

(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2004-500811

(P2004-500811A)

(43) 公表日 平成16年1月15日(2004.1.15)

(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	2 G O 4 5
A 6 1 K 38/43	A 6 1 P 25/28	4 B O 2 4
A 6 1 P 25/28	C 1 2 N 1/15	4 B O 5 0
C 1 2 N 1/15	C 1 2 N 1/19	4 B O 6 3
C 1 2 N 1/19	C 1 2 N 1/21	4 B O 6 5
	審査請求 未請求 予備審査請求 有 (全 161 頁) 最終頁に続く	

(21) 出願番号 特願2001-546896 (P2001-546896)
 (86) (22) 出願日 平成12年12月12日 (2000.12.12)
 (85) 翻訳文提出日 平成14年6月19日 (2002.6.19)
 (86) 国際出願番号 PCT/US2000/033743
 (87) 国際公開番号 W02001/046398
 (87) 国際公開日 平成13年6月28日 (2001.6.28)
 (31) 優先権主張番号 09/470,954
 (32) 優先日 平成11年12月22日 (1999.12.22)
 (33) 優先権主張国 米国 (US)

(71) 出願人 301005050
 インサイト・ゲノミックス・インコーポレ
 イテッド
 アメリカ合衆国カリフォルニア州9430
 4・パロアルト・ポータードライブ 31
 60
 (74) 代理人 100089266
 弁理士 大島 陽一
 (72) 発明者 カイザー、マシュー・アール
 アメリカ合衆国カリフォルニア州9454
 6・カストロバレー・ユーイングロード
 4793

最終頁に続く

(54) 【発明の名称】 アスパルチルプロテアーゼ

(57) 【要約】

本発明は、哺乳動物核酸分子及びその断片を提供する。
 本発明はまた、この哺乳動物核酸分子を用いた、アルツ
 ハイマー病やダウン症関連する症状や疾患、または障害
 の特徴付け・診断・評価・治療・予防方法を提供する。
 更に、本発明は、この哺乳動物核酸分子によってコード
 されるタンパク質を産生させるための発現ベクター及び
 宿主細胞を提供する。

```

5' 11 20 29 38 47 56
ACG CCT CAC ATA GGG AAT TTG GCC CTC GAG GCA AGA ATT CGG CAG AGC TCA CTG
65 74 83 92 101 110
CCT TAA GGT GGT CTT TCC CAA GGA GCA TCC CCC AAG CCT TCA AAT TTG CCT TCC
119 128 137 146 155 164
AGT TTA GAC AGG AAG CTG GCC TCT GCC TTT AGG GCC GTA TGA GGA CTC ACT CCT
173 182 191 200 209 218
TGT CTC TTT CTG CCA ANG CCT GAC GAC TCC CTG GAG CCT TTC TTT GAC TCT CTG
C L F L P K P D D S L E P F F D S L
227 236 245 254 263 272
GTA AAG CAG ACC CAC GTT CCC AAC CTC TTC TCC CTG CAG CTT TGT GGT GCT GGC
V K Q T H V P N L F S L Q L C G A G
281 290 299 308 317 326
TTC CCC CTC AAC CAG TCT GAA GTG CTG GCC TCT GTC GGA GGG AGC ATG ATC ATT
F P L N Q S E V L A S V G G S M I I
    
```

【特許請求の範囲】**【請求項 1】**

(a) SEQ ID NO: 46 のアミノ酸配列、
(b) SEQ ID NO: 46 のアミノ酸配列と比較した時の N C I - B L A S T X (バージョン 2.0.4) を用いて計算された B L A S T スコアが少なくとも 250 ビットである天然のアミノ酸配列、
(c) SEQ ID NO: 46 のアミノ酸配列の生物学的に活性な断片、
および (d) SEQ ID NO: 46 のアミノ酸配列の免疫学的に活性な断片から選択されることを特徴とする A S P (アスパルチルプロテアーゼ) をコードする実質的に精製された哺乳動物核酸分子およびその断片、変異体、およびその一部。

10

【請求項 2】

(a) SEQ ID NO: 1 - 5、7 - 29、31 - 33、37、および 43 の核酸配列、
(b) SEQ ID NO: 1 - 43 の配列と比較した時の N C I - B L A S T N (バージョン 2.0.4) を用いて計算された B L A S T スコアが少なくとも 200 ビットである天然の核酸配列、
および (c) SEQ ID NO: 44 の配列と比較した時の N C I - B L A S T N (バージョン 2.0.4) を用いて計算された B L A S T スコアが少なくとも 200 ~ 400 ビットの範囲である天然の核酸配列から選択されることを特徴とする単離され精製された哺乳動物核酸分子またはその相補配列。

20

【請求項 3】

請求項 2 の (a) の哺乳動物核酸配列の断片。

【請求項 4】

高いストリンジェンシーの条件下で、請求項 2 の哺乳動物核酸分子、相補配列、またはその断片とハイブリダイズする少なくとも 16 の連続する核酸を含むプローブ。

【請求項 5】

請求項 2 の哺乳動物核酸分子に機能的に結合したプロモーターを含む組換え核酸分子。

【請求項 6】

請求項 5 の組換え核酸分子で形質転換された細胞。

【請求項 7】

ポリペプチドを生産する方法であって、

(a) 前記ポリペプチドが発現する条件下で、請求項 6 の細胞を培養するステップと、
(b) 発現した前記ポリペプチドを回収するステップとを含むポリペプチド生産方法。

30

【請求項 8】

請求項 5 の組換え核酸分子を含む遺伝子組換え生物。

【請求項 9】

サンプルにおいて哺乳動物核酸分子を検出するための方法であって、

(a) 請求項 2 のプローブを前記サンプルとハイブリダイズさせて、ハイブリダイゼーション複合体を形成させるステップと、
(b) 前記ハイブリダイゼーション複合体を検出するステップとを含み、
前記ハイブリダイゼーション複合体が、前記サンプルにおける前記哺乳動物核酸分子の存在を示すことを特徴とする哺乳動物核酸分子の検出方法。

40

【請求項 10】

前記ハイブリダイゼーションの前に、前記核酸分子またはその断片を増幅するステップを更に含むことを特徴とする請求項 9 に記載の方法。

【請求項 11】

哺乳動物核酸分子またはその断片を用いて分子のライブラリをスクリーニングし、前記哺乳動物核酸分子と特異的に結合する少なくとも 1 つのリガンドを同定する方法であって、
(a) 特異的な結合が許容される条件下で、請求項 2 の哺乳動物核酸分子を前記分子のライブラリと結合させるステップと、

50

(b) 特異的な結合を検出して、前記哺乳動物核酸分子と特異的に結合するリガンドを同定するステップとを含むことを特徴とするリガンド同定方法。

【請求項 12】

前記ライブラリが、DNA 分子、RNA 分子、PNA、ペプチド、およびタンパク質から選択されることを特徴とする請求項 11 に記載の方法。

【請求項 13】

哺乳動物核酸分子またはその断片を用いて、サンプルから前記哺乳動物核酸分子と特異的に結合するリガンドを精製する方法であって、

(a) 特異的な結合が許容される条件下で、請求項 2 の哺乳動物核酸分子またはその断片とサンプルとを結合させるステップと、

10

(b) 結合した前記哺乳動物核酸分子を回収するステップと、

(c) 前記リガンドから前記哺乳動物核酸分子を分離して、精製したリガンドを得るステップとを含むことを特徴とするリガンド精製方法。

【請求項 14】

(a) SEQ ID NO: 46 のアミノ酸、

(b) SEQ ID NO: 46 のアミノ酸配列と比較した時の NCBI - BLASTX (バージョン 2.0.4) を用いて計算された BLAST スコアが少なくとも 250 ビットである天然のアミノ酸配列、

