BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US); BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US).

(21) International Application Number: PCT/US01/06806

(22) International Filing Date: 1 March 2001 (01.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/186,307 1 March 2000 (01.03.2000) US
60/192,532 28 March 2000 (28.03.2000) US
60/193,578 30 March 2000 (30.03.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/186,307 (CIP)
Filed on 1 March 2000 (01.03.2000)
US 60,192,532 (CIP)
Filed on 28 March 2000 (28.03.2000)
US 60,193,578 (CIP)
Filed on 30 March 2000 (30.03.2000)

(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): TANG, Y., Tam [US/US]; 4230 Ranwic Court, San Jose, CA 95118 (US); LI, Dyoung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US); BANDMAN, Olga [US/US]; 366 Anna

(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, CU, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TH, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

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

Published:
with international search report

(88) Date of publication of the international search report:
4 April 2002

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



WO 01/64896 A3

(54) Title: HUMAN ENZYME MOLECULES

(57) Abstract: The invention provides human enzyme molecules (HEM) and polynucleotides which identify and encode HEM. 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 HEM.

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

Info: International Application No
PC1/US 01/06806

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12H3/52 C12H9/00 C12H5/10 A01K67/027 C12P21/00
C07K16/40 C12Q1/68 G01N33/573 C12Q1/60

According to International Patent Classification (IPC) or in both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification scheme followed by classification symbols)
IPC 7 C12H

Documents searched other than in the IPC scheme (classification scheme followed by classification symbols)

Electronic data base consulted during the international search (name of database and where electronic search was used)
SEQUENCE SEARCH, EPD-Internal, WPI Data, PAJ, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Classification of document, with the nation, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 55858 A (SCHMITT ARMIN ; SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN BERND) 4 November 1999 (1999-11-04) abstract; claims page 1 -page 5 seq id nos 38, 269	1-19, 22, 25-28
X	WO 99 38972 A (CRKVENJAKOV RADOMIR ; JONES WILLIAM LEE (US); STACHE CRATI BIRGIT ()) 5 August 1999 (1999-08-05)	12
A	abstract; claims; tables 2.20 ---	1-19, 22, 25-28

Further documents are listed in the continuation of this C. Patent family members are listed in annex

Special categories of cited documents

"A" document referring to the general state of the art which is not considered to be of particular relevance

"C" non-relevant document but published on or after the international filing date

"I" document which may throw doubt on priority claim(s) or which is cited to assist in the prosecution of a national phase or other special right(s) specified

"O" document referred to in an oral presentation, written description or other means

"P" document published prior to the international filing date but which is not publicly available

"T" document cited after the international filing date or priority date and not in conflict with the application but cited to inform about the state of the art underlying the invention

X document of particular relevance: the claimed invention cannot be considered to be novel because of the disclosure in this document

Y document of particular relevance: the claimed invention cannot be considered to be novel because of the disclosure in this document

Z document of particular relevance: the claimed invention cannot be considered to be novel because of the disclosure in this document

Date of the actual completion of the international search: 24 September 2001

Date of issuance of the international search report: 03 Jan 2002

Name and mailing address of the ISA:
European Patent Office, P.O. Box 1, 7000 Lausanne 2
CH-1000 Lausanne, Switzerland
Tel: +41 (0) 21 611 9511
Fax: +41 (0) 21 611 9595

Name and address of the authorized officer:
CEDER O.