(c) SEQ ID NO: 46 のアミノ酸配列の生物学的に活性な断片、

および (d) SEQ ID NO: 46 のアミノ酸配列の免疫学的に活性な断片から選択されることを特徴とする実質的に精製されたタンパク質またはその一部。

20

【請求項 15】

タンパク質またはその一部を用いて分子のライブラリをスクリーニングし、前記タンパク質と特異的に結合する少なくとも 1 つのリガンドを同定する方法であって、

(a) 特異的な結合が許容される条件下で、請求項 14 のタンパク質またはその一部を前記分子のライブラリと結合させるステップと、

(b) 特異的な結合を検出して、前記タンパク質と特異的に結合するリガンドを同定するステップとを含むことを特徴とするリガンド同定方法。

【請求項 16】

前記ライブラリが、DNA 分子、RNA 分子、PNA、ペプチド、タンパク質、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、インヒビター、薬剤化合物、および医薬品から選択されることを特徴とする請求項 15 に記載の方法。

30

【請求項 17】

タンパク質またはその一部を用いて分子のライブラリをスクリーニングし、サンプルから前記タンパク質と特異的に結合する少なくとも 1 つのリガンドを精製する方法であって、

(a) 特異的な結合が許容される条件下で、請求項 14 のタンパク質またはその一部とサンプルとを結合させるステップと、

(b) 結合した前記タンパク質を回収するステップと、

(c) 前記リガンドから前記タンパク質を分離して、精製したリガンドを得るステップとを含むことを特徴とするリガンド精製方法。

40

【請求項 18】

効果的な量の請求項 14 のタンパク質および医薬的に許容される賦形剤を含む医薬組成物。

【請求項 19】

ASP の発現の変化に関連する疾患や症状を治療する方法であって、そのような治療が必要な患者に請求項 18 の医薬組成物を投与することを特徴とする治療方法。

【請求項 20】

前記疾患や症状がアルツハイマー病およびダウン症から選択されることを特徴とする請求項 19 に記載の方法。

【発明の詳細な説明】

50

【0001】

本願はPCT出願であり、1999年12月22日に出願の米国非仮出願第09/470,954号(整理番号PC-0008US)の恩典を請求するものである。

【0002】

(発明の技術分野)

本発明は、新規の哺乳動物タンパク質の核酸配列及びアミノ酸配列、およびその核酸配列変異体、並びにこれらの配列を用いたアルツハイマー病及びダウン症の特徴付け・診断・予防・治療に関連する。

【0003】

(発明の背景)

生物間の系統発生的な関連が何度も実証され、様々な原核生物及び真核生物の研究によって、生化学的及び生理学的な機構や代謝経路の実質的に漸進的な進化が示された。異なった進化の圧力にもかかわらず、酵母及び線虫、ハエ、ラット及びヒトにおける細胞周期を調節するタンパク質が、共通の化学的または構造的な特徴を有し、同一の一般的な細胞活性を調節する。ヒト遺伝子配列を、構造及び/または機能が既知である他の生物の遺伝子配列と比較することによって、研究者は類似性を引き出し、仮説を検証するためのモデル系を作製することができる。モデル系は、ヒトの症状、疾患、及び障害の診断薬及び治療薬の開発に極めて重要である。

【0004】

死後の検査におけるアルツハイマー病の神経病理学的なシグネチャである老人性神経プラーク(senile neuritic plaques)が、脳の海馬および新皮質領域で見ついている。海馬は、大脳辺縁系の一部であって、学習および記憶に重要な役割を果たす。アルツハイマー病では、プラークの蓄積が海馬の神経構造に損傷を与え、最終的に記憶プロセスを不全にする。この疾患は家族性あるいは散発性であって、40歳の人々が感染するなど早期に発症することがある。家族性アルツハイマー病は、遺伝性の強い常染色体の優勢形質であると考えられてきた。しかしながら、より多くの遺伝的決定基が単離されるにつれこの考えが変わってきた。例えば、老人性プラークに見られるタンパク質である、アポリポタンパク質E(ApoE)の或る正常なアレル変異体が、この疾患の発症リスクを抑制するか或いは高めるかの何れかに作用する(Strittmatterら(1993) Proc. Natl. Acad. Sci. 90: 1977-198)。

アミロイド前駆体タンパク質(APP)、プレセニリン1、およびプレセニリン2の3種類の別の遺伝子の突然変異も、人がアルツハイマー病を発症する素因であることが分かった(Selkoe(1999) Nature 399: A23-A31)。

【0005】

アミロイドタンパク質(A)が老人性プラークの主な成分であって、通常はセクレターゼおよびセクレターゼがAPPを切断して形成される。アルツハイマー病の患者では、多量のAが生成され、細胞外のこれらの神経病理学的なプラークに蓄積されている。Aが蓄積するメカニズムを究明するために、2種類の酵母アスパルチルプロテアーゼ(aspartyl proteases)が*in vitro*でAPPを処理することが見出されたため、APP切断セクレターゼに焦点を絞った(Zhangら(1997) Biochim Biophys Acta 1359: 110-122)。

ペプチド様プローブを用いた近年の研究により、これらのセクレターゼが膜内切断アスパルチルプロテアーゼであることが分かった(Wolfeら(1999) Biochemistry 38: 4720-4727)。

【0006】

プレセニリン1遺伝子は、APPカルボキシル末端を切断するセクレターゼの候補である。この疾患プロセスにおけるプレセニリンの関与を支持する幾つかの証拠がある。その証拠の中には、プレセニリンおよびAPPの免疫共沈降、並びにプレセニリン遺伝子におけるミスセンス点突然変異が、アルツハイマー病の特に攻撃性の早期発症を引き起こすことなどがある(HaasおよびDestrooper(1999) Science

10

20

30

40

50

286: 916-919)。プレセニリンは、notchタンパク質を切断すると考えられている。APPがセクターゼおよびセクターゼによってその外部ドメインに切断されると同様に、プレセニリンがこのnotchタンパク質を切断して、セクターゼの基質であるカルボキシル末端断片を生成する(HaasおよびDe Strooper, 前出)。

【0007】

アスパラギン酸プロテアーゼはリソソームプロテアーゼのカテプシンファミリーのメンバーであって、ペプシンA、ガストリクシン、キモシン、レニン、およびカテプシンDおよびEを含む。アスパラギン酸プロテアーゼは活性部位における2つのアスパラギン酸残基を含み、pH2~3の範囲で最も活性であって、その状態では一方のアスパラギン酸がイオン化し、他方はイオン化していない。アスパルチルプロテアーゼ(aspartyl protease)活性部位の触媒アスパラギン酸残基は、アミノ酸残基配列D-T/S-G-T/S-T/S(アスパルチルプロテアーゼシグネチャPS00141; PROSITE PDOC00128)によって特徴付けられる。ヘキサペプチドペプスタチンはアスパラギン酸プロテアーゼの強力なプロモーターであって、遷移状態にある時は通常の前駆体と類似している。またアスパラギン酸プロテアーゼには、細菌ペニシロペプシン、哺乳動物ペプシンおよびキモシン、および或る種の真菌プロテアーゼが含まれる。カテプシンの異常な調節および発現は、関節炎などの様々な炎症性疾患の症状を示すものである。

10

【0008】

本発明は、アスパルチルプロテアーゼをコードする更なる核酸配列およびその核酸配列変異体の発見に基づき、アルツハイマー病およびダウン症の特徴付け・診断・予防・治療に有用な組成物を提供する。

20

【0009】

(発明の開示)

本発明は、アスパルチルプロテアーゼ(ASP)をコードする哺乳動物核酸分子、およびその核酸配列変異体の発見に基づき、アルツハイマー病およびダウン症など症状の特徴付け・診断・予防・治療に有用な組成物を提供することで当分野の要求に応える。

【0010】

本発明は、SEQ ID NO: 46及びSEQ ID NO: 47のアミノ酸配列またはそれらの一部、SEQ ID NO: 46の生物学的に活性な断片、およびSEQ ID NO: 46の免疫学的に活性な断片を含む本哺乳動物タンパク質をコードするSEQ ID NO: 1-43またはそれらの断片を含む単離され精製されたヒト、サル、およびラットの核酸分子を提供する。

30

【0011】

本発明はまた、高いストリンジェンシーの条件下で、本哺乳動物核酸分子またはその断片とハイブリダイズするプローブを提供する。本発明はまた、SEQ ID NO: 1-43の核酸分子に相補的な単離され精製された核酸分子を提供する。一実施態様では、このプローブは、一本鎖の相補的なRNA分子またはDNA分子である。

【0012】

本発明は更に、サンプルにおいて核酸分子を検出する方法であって、プローブをそのサンプルの少なくとも1つの核酸分子とハイブリダイズするステップと、ハイブリダイゼーション複合体を形成するステップと、そのハイブリダイゼーション複合体を検出するステップとを含み、そのハイブリダイゼーション複合体の存在がそのサンプルに核酸分子存在することと相関する方法を提供する。一実施態様では、この方法は、ハイブリダイゼーションの前にその核酸分子を増幅するステップを更に含む。この核酸分子若しくはその断片を、マイクロアレイ上のエレメント若しくは標的として用いることができる。

40

【0013】

本発明はまた、核酸配分子若しくはその断片を用いて、その核酸分子と特異的に結合する少なくとも1つのリガンドを同定するべく、分子のライブラリをスクリーニングする方法であって、特異的な結合が許容される条件下で、その核酸分子をその分子のライブラリと結

50

合させるステップと、その特異的な結合を検出して、その核酸分子と特異的に結合するリガンドを同定するステップとを含むスクリーニング方法を提供する。このようなライブラリには、DNA分子及びRNA分子、ペプチド、PNA、タンパク質等が含まれる。類似の方法で、核酸分子またはその断片を用いてリガンドを精製することができる。

【0014】

本発明はまた、本核酸分子の少なくともある断片を含む発現ベクターを提供する。別の実施態様では、この発現ベクターは宿主細胞内に含まれている。本発明は更に、タンパク質が発現する条件下で宿主細胞を培養するステップと、そのタンパク質をその宿主細胞から回収するステップとを含むタンパク質生産方法を提供する。

【0015】

本発明はまた、実質的に精製され哺乳動物ASPまたはその一部を提供する。更に、本発明は、SEQ ID NO: 46のアミノ酸配列、SEQ ID NO: 46の生物学的に活性な断片、およびSEQ ID NO: 46の免疫学的に活性な断片を有する単離され精製されたタンパク質を提供する。本発明はまた、医薬用担体と共に実質的に精製された哺乳動物タンパク質またはその一部を含む医薬組成物を提供する。

10

【0016】

本発明は更に、その哺乳動物タンパク質の一部を用いて抗体を産生させる方法を提供する。本発明はまた、哺乳動物タンパク質若しくはその一部を用いて、そのタンパク質に特異的に結合する少なくとも1つのリガンドを同定するべく、分子のライブラリをスクリーニングする方法であって、特異的な結合が許容される条件下で、そのタンパク質を分子のライブラリと結合させるステップと、その特異的な結合を検出して、そのタンパク質と特異的に結合するリガンドを同定するスクリーニング方法を提供する。このようなライブラリには、DNA分子及びRNA分子、ペプチド、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、薬剤、医薬品、その他のリガンドが含まれる。一実施態様では、この方法で同定されたりガンドがその哺乳動物タンパク質の活性を調節する。類似の方法で、そのタンパク質若しくはその一部を用いてリガンドを精製する。本発明はまた、特異的な結合が許容される条件下で、その哺乳動物タンパク質またはその一部をサンプルと結合させて、そのタンパク質とリガンドとの特異的な結合を検出して、その結合したタンパク質を回収し、そのリガンドからそのタンパク質を分離して、精製されたりガンドを得る方法を提供する。

20

30

【0017】

本発明はまた、哺乳動物のゲノムDNAの中にマーカー遺伝子を挿入して、その天然の哺乳動物核酸分子の発現を阻害する方法を提供する。本発明はまた、その哺乳動物核酸分子を用いて哺乳動物モデル系を作製する方法であって、その哺乳動物核酸分子を含むベクターを作製するステップと、そのベクターを全能性哺乳動物胚性幹細胞の中に導入するステップと、ゲノムDNAの中にそのベクターが組み込まれた胚性幹細胞を選択するステップと、この選択した細胞を哺乳動物胚盤胞の中に微量注入して、キメラ胚盤胞を形成するステップと、偽妊娠メスにキメラ胚盤胞を移植するステップと、このメスが、生殖細胞系の中に少なくとも1つの哺乳動物核酸分子の別の複製を含むキメラ哺乳動物を出産し、そのキメラ哺乳動物を交配して同型接合哺乳動物モデル系を作製するステップとを含む哺乳動物モデル系作製方法を提供する。

40

【0018】

(本発明の実施方法について)

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るものであり、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。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」は当業者には周知の複数の宿主細胞を含む。

50

【0019】

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。本明細書に記載の全ての文献は、本発明に関連して使用する可能性のある文献に記載された細胞系、プロトコル、試薬、ベクターを記述し開示するために引用した。従来の特許を引用したからと言って、本発明の新規性が損なわれると解釈されるものではない。

【0020】

(定義)

「アスパルチルプロテアーゼ (aspartyl protease)」または「ASP」は、天然、合成、半合成或いは組み換え体などの全ての種(マウス、ウシ、ヒツジ、ブタ、齧歯類、イヌ、サルを含むが好ましくはヒトである哺乳動物)から得られる、実質的に精製されたタンパク質を指す。

10

【0021】

「遺伝子発現の変化」および「発現の変化」は、サンプルにおける遺伝子発現の増大および減少、並びに転写されたメッセンジャーRNA転写産物の存在または不在を指す。

【0022】

「生物学的に活性」は、構造的または免疫学的、調節的、化学的な機能を有する、天然或いは組み換え、または合成分子であるタンパク質を指す。

【0023】

「相補的な」は、プリン塩基とピリミジン塩基との自然な水素結合による塩基対形成を指す。例えば、配列A-C-G-Tは、その相補配列T-G-C-AまたはU-G-C-Aと水素結合する。2つの一本鎖分子の相補性は、ヌクレオチドの幾つかのみが結合する部分的な相補性と、ヌクレオチドの殆ど全てが結合する完全な相補性とがある。核酸鎖間の相補性の程度は、ハイブリダイゼーション及び増幅反応の効率及び強度に影響する。

20

【0024】

「誘導体」は、化学修飾された核酸分子やタンパク質配列を指す。化学修飾には、分子や配列の生物学的活性または寿命を維持或いは増大する、アルキル基またはアシル基、アミノ基による水素の置換や、グリコシル化、ポリエチレングリコール化、または類似の任意のプロセスが含まれる。

【0025】

「断片」は、有用な機能的特性を維持する核酸分子の任意の一部またはインサイトシクローンを指す。有用な断片には、ハイブリダイゼーションや増幅技術、または複製や転写、翻訳の調節に有用なオリゴヌクレオチドを含む。

30

【0026】

「ハイブリダイゼーション複合体」は、プリン塩基とピリミジン塩基との間の水素結合の形成による2つの核酸分子の複合体を指す。

【0027】

「リガンド」は、核酸分子やタンパク質の相補部位に特異的に結合する任意の分子や物質、または化合物を指す。このようなリガンドは、核酸及びタンパク質、炭水化物、脂肪、脂質を含む無機及び有機物質の少なくとも1つからなり、本発明の核酸分子やタンパク質の活性を安定化させたり、調節したりする。

40

【0028】

「調節する」は、分子と核酸分子がタンパク質の何れかとの特異的な結合によって、寿命や活性(生物学的または化学的、免疫学的)を変化させることを指す。

【0029】

「核酸分子」は、核酸、オリゴヌクレオチド、ヌクレオチド、ポリヌクレオチド、またはそれらの任意の断片を指す。また、「核酸分子」は、ゲノム若しくは合成起源の二本鎖若しくは一本鎖のDNA或いはRNAであり、炭水化物または脂質、タンパク質、その他の物質と結合して、形質転換などの特殊な作用を起こしたり、ペプチド核酸(PNA)等の有用な成分を形成する。「オリゴヌクレオチド」は、アンプリマー(amplimer)

50

及びプライマー、オリゴマー、エレメント、標的、プローブと実質的に同一であり、好ましくは一本鎖である。核酸分子は別の核酸分子のスプライスバリエーションであり得る。

【0030】

「タンパク質」は、天然或いは合成のアミノ酸配列、オリゴペプチド、ペプチド、ポリペプチド、またはそれらの一部を指す。

【0031】

ここで用いる「一部」は、あらゆる目的に用いられるタンパク質の任意の一部を指すが、特に、その一部と特異的に結合する分子のライブラリのスクリーニングや抗体の生産に用いられる。「一部」の典型的な例は、SEQ ID NO: 46の初めの40の連続するアミノ酸である。