1

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

INTERNATIONAL SEARCH REPORT

International Application No.
PC/US 01/06806

C/(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Classification of documents, with indication where appropriate, of the relevant passages	Relevant to claim no.
X	<p>DATABASE EM EST [Online] EMBL: 25 March 1999 (1999-03-25) NCI-CGAP: "tq42c04.x1 NCI CGAP Ut1 Homo sapiens cDNA clone IMAGE:2211462 3' similar to TR:023034 023034 SIMILAR TO CAEORRHABDITIS UNKNOWN PROTEIN T03F1.1. ; , mRNA sequence" retrieved from EBI, accession no. AI559155 Database accession no. AI559155 XP002178282 the whole document</p>	12
P,X	<p>DATABASE EM HUM [Online] EMBL: 29 September 2000 (2000-09-29) SUGAHO ET AL.: "Homo sapiens cDNA: FLJ23251 fis, clone COL04590" retrieved from EBI, accession no. AK026904 Database accession no. AK026904 XP002178283 the whole document</p>	1-5,11, 12
T	<p>-& DATABASE SWALL [Online] 1 March 2001 (2001-03-01) BLDECKER ET AL.: "Hypothetical 44.9 kDa protein (cDNA FLJ23251 fis, clone COL04590)" retrieved from EBI, accession no. Q96Z29 Database accession no. Q96Z29 XP002178284 the whole document</p>	

1

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 01/06806
Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)		
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1.	<input checked="" type="checkbox"/> Claims Nos.:	
	because they relate to subject-matter not intended to be searched by this Authority, namely:	
		see FURTHER INFORMATION sheet PCT/ISA/210
2.	<input checked="" type="checkbox"/> Claims Nos.:	20, 21, 23, 24
	because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:	
		see FURTHER INFORMATION sheet PCT/ISA/210
3.	<input type="checkbox"/> Claims Nos.:	
	because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 8.4(a).	
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows:		
see additional sheet		
1.	<input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.	
2.	<input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.	
3.	<input type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:	
4.	<input checked="" type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims, it is covered by claims Nos.:	1-19, 22, 25-28
Remark on Protest		
	<input type="checkbox"/> The additional search fees were accompanied by the applicant's protest.	
	<input type="checkbox"/> No protest accompanied the payment of additional search fees.	

International Application No. PCT/US 01/06806

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 1 and 27, respectively.

2. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 2 and 28, respectively.

3. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 3 and 29, respectively.

4. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 4 and 30, respectively.

5. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 5 and 31, respectively.

6. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 6 and 32, respectively.

7. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 7 and 33, respectively.

8. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide

International Application No. PCT/US 01/06806

FURTHER INFORMATION CONTINUED FROM PCT/ISA 210

encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 8 and 34, respectively.

9. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 9 and 35, respectively.

10. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 10 and 36, respectively.

11. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 11 and 37, respectively.

12. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 12 and 38, respectively.

13. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 13 and 39, respectively.

14. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 14 and 40, respectively.

15. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 15 and 41, respectively.

16. Claims: 1-19 22 25-28

International Application No. PCT/US 01/06806

FURTHER INFORMATION CONTINUED FROM PCT/ISA: 210

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 15 and 42, respectively.

17. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 17 and 43, respectively.

18. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 18 and 44, respectively.

19. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 19 and 45, respectively.

20. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 20 and 46, respectively.

21. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 21 and 47, respectively.

22. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 22 and 48, respectively.

23. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 23 and 49, respectively.

24. Claims: 1-19 22 25-28

International Application No. PCT/US 01/06806

FURTHER INFORMATION CONTINUED FROM PCT/SA/ 210

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 24 and 50, respectively.

25. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 25 and 51, respectively.

26. Claims: 1-19 22 25-28

An isolated polypeptide and an isolated polynucleotide encoding it and their uses, wherein the polynucleotide and polypeptide are Seq Id Nos 26 and 52, respectively.

International Application No. PCT/US 01/06806

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claim 18 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Rule 39.1(iv) PCT - Method for treatment of the human or animal body by therapy

Continuation of Box I.2

Claims Nos.: 20, 21, 23, 24

Claims 20,21,23,24 refer to agonists/antagonists of the polypeptide of claim 1, identified by the method of claim 19 or 22, without giving a true technical characterization. Moreover, no such compounds are specifically defined in the description. It is only indicated that they "may include proteins" (such as antibodies), nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of HEM either by directly interacting with HEM or by acting on components of the biological pathway in which HEM participates." (page 24 lines 13-15; page 25 lines 13-15). In consequence the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such claims whose wording is, in fact, a mere recitation of the results to be achieved.

Claim 8 has only been search in so far as it concerns non-human transgenic organisms.