10

【0032】

ポリヌクレオチド配列に用いる「類似性」とは、標準的なアルゴリズムを用いてアラインメントされた少なくとも2つの配列間の一致する残基数から求められる。このようなプログラムは、2つの配列間のアラインメントを最適化するべく、標準的かつ再現性のある方法で比較する配列内にギャップを挿入可能なため、2つの配列間のより有意な比較を達成することが可能である。

【0033】

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、National Center for Biotechnology Information (NCBI)、(Bethesda, MD)によって提供される。Basic Local Alignment Search Tool (BLAST) (Altschul, ら (1990) J. Mol. Biol. 215: 403-410; Altschul (1993) J. Mol. Evol. 36: 290-300)は、NCBI (<http://WWW.ncbi.nlm.nih.gov/BLAST/>) および幾つかのほかの情報源から入手可能である。このBLASTソフトウェア一式には、既知のポリヌクレオチド配列と様々なデータベースの別のポリヌクレオチド配列とのアラインメントに用いられる「blastn」を含む、様々な配列分析プログラムが含まれる。「BLAST 2 Sequences」と呼ばれるツールが入手可能であり、2つのヌクレオチド配列の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、対話形式で利用できる。「BLAST 2 Sequences」ツールは、blastn及びblastp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ーと共に用いられる。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 2.0.09 (May - 07 - 1999)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。

20

30

【0034】

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

類似性は、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

40

50

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。

【0035】

「サンプル」は、その最も広い意味で用いられる。核酸分子を含むサンプルは、体液や、細胞抽出物、細胞から単離された染色体、細胞小器官、膜や、溶液中に遊離している或いは基板に結合されたゲノムDNAまたはRNA、cDNAや、組織、細胞、組織プリント、またはフィンガープリント等を含み得る。

【0036】

「実質的に精製された」は、その自然環境から切り離されてから分離或いは単離された、自然環境では結合しているその他の成分が少なくとも約60%、好適には約75%、最も好適には、約90%取り除かれた核酸分子またはタンパク質を指す。 10

【0037】

「基板」は、核酸分子またはタンパク質が結合した任意の固体或いは半固体の支持物を指し、膜またはフィルター、チップ、スライド、ウエハ、ファイバー、磁気または非磁気ビーズ、ゲル、毛細管または他のチューブ、プレート、ポリマー、微小粒子が含まれる。この基板は、孔または溝、ピン、チャンネル、細孔を含む様々な表面形態を有する。

【0038】

「変異体」、「核酸分子変異体」または「核酸配列変異体」は、ASPをコードする核酸分子のサブライソバリエーションであって、少なくとも100最も好ましくは400以上のBLASTスコアによって決定され得る核酸分子を指す。 20

【0039】

(発明)

本発明は、ASPをコードする新規の哺乳動物核酸分子およびその核酸分子変異体の発見に基づき、アルツハイマー病およびダウン症などの病態の特徴付け・診断・治療・予防においてその核酸分子やその断片、またはそのタンパク質やその一部を用いることに関連する。

【0040】

本発明のASP1をコードする核酸分子は、アミノ酸配列アライメント用のコンピュータ検索を用いて結腸組織cDNAライブラリ(COLNTUT06)からのインサイトクローン1611565H1において初めに同定された。コンセンサス配列SEQ ID NO 30 : 43は、インサイトクローン1863920X321F1 (PROSNOT19)、2233901H1 (PANCTUT02)、1611565T1 (COLNTUT06)、336541H1 (EOSIHET02)、およびショットガン配列SBA A03902F1に由来する。

【0041】

ASP2をコードする核酸は、アミノ酸配列アライメント用のコンピュータ検索を用いて皮膚組織cDNAライブラリ(SKINBIT01)からのインサイトクローン1869869H1において初めに同定された。コンセンサス配列SEQ ID NO : 44は、インサイトクローン1869869F6 (SKINBIT01)、2549562F6 (LUNG TUT06)、3338603H1 (SPLNNOT10)、およびショットガン配列70317925D1、70317570D1、および70319214D1に由来する。 40

【0042】

本発明の哺乳動物アスパルチルプロテアーゼ(aspartyl protease)をコードする核酸分子およびその核酸分子変異体は、本発明のASP1およびASP2をコードする核酸配列およびアミノ酸配列をLIFESEQデータベースおよびZOOSEQデータベース(Incyte Genomics, Palo Alto CA)における核酸配列と比較して初めに同定した。

【0043】

ASPは、423のアミノ酸からなるSEQ ID NO : 46のアミノ酸配列および32 50

2のアミノ酸からなるSEQ ID NO: 47のアミノ酸配列を含む。次に示すASP1 (SEQ ID NO: 46)の化学的および構造的な特徴付けは、図3A - 図3Dに示されている絶対位置(1列につき30残基)に基づいて行われた。

【0044】

ヒトASP1タンパク質(SEQ ID NO: 46)は、N75およびN271残基における2つの潜在的なNグリコシル化部位と、K102残基におけるロイシンジッパー、A29、H33、L84、L114、S118、Q251、L315、I354、およびC367残基における9箇所のカゼインキナーゼIIリン酸化部位と、D37、T42、K51、L272、およびE275残基における5箇所のプロテインキナーゼCリン酸化部位とを有する。加えて、SEQ ID NO: 46は、P12残基におよびY205残基における2つの真核およびウイルスアスパルチルプロテアーゼを有する。SEQ ID NO: 46は、SEQ ID NO: 47とは43.5%の類似性、ヒトガストリクシン(SEQ ID NO: 48; GI 1658286)とは18.8%の類似性、Neurospora crassa空胞プロテアーゼ(vacuolar protease)(SEQ ID NO: 49; GI 1039445)とは14.1%の類似性とを有する。類似性のパーセントは、LASERGENEソフトウェアのMEGALIGNプログラム(DNASTAR)を用いて計算した。SEQ ID NO: 46のT35残基からG47残基までの抗体の生産に有用であるアミノ酸は、LASERGENEソフトウェアのPROTEANプログラム(DNASTAR)で選択した。BLOCKS分析($p < 10^{-12}$)によって、SEQ ID NO: 46のL40からG55、G129からT137、G187からG196、A235からL244、およびV341からA364のアミノ酸配列同一性に基づいてSEQ ID NO: 46がアスパルチルプロテアーゼと同定された。更に、PRINTS分析($p < 10^{-8}$)によって、SEQ ID NO: 46のI33からV53、G182からK195、A235からL244、およびV240からD354の残基から、SEQ ID NO: 46において2つのアスパルチルプロテアーゼシグネチャが同定された。

【0045】

SEQ ID NO: 44を用いて、インサイトLIFESEQデータベース(Incyte Genomics)においてヒトASP2核酸分子変異体が同定された。1499218H1(SINTBST01)(インサイトクローン番号(ライブラリ))、1438719T6(PANCN08)、3642853T6(LUNGN034)、1499218F6(SINTBST01)、1377178H1および1377178F1(LUNGN010)、3642853F6(LUNGN034)、5577513H1(BRAPN04)、5094633H1(EPIMN05)、2293393H1(BRAIN01)、2908594H1(THYMN05)、2007269H1(TESTN03)、1438613F1(PANCN08)、3484819H1(KIDN031)、2757870H1(THP1AZS08)、2900063H1(DRGCN01)、2123411H1(BRSTN07)、1438719H1(PANCN08)、および1438613H1(PANCN08)、SEQ ID NO: 17-28はそれぞれ、コンセンサス配列であるSEQ ID NO: 44の変異体である(図4A - 図4H)。推定上のスプライス部位は記号で示した。これらのクローンは、生殖(3/15)、心血管(2/15)、消化系(4/15)、神経(3/15)、造血(2/15)、および泌尿器(1/15)の組織、臓器、または関連する系に由来するライブラリで発現した。

【0046】

SEQ ID NO: 43を用いて、インサイトLIFESEQデータベース(Incyte Genomics)においてヒトASP1核酸分子変異体が同定された。5157179F6(BRSTTMT02)(インサイトクローン番号(ライブラリ))、3985758H1(UTRSTUT05)、2846604H1(DRGLN01)、5743028H1(LUNGN03)、3751907H1(UTRSNOT1)

8)、4148749H1 (SINITUT04)、1855226H1および1855420H1 (HNT3AZT01)、5114558H1 (ENDITXT01)、839538R1 (PROSTUT05)、5588490H1 (ENDINOT02)、4999662H1 (MYEPTXT02)、3630054H1 (COLNNOT38)、および4108264H1 (PROSBPT07)、SEQ ID NO: 29-42はそれぞれ、コンセンサス配列であるSEQ ID NO: 43の変異体である(図5A - 図5G)。推定上のスプライス部位は記号で示した。これらのクローンは、生殖(5/15)、心血管(4/15)、消化系(2/15)、胸部神経節(1/15)、造血(2/15)、および奇形癌細胞系(2/15)の細胞、組織、臓器、または関連する系に由来するライブラリで発現した。

10

【0047】

本発明の哺乳動物アスパルチルプロテアーゼをコードする更なる核酸分子は、BLAST若しくはBLAST2 (Altschulら(1997) Nucleic Acid Res. 25: 33893402)を用い、ZOOSEQ データベース (Incyte Genomics) においてラットおよびサルクローンをヒトアスパルチルプロテアーゼ (SEQ ID NO: 43および44) とアラインメントして同定された。SEQ ID NO: 1-16の配列表に存在するクローン(様々なライブラリに由来する): 700251567H1 (RABMNOT01)、701432228T1 (RALITXT06)、700769871H1 (RAHYNOT02)、700769257H1 (RAHYNOT02)、701188357H1 (RACONON02)、701918575H1 (RABYUNS03)、700921566H1 (RANPNOT02)、700494841H1 (RABCNOT01)、700885432H1 (RAVANOT01)、701901708H1 (RABYUNS09)、701911752H1 (RABYUNS03)、701823687H1 (RAKITXT08)、701596150H1 (RALITXT26)、700542176H1 (RACONOT01)、701031389H1 (RABMNON02)、700376870H1 (RALANOT01)、および700709032H1 (MNBFNOT01)は、ASPsをコードする核酸分子SEQ ID NO: 43および44の非ヒト変異体である(図6A - 図6Eおよび図7A - 図7G)。推定上のスプライス部位は記号で示した。これらのクローンは、神経(10/16)、消化系(4/16)、心血管(1/16)、および泌尿器系(1/16)の組織、臓器、または関連する系に由来するライブラリで発現した。

20

30

【0048】

核酸配列、SEQ ID NO: 17-42は、サンプルにおいてSEQ ID NO: 43および44または類似の分子の同定および区別をするためのハイブリダイゼーションおよび増幅技術に用いることができる。この分子を用いて遺伝子組換えによりヒトの症状、疾患、または障害を模した疾患動物モデルを作製したり、動物毒素研究、臨床検査、および被験動物/患者の治療プロファイルをモニタリングすることが可能である。

【0049】

(本発明の特徴及び使用)

40

cDNAライブラリ

ここに開示する特定の実施例では、当分野で周知の方法を用いて哺乳動物の細胞及び組織からmRNAを単離し、これを用いてcDNAライブラリを作製する。上記したインサイト社クローンは、哺乳動物cDNAライブラリから単離された。本発明の代表的な少なくとも1つのライブラリの作製方法を以降に記載する実施例に示す。コンセンサス哺乳動物配列は、Phrap (P. Green, 前出)及びGELVIEW断片構築システム (Genetics Computer Group, Madison WI)、AUTOASSEMBLERアプリケーション (Applied Biosystems, Foster City CA)などのコンピュータプログラムを用いて、インサイト社クローンを含む断片、伸長、及び/またはショットガン配列から化学的かつ/または電子的に構築

50

した。クローン、伸長配列および/またはショットガン配列は、クラスターおよび/または主クラスターの中に電子的に組み入れられた。

【0050】

シーケンシング

核酸をシーケンシングする方法は当分野で周知であり、そのような方法を用いて本発明の任意の実施例を実施することができる。これらの方法は、DNAポリメラーゼIであるクレノウフラグメント、SEQUENASE、Taq DNAポリメラーゼおよび熱耐性T7 DNAポリメラーゼ(Amersham Pharmacia Biotech (APB), Picataway NJ)、或いはELONGASE増幅システム(Life Technologies, Rockville MD)に用いられるような校正エクソヌクレアーゼとポリメラーゼとの組み合わせを用いることができる。配列の準備は、HYDRAMICRODISPENSER(Robbins Scientific, Sunnyvale CA)、MICROLAB 2200(Hamilton, Reno NV)、及びDNA ENGINEサーマルサイクラー(PTC200; MJ Research, Watertown MA)などの装置を用いて自動的に行うのが望ましい。シーケンシングに用いる装置には、ABI 3700、377または373 DNAシーケンシングシステム(Applied Biosystems)、及びMEGABACE 1000 DNAシーケンシングシステム(APB)等がある。当分野で周知の様々なアルゴリズムを用いて、シーケンシングした配列を解析することができる。これらのアルゴリズムは、Ausubel(1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7) 及びMeyers(1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856 - 853)に記載されている。

【0051】

ショットガンシーケンシングを用いて、多数の生物に由来するクローニング挿入断片からより多くの配列を作り出す。ショットガンシーケンシング方法は当分野で周知であり、熱耐性DNAポリメラーゼ類や非熱耐性DNAポリメラーゼ類、及び目的の核酸分子に隣接する代表的な領域から選択されたプライマーを用いる。当分野で周知のCONSED(Gordon(1998) Genome Res. 8:195-202)などの様々なアルゴリズムやプログラムを用いて、組み立てが不完全な配列を調べる。ベクターやキメラ配列、または欠失配列を含む汚染配列を除去して、組み立てが不完全な配列を完全な配列に組み立てる。

【0052】

核酸配列の伸長

本発明の配列は、当分野で周知の様々なPCR法を用いた方法で伸長することができる。例えば、XL-PCRキット(Applied Biosystems)及び入れ子プライマー(nested primer)、市販のcDNAまたはゲノムDNAライブラリを用いてヌクレオチド配列を伸長することが可能である。全てのPCR系の方法に用いることができるように、プライマーは、OLIGO 4.06プライマー分析ソフトウェア(National Biosciences, Plymouth MN)等の市販のソフトウェアを用いて、ヌクレオチドの長さが約22~30個、GC含量が約50%以上、約55~68の温度で標的配列とアニールするように設計することが可能である。調節エレメントを復活させるために配列を伸長する場合は、cDNAライブラリよりゲノムライブラリを用いる方が良い。

【0053】

(本哺乳動物核酸分子の使用)

ハイブリダイゼーション

本哺乳動物核酸分子及びその断片は、様々な目的のための様々なハイブリダイゼーション技術に用いることができる。プローブは、5'調節領域やアスパルチルプロテアーゼグネ

10

20

30

40

50

チャなどの保存されたモチーフから作製可能であり、これらのプローブを、この哺乳動物タンパク質やアレル変位配列、または関連分子をコードする天然の分子を同定するためのプロトコルに用いることができる。通常は一本鎖であるDNA若しくはRNAからなるこのプローブは、任意のこの核酸配列と少なくとも50%の配列同一性を有することが望ましい。ハイブリダイゼーションプローブは、標識されたヌクレオチドの存在下でのPCR増幅、オリゴ標識化、ニックトランスレーション法、または末端標識化を利用して作製することができる。この核酸分子またはその断片を含むベクターを用いて、RNAポリメラーゼ及び標識したヌクレオチドを加えて *in vitro* で mRNA プローブを作製することができる。これらの方法は *Amersham Pharmacia Biotech (APB)* 社が販売するキットを用いて行うことができる。