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

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 01/06806

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9955858 A	04-11-1999	DE 19820190 A1	04-11-1999
		WO 9955858 A2	04-11-1999
		EP 1076700 A2	21-02-2001
WO 9938972 A	05-08-1999	AU 2471699 A	16-08-1999
		EP 1053319 A2	22-11-2000
		WO 9938972 A2	05-08-1999
		AU 2095599 A	19-07-1999
		AU 4187499 A	29-11-1999
		EP 1105474 A2	13-06-2001
		WO 9933982 A2	08-07-1999
		WO 9958675 A2	18-11-1999
		AU 6263999 A	17-04-2000
		EP 1144636 A2	17-10-2001
		WO 0618916 A2	06-04-2008

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 35/00	A 6 1 P 37/02	4 C 0 8 4
A 6 1 P 37/02	A 6 1 P 43/00	4 H 0 4 5
A 6 1 P 43/00	C 0 7 K 16/40	
C 0 7 K 16/40	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 9/00	
C 1 2 N 5/10	C 1 2 Q 1/68	A
C 1 2 N 9/00	C 1 2 Q 1/68	Z
C 1 2 Q 1/68	G 0 1 N 33/15	Z
G 0 1 N 33/15	G 0 1 N 33/48	M
G 0 1 N 33/48	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 33/566	
G 0 1 N 33/566	C 1 2 N 5/00	A
	A 6 1 K 37/48	

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

- (72) 発明者 バンドマン、オルガ
アメリカ合衆国カリフォルニア州 9 4 0 4 3 ・ マウンテンビュー ・ アンナアベニュー 3 6 6
- (72) 発明者 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 ・ サニーベイル ・ ルイスアベニュー 8 2 6
- (72) 発明者 アジムザイ、ヤルダ
アメリカ合衆国カリフォルニア州 9 4 5 5 2 ・ カストロパレー ・ ボールダーキャニオンドライブ 5 5 1 8
- (72) 発明者 ラル、プリーティ
アメリカ合衆国カリフォルニア州 9 5 0 5 6 ・ サンタクララ ・ ピーオーボックス 5 1 4 2
- (72) 発明者 パーフォード、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 ボーゲン、マライア・アール
アメリカ合衆国カリフォルニア州 9 4 5 7 7 ・ サンレアンドロ ・ サンティアゴロード 1 4 2 4 4

F ターム(参考) 2G045 AA34 AA35 BB01 BB07 BB20 BB50 CB01 DA13 DA36 FB01
FB02 FB05 FB06 FB09
4B024 AA01 AA11 BA07 BA61 CA04 DA01 DA02 DA05 DA11 EA04
HA08 HA11 HA19
4B050 CC03 DD11 LL01 LL03
4B063 QA01 QA18 QQ44 QQ52 QR08 QR14 QR20 QR31 QR32 QR42
QR51 QR55 QR62
4B065 AA01X AA57X AA87X AA93Y AB01 AB02 BA01 BA08 CA23 CA24
CA27 CA44 CA46
4C084 AA02 AA06 AA07 AA17 BA01 BA02 BA08 BA22 DC01 NA14
ZA012 ZB072 ZB112 ZB262