10

【0054】

ハイブリダイゼーションのストリンジェンシー（厳密性）は、プローブのGC含量、塩濃度、及び温度によって決まる。特に、塩濃度を下げる、またはハイブリダイゼーションの温度を上げて、ストリンジェンシーを高めることができる。ある膜系のハイブリダイゼーション用の溶液にホルムアミドなどの有機溶媒を加えて、反応が低い温度で起こるようにすることができる。ハイブリダイゼーションは、低いストリンジェンシーの緩衝液（5xSSC、1%のドデシル硫酸ナトリウム（SDS））で、60℃で行うことができるが、核酸配列間に不適正塩基対を含む複合体の形成を許容し得る。続く洗浄は、45℃（中程度のストリンジェンシー）或いは68℃（高いストリンジェンシー）の何れかの温度、0.2xSSC、0.1% SDSなどの高いストリンジェンシーで行う。高いストリンジェンシーでは、ハイブリダイゼーション複合体は、完全に相補的な核酸分子部分のみが安定して保持される。ある膜系のハイブリダイゼーションにおいて、好ましくは35%、最も好ましくは50%のホルムアミドをハイブリダイゼーション溶液に加えて、ハイブリダイゼーションを行う温度を下げたり、または、SarkosylやTriton X-100（Sigma Aldrich, St. Louis MO）などの界面活性剤及び変性したサケ精子DNAなどのブロッキング試薬を用いてバックグラウンドシグナルを低減することが可能である。ハイブリダイゼーションの条件や要素の選択については当分野で周知であり、Ausubel (*supra*) 及び Sambrook 他（1989）*Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY. に記載されている。

20

30

【0055】

当分野で周知の方法でマイクロアレイを準備して分析することができる。オリゴヌクレオチドをマイクロアレイのプローブや標的として用いることができる。マイクロアレイを用いて、同時に極めて多数の遺伝子の発現レベルをモニタリングし、遺伝子変異体、突然変異及びSNP（一塩基多型）を同定することができる。このようなデータを用いて、遺伝子機能の解明や、症状及び疾患、または障害における遺伝子原理の解明や、症状及び疾患、障害の診断または治療、治療薬の開発、並びにこれらの治療薬の活性のモニタリングが可能である（例えば、Brennan 他（1995）USPN 5, 474, 796；Schena 他（1996）Proc. Natl. Acad. Sci. 93:10614-10619；Baldeschweiler 他（1995）PCT出願 WO 95/251116；Shalon 他（1995）PCT出願 WO 95/35505；Heller 他（1997）Proc. Natl. Acad. Sci. 94:2150-2155；and Heller 他（1997）USPN 5, 605, 662を参照）。

40

【0056】

ハイブリダイゼーションプローブはまた、天然のゲノム配列のマッピングに有用である。このプローブは、（1）特定の染色体、（2）染色体の特定の領域、（3）ヒト人工染色体（HAC）や酵母人工染色体（YAC）、細菌人工染色体（BAC）、細菌P1作製物、または単一の染色体、（5）これらの試料の任意のものから作製されたDNAライブラ

50

りなどの人工染色体作製物、にハイブリダイズすること可能である。

【0057】

発現

A S Pをコードする多数の核酸分子をベクターにクローニングして、このタンパク質若しくはその一部を宿主細胞で発現させることができる。この核酸配列を、DNAシャフリング (Stemmer and Cramer i (1996) USPN 5, 830, 721) や部位特異的変異誘発などの方法によって、新規の制限部位を作り出したり、グリコシル化パターンを変えたり、優先コドンを変えて特定の宿主における発現を増大させたり、スプライスバリエーションを作り出したり、半減期を延長する等の操作が可能である。この発現ベクターは、特定の宿主における各要素の効率に基づいて選択された様々なサンプルに由来する転写及び翻訳調節エレメント (プロモーター及びエンハンサー、特定の開始シグナル、ポリアデニル化3'配列) を含み得る。in vitro組み換えDNA技術、合成技術及び/または in vivo 遺伝子組み換え技術を組み合わせ、このベクターに核酸配列と調節エレメントをつなぐことができる。このような技術は、当分野で周知であり、Sambrook (前出、ch. 4, 8, 16 and 17) に記載されている。

10

【0058】

様々な宿主系を発現ベクターで形質転換することができる。以下に限定するものではないが、これらの中には組み換えバクテリオファージやプラスミド、またはコスミドDNA発現ベクターで形質転換された細菌と、酵母発現ベクターで形質転換された酵母と、バキュロウイルス発現ベクターで形質転換された昆虫細胞系と、ウイルスエレメント及び/または細菌エレメントを含む発現ベクターで形質転換された植物細胞系や動物細胞系が含まれる (Ausubel 前出、unit 16)。例えば、アデノウイルス転写/翻訳複合体を哺乳動物細胞に用いることができる。配列をウイルスのゲノムのE1若しくはE3領域に結合させた後、この感染ウイルスを用いて形質転換させ、宿主細胞でタンパク質を発現させることができる。また、ラウス肉腫ウイルスエンハンサーやSV40、またはEBV系のベクターを用いてタンパク質を高発現させることができる。

20

【0059】

核酸配列のルーチンのクローニング及びサブクローニング、増殖は、多機能PBLUES CRIP Tベクター (Stratagene, La Jolla CA) またはPSPO RT1プラスミド (Life Technologies) を用いて行うことができる。核酸配列をこれらのベクターの多数のクローニング部位に導入すると、lacZ遺伝子が破壊され、形質転換された細菌を確認するための比色法によるスクリーニングが可能となる。更に、これらのベクターは、クローニングされた配列におけるin vitroでの転写及びジデオキシ法によるシーケンシング、ヘルパーファージによる一本鎖の調整、入れ子状欠失の作製において有用である。

30

【0060】

長期に渡って組み換えタンパク質を産生させるために、同一或いは別のベクター上の選択マーカー遺伝子或いは可視マーカー遺伝子と共にこのベクターを持続的に細胞株に形質転換することができる。形質転換後、細胞を強化培地で約1~2日間増殖させてから選択培地に移す。代謝拮抗物質、抗生物質、除草剤耐性遺伝子を含む選択マーカーは、適正な選択薬剤に対する抵抗性を与え、導入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。アントシアニン、緑色蛍光タンパク質 (GFP)、グルクロニダーゼ、ルシフェラーゼなどの可視マーカーの発現によって同定された、或いは選択培地に生存することによって同定された耐性クローンを、培養技術を用いて増殖することができる。また、可視マーカーを用いて、導入された遺伝子によって発現するタンパク質を定量することができる。宿主細胞が目的の哺乳動物核酸分子を含むか否かの決定は、DNA-DNAまたはDNA-RNAハイブリダイゼーション、或いはPCR増幅技術に基づいて行うことができる。

40

【0061】

50

宿主細胞は、組み換えタンパク質を目的の形に修飾する能力に基づいて選択することができる。このような修飾には、アセチル化及びカルボキシル化、グリコシル化、リン酸化、脂質化、及びアシル化等が含まれる。「プレプロ」型を切断する翻訳後プロセッシングを利用して、タンパク質のターゲティング、折り畳み及び/または活性を特定することができる。翻訳後活性のための特定の細胞装置及び特徴的な機構を有する異なった宿主細胞 (Bethesda, MD) を選択して、組み換えタンパク質の正確な修飾及びプロセッシングを確実にすることが可能である。

【0062】

細胞培地からのタンパク質の回収

精製を容易にするために、ベクターに導入する異種部分は、グルタチオンSトランスフェラーゼ (GST)、カルモジュリン結合ペプチド (CBP)、6-His、FLAG、MYC等を含む。GST及びCBP、6-Hisはそれぞれ、グルタチオン及びカルモジュリン、金属キレート樹脂が結合した市販のアフィニティマトリックスを用いて精製される。FLAG及びMYCは、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て精製される。目的のタンパク質配列と異種部分との間にタンパク分解切断部位を設けて、生成の後の分離が容易にすることができる。組み換えタンパク質の発現及び精製の方法はAusubel (前出、unit 16) に記載され市販されている。

【0063】

ペプチドの化学合成

タンパク質若しくはその一部は、組み換え方法以外の当分野で周知の化学的方法によって合成することもできる。固相技術を用いるペプチド合成は、バッチ式或いは連続的なフロープロセスによって行うことができる。連続的なフロープロセスでは、アミノ保護及び側鎖保護アミノ酸残基をリンカーを介して不溶性の高分子支持物に連続的に追加する。メチルアミン誘導体化ポリエチレングリコールなどのリンカーを、ポリ(スチレン-co-ジビニルベンゼン)に結合させて支持レジンを形成する。このアミノ酸残基は、酸不安定Boc (t-butylloxycarbonyl) 法若しくは塩基不安定Fmoc (9-fluorenylmethoxycarbonyl) 法によって保護されたN-である。保護されたアミノ酸のカルボキシル基をリンカーのアミンに結合して、この残基を固相支持レジんに結合させる。Boc若しくはFmocを用いた場合、トリフルオロ酢酸若しくはピペリジンを用いて保護基を除去する。カップリング試薬若しくは予め活性化されたアミノ酸誘導体を用いて、追加する各アミノ酸を結合された残基に付加してから、レジンを洗浄する。完全長のペプチドは、連続的な保護の停止、即ち誘導体化アミノ酸を結合させて合成し、ジクロロメタン及び/またはN,N-ジメチルホルムアミドで洗浄する。このペプチドは、ペプチドカルボキシル末端とリンカーとの間で切断され、ペプチド酸またはペプチドアミドが作られる (Novabiochem 1997/98 Catalog and Peptide Synthesis Handbook, San Diego CA pp. S1-S20)。ABI 431A ペプチドシンセサイザー (Applied Biosystems) などの装置を用いて、ペプチドを自動合成することができる。タンパク質またはその一部は調整用の高性能液体クロマトグラフィーによって実質的に精製し、その組成をアミノ酸解析またはシーケンシングによって確認することができる (Creighton (1984) Proteins. Structures and Molecular Properties, WH Freeman, New York NY)。

【0064】

抗体の準備及びスクリーニング

ヤギ及びウサギ、ラット、マウス、ヒト等を含む様々な宿主は、ASP若しくはその任意の一部を注入して免疫することができる。フロイントなどのアジュバント及びミネラルゲルと、リゾレシチン及びpluronic polyol、ポリアニオン、ペプチド、油乳剤、キーホールリンペットヘモシニアン (KLH)、ジニトロフェノールなどの表面活性物質とを用いて免疫反応を高めることができる。オリゴペプチドやペプチド、またはタ

ンパク質の一部を用いて、少なくとも約5個のアミノ酸、より好ましくは10個の天然のタンパク質の一部と同一のアミノ酸を含む抗体を誘発させる。キメラ分子に対する抗体を産生させるために、オリゴヌクレオチドをK L Hなどのタンパク質と融合させることができる。

【0065】

モノクローナル抗体は、培地の連続細胞株によって抗体を産生させる任意の技術を用いて準備する。以下に限定するものではないが、このような技術には、ハイブリドーマ技術及びヒトB細胞ハイブリドーマ技術、EBV-ハイブリドーマ技術が含まれる(例えば、Kohler他(1975) *Nature* 256:495-497; Kozbor他(1985) *J. Immunol. Methods* 81:31-42; Cote他(1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole他(1984) *Mol. Cell Biol.* 62:109-120. を参照)。

10

【0066】

別法では、当分野で周知の方法を用いる上記した一本鎖抗体を生産する技術で、エピトープ特異的一本鎖抗体を生産する。本哺乳動物タンパク質のエピトープに対して特異的に結合する部位を含む抗体断片を生産することが可能である。限定するものではないが、このような断片には、例えば、抗体分子のペプシン消化によって作製されたF(ab')₂断片及びこのF(ab')₂断片のジスルフィド架橋を減少させて作製したFab断片が含まれる。別法では、Fab発現ライブラリを作製して、目的の特異性を有するモノクローナルFab断片の高速かつ容易に同定できるようにする(例えば、Huse他(1989) *Science* 246:1275-1281を参照)。

20

【0067】

本哺乳動物タンパク質若しくはその一部を用いて、ファージミドまたはBリンパ球免疫グロブリン・ライブラリをスクリーニングして、目的の特異性を有する抗体を同定する。確立された特異性を有するモノクローナル抗体或いはポリクローナル抗体のいずれか一方を用いる、競合的結合またはイムノアッセイの様々なプロトコルが当分野で周知である。このようなイムノアッセイは通常、このタンパク質とその特異的な抗体との複合体形成の測定を行う。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る2部位モノクローナル系イムノアッセイが好ましいが、競合的結合によるアッセイを用いることもできる(Pound(1998) *Immunochemical Protocols*. Humana Press, Totowa NJ)。

30

【0068】

アッセイのための分子の標識化

多様な標識化及び接合技術は当分野で周知であり、様々な核酸やアミノ酸、及び抗体のアッセイに用いることができる。標識した分子の合成は、³²P-dCTPまたはCy3-dCTP、Cy5-dCTPなどの標識したヌクレオチドや³⁵Sメチオニンなどのアミノ酸を組み込むためのAmersham Pharmacia BiotechキットまたはPromega(Madison WI)を用いて行うことができる。ヌクレオチド及びアミノ酸は、BIODIPYまたはFITC(Molecular Probes, Eugene OR)などの試薬を用いて分子中に存在するアミン及びチオール基または他の基に化学的に結合させることで、様々な物質(蛍光剤または化学発光剤、色素産生剤など)で直接標識することができる。

40

【0069】

(診断)

本核酸分子、断片、オリゴヌクレオチド、相補的なRNA及びDNA分子、PNAを用いて、遺伝子発現の変化やmRNAの過剰な発現の不在/存在を検出及び定量、または治療期間中のmRNAレベルのモニタリングを行うことができる。発現の変化に関連する症状や疾患には、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側索硬化

50

及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクロイツフェルト ヤコブ病、ゲルストマン症候群、Gerstmann-Straussler-Scheinker症候群を含むプリオン病 (prion disease) と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫 (cerebelloretinal hemangioblastomatosis)、脳3叉神経血管症候群、ダウン症を含む中枢神経系性の精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経障害、脊髄病、筋ジストロフィー及び他の神経筋障害、末梢神経疾患、皮膚筋炎及び多発性筋炎と、遺伝性、代謝性、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性精神障害、及び妄想性精神病と、季節性の感情の障害 (SAD)、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病、進行性核上性麻痺、皮質基底変性 (corticobasal degeneration)、家族性前頭側頭痴呆、プロラクチンの産生異常、卵管病及び排卵異常、子宮内膜症、発情期異常、月経周期異常、多嚢胞卵巣症候群、卵巣過剰刺激症候群、子宮内膜癌及び卵巣癌、子宮筋腫、自己免疫異常、異所性妊娠、奇形発生を含む不妊症、乳癌、線維嚢胞性乳腺症、乳漏症、精子形成異常、生理学上の精子異常、精巣癌、前立腺癌、良性の前立腺過形成、前立腺炎、ペーロニー病、インポテンス、男性乳房及び女性化乳房癌、嚥下障害、消化性食道炎、食道痙攣、食道狭窄、食道癌、消化不良、消化障害、胃炎、胃癌、食欲不振、悪心、嘔吐、胃不全麻痺、洞または幽門の浮腫、腹部アングナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリー ヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、及び後天性免疫不全症候群 (AIDS) 腸症、黄疸、肝性脳症、肝腎症候群、肝炎、肝脂肪症、血色素症、ウィルソン病、 α -1-アンチトリプシン欠損症、ライ症候群、原発性硬化性胆管炎、肝梗塞、門脈循環閉塞及び血栓、小葉中心壊死、肝臓紫斑病、肝静脈血栓、肝静脈閉塞症、子癇前症、子癇、妊娠性急性肝脂肪、妊娠性肝臓内胆汁うっ滞と、結節性再生及び腺腫、癌腫を含む肝癌とが含まれる。診断アッセイにハイブリダイゼーションまたは増幅技術を用いて、遺伝子発現の変化を検出するべく、患者からの生体サンプルの遺伝子発現レベルを標準的なサンプルの値と比較する。質的または量的なこのような比較法は当分野で周知である。

【0070】

例えば、本核酸分子またはプローブを標準的な方法で標識して、これをハイブリダイゼーション複合体の形成に好適な条件下で、患者からの生体サンプルに加える。インキュベーションの後、このサンプルを洗浄し、ハイブリダイゼーション複合体に関連する標識 (シグナル) の量を定量して標準値と比較する。患者のサンプルにおける標識の量が標準値と著しく異なっている場合は、関連する症状や疾患、または異常症の存在が示唆される。