4H045 AA10 AA11 AA20 AA30 BA09 CA40 DA89 EA20 FA72 FA73
FA74

专利名称(译)	人体酶分子		
公开(公告)号	JP2004504008A	公开(公告)日	2004-02-12
申请号	JP2001563585	申请日	2001-03-01
[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	タングワイトム リュデュングアイナエム バンドマンオルガ ユエヘンリー アジムザイヤルダ ラルプリーティ バーフォードニール ボーグンマライアール		
发明人	タング、ワイトム リュ、デュング・アイナ・エム バンドマン、オルガ ユエ、ヘンリー アジムザイ、ヤルダ ラル、プリーティ バーフォード、ニール ボーグン、マライア・アール		
IPC分类号	G01N33/48 A61K38/00 A61K38/43 A61K45/00 A61P25/00 A61P29/00 A61P35/00 A61P37/02 A61P43/00 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/00 C12N15/09 C12Q1/68 G01N33/15 G01N33/50 G01N33/53 G01N33/566		
CPC分类号	A01K2217/05 A61K38/00 A61P25/00 A61P29/00 A61P35/00 A61P37/02 A61P43/00 C12N9/00		
FI分类号	C12N15/00.ZNA.A A61K45/00 A61P25/00 A61P29/00 A61P35/00 A61P37/02 A61P43/00.111 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N9/00 C12Q1/68.A C12Q1/68.Z G01N33/15.Z G01N33/48.M G01N33/50.Z G01N33/53.D G01N33/566 C12N5/00.A A61K37/48		
F-TERM分类号	2G045/AA34 2G045/AA35 2G045/BB01 2G045/BB07 2G045/BB20 2G045/BB50 2G045/CB01 2G045/DA13 2G045/DA36 2G045/FB01 2G045/FB02 2G045/FB05 2G045/FB06 2G045/FB09 4B024/AA01 4B024/AA11 4B024/BA07 4B024/BA61 4B024/CA04 4B024/DA01 4B024/DA02 4B024/DA05 4B024/DA11 4B024/EA04 4B024/HA08 4B024/HA11 4B024/HA19 4B050/CC03 4B050/DD11 4B050/LL01 4B050/LL03 4B063/QA01 4B063/QA18 4B063/QQ44 4B063/QQ52 4B063/QR08 4B063/QR14 4B063/QR20 4B063/QR31 4B063/QR32 4B063/QR42 4B063/QR51 4B063/QR55 4B063/QR62 4B065/AA01X 4B065/AA57X 4B065/AA87X 4B065/AA93Y 4B065/AB01 4B065/AB02 4B065/BA01 4B065/BA08 4B065/CA23 4B065/CA24 4B065/CA27 4B065/CA44 4B065/CA46 4C084/AA02 4C084/AA06 4C084/AA07 4C084/AA17 4C084/BA01 4C084/BA02 4C084/BA08 4C084/BA22 4C084/DC01 4C084/NA14 4C084/ZA012 4C084/ZB072 4C084/ZB112 4C084/ZB262 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA09 4H045/CA40 4H045/DA89 4H045/EA20 4H045/FA72 4H045/FA73 4H045/FA74		
優先権	60/186307 2000-03-01 US 60/192532 2000-03-28 US 60/193578 2000-03-30 US		
外部链接	Espacenet		
摘要(译)			

本发明提供了鉴定和编码HEM的人酶分子 (HEM) 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HEM异常表达有关的疾病的方法。

インタクト アミノ酸残基番号	番号	アミノ酸 残基	番号	アミノ酸 残基	番号	アミノ酸 残基
1	2783442001	27	2783442002	27	2783442003	27
2	2116390001	28	2116390002	28	2116390003	28
3	578524001	29	578524002	29	578524003	29
4	187096001	30	187096002	30	187096003	30
5	138841001	31	138841002	31	138841003	31
6	1485405001	32	1485405002	32	1485405003	32
7	2024617001	33	2024617002	33	2024617003	33
8	4721827001	34	4721827002	34	4721827003	34
9	5406614001	35	5406614002	35	5406614003	35
10	1252792001	36	1252792002	36	1252792003	36
11	1295305001	37	1295305002	37	1295305003	37
12	138519001	38	138519002	38	138519003	38
13	2133162001	39	2133162002	39	2133162003	39
14	254557001	40	254557002	40	254557003	40
15	3903488001	41	3903488002	41	3903488003	41
16	5391816001	42	5391816002	42	5391816003	42
17	5885989001	43	5885989002	43	5885989003	43
18	723432001	44	723432002	44	723432003	44
19	1332863001	45	1332863002	45	1332863003	45
20	1559410001	46	1559410002	46	1559410003	46
21	1752387001	47	1752387002	47	1752387003	47
22	1911509001	48	1911509002	48	1911509003	48
23	2213170001	49	2213170002	49	2213170003	49
24	234664001	50	234664002	50	234664003	50
25	288414001	51	288414002	51	288414003	51
26	4103559001	52	4103559002	52	4103559003	52