【0071】

遺伝子発現に関連する症状や疾患、または異常症の診断のための基準を設けるために、正常或いは標準的な発現プロフィールを確立する。この発現プロフィールは、動物かヒトの正常な被験体から採取した生体サンプルを、ハイブリダイゼーションまたは増幅に好適な条件下で、プローブと結合させることによって確立することができる。標準的なハイブリダイゼーションは、正常な被験体から得た値と、実質的に生成された標的配列を所定量用いた実験値とを比較することによって定量することができる。このように求めた標準値を、特定の症状や疾患、または異常症を示す患者のサンプルから得た値と比較することができる。標準値と特定の症状に関連する値との偏差からその症状を診断する。

【0072】

またこのようなアッセイを用いて、動物実験や臨床検査における特定の治療計画の効果を評価したり、患者個人の治療をモニタリングすることができる。病態が確認されると治療プロトコルを開始し、通常ペースで診断アッセイを繰り返して、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否かを調べることが可能である。連続して行ったアッセイの結果から、数日から数ヶ月に渡る期間の治療効果を調べることができる。

【0073】

免疫学的方法

特異的なポリクローナル抗体若しくはモノクローナル抗体の何れかを用いるタンパク質の検出及び定量は当分野で周知である。このような技術には、ELISA（酵素結合免疫吸着検定法）及びラジオイムノアッセイ（RIA）、蛍光活性化セルソーター法（FACS）が含まれる。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る2部位モノクローナル系イムノアッセイが好ましいが、競合的結合アッセイを用いることもできる（例えば、Coligan 他（1997）*Current Protocols in Immunology*, Wiley-Interscience, New York NY; and Pound を参照）。

【0074】

（治療）

SEQ ID NO: 46、SEQ ID NO: 47、その他のASPsであるヒトガストリクシン（g1658286）、および *Neurospora crassa* 空胞プロテアーゼA（*Neurospora crassa vacuolar protease A*）（g1039445）のそれぞれの所定の領域間に、例えば配列及びモチーフの文脈において、化学的及び構造的類似性が存在する。更に、遺伝子の発現は、消化系、生殖組織、および脳組織に密接に関連し、アルツハイマー病およびダウン症などの症状にある役割を果たすと考えられる。発現若しくは活性の増大に関連する症状の治療においては、発現またはタンパク質の活性を低下させることが望ましい。また、発現または活性の低下に関連する症状の治療においては、発現またはタンパク質の活性を増大させることが望ましい。

【0075】

核酸分子変異体であるSEQ ID NO: 1-42およびSEQ ID NO: 46および47の所定の領域間に、配列の文脈における化学的および構造的な類似性が存在する。更に、核酸分子変異体の遺伝子発現が、消化系組織、生殖組織、脳組織、および心血管組織に密接に関連し、アルツハイマー病およびダウン症などの症状にある役割を果たすと考えられる。

【0076】

一実施例において、本哺乳動物タンパク質の発現または活性の変化に関連する症状の治療または予防のために、本哺乳動物タンパク質、その一部、またはその誘導体を患者に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、癲癇、虚血性脳血管障害、脳卒中、大脳新生物、アルツハイマー病、ピック病、ハンチントン病、痴呆、パーキンソン病及びその他の錐体外路障害、筋萎縮性側策硬化及びその他の運動ニューロン障害、進行性神経性筋萎縮症、色素性網膜炎、遺伝性運動失調、多発性硬化症及び他の脱髄疾患、細菌性及びウイルス性髄膜炎、脳膿瘍、硬膜下蓄膿症、硬膜外膿瘍、化膿性頭蓋内血栓性静脈炎、脊髄炎及び神経根炎、ウイルス性中枢神経系疾患と、クールー及びクロイツフェルト ヤコブ病、ゲルストマン症候群、Gerstmann-Straussler-Scheinker 症候群を含むプリオン病（prion disease）と、致死性家族性不眠症、神経系性栄養病及び代謝病、神経線維腫症、結節硬化症、小脳網膜血管芽腫（cerebelloretinal hemangioblastomatosis）、脳3叉神経血管症候群、ダウン症を含む中枢神経系性の精神薄弱及び他の発生障害、脳性麻痺、神経骨格異常症、自律神経系障害、脳神経障害、脊髄病、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

10

20

30

40

50

、内分泌性、及び中毒性ミオパシーと、重症筋無力症、周期性四肢麻痺と、気分性及び不安性精神障害、及び妄想性精神病と、季節性の感情の障害（SAD）、静座不能、健忘症、緊張病、糖尿病性ニューロパシー、錐体外路性終末欠陥症候群、ジストニー、分裂病性精神障害、帯状疱疹後神経痛、及びトゥレット病、進行性核上性麻痺、皮質基底変性（corticobasal degeneration）、家族性前頭側頭痴呆、プロラクチンの産生異常、卵管病及び排卵異常、子宮内膜症、発情期異常、月経周期異常、多嚢胞卵巣症候群、卵巣過剰刺激症候群、子宮内膜癌及び卵巣癌、子宮筋腫、自己免疫異常、異所性妊娠、奇形発生を含む不妊症、乳癌、線維嚢胞性乳腺症、乳漏症、精子形成異常、生理学上の精子異常、精巣癌、前立腺癌、良性の前立腺過形成、前立腺炎、ペーロニー病、インポテンス、男性乳房及び女性化乳房癌、嚥下障害、消化性食道炎、食道痙攣、食道狭窄、食道癌、消化不良、消化障害、胃炎、胃癌、食欲不振、悪心、嘔吐、胃不全麻痺、洞または幽門の浮腫、腹部アンギナ、胸焼け、胃腸炎、イレウス、腸管感染、消化性潰瘍、胆石症、胆嚢炎、胆汁うっ滞、膵臓炎、膵臓癌、胆道疾患、肝炎、高ビリルビン血症、硬変症、肝臓の受動性うっ血、ヘパトーム、感染性大腸炎、潰瘍性大腸炎、潰瘍性直腸炎、クローン病、ホイップル病、マロリー ヴァイス症候群、結腸癌、結腸閉塞、過敏性腸症候群、短小腸症候群、下痢、便秘、胃腸出血、及び後天性免疫不全症候群（AIDS）腸症、黄疸、肝性脳症、肝腎症候群、肝炎、肝脂肪症、血色素症、ウィルソン病、
- 1
- アンチトリプシン欠損症、ライ症候群、原発性硬化性胆管炎、肝梗塞、門脈循環閉塞及び血栓、小葉中心壊死、肝臓紫斑病、肝静脈血栓、肝静脈閉塞症、子癇前症、子癇、妊娠性急性肝脂肪、妊娠性肝臓内胆汁うっ滞と、結節性再生及び腺腫、癌腫を含む肝癌とが含まれる。 10 20

【0077】

別の実施例では、医薬用担体と共に実質的に精製された哺乳動物タンパク質を含む医薬組成物を患者に投与して、限定するものではないが上記した疾患を含むこの内在性タンパク質の活性、または発現の変化に関連する症状の治療または予防を行うことが可能である。

【0078】

更なる実施例では、この哺乳動物タンパク質の活性を調節するリガンドを患者に投与して、限定するものではないが上記した疾患を含むこのタンパク質の寿命や発現、または活性の変化に関連する症状の治療または予防を行うことが可能である。一実施態様では、この哺乳動物タンパク質に特異的に結合する抗体を、この哺乳動物タンパク質を発現する組織や細胞に薬剤を送達するためのターゲッティング或いは送達機構として用いることもできる。 30

【0079】

更なる実施例では、この哺乳動物タンパク質またはその一部や誘導体を発現可能なベクターを患者に投与して、限定するものではないが上記した疾患を含むこのタンパク質の寿命や発現、または活性の変化に関連する症状の治療または予防を行うことが可能である。

【0080】

更なる実施例では、この核酸分子やその断片の相補配列を発現するベクターを患者に投与して、限定するものではないが上記した疾患を含むこのタンパク質の寿命や発現、または活性の変化に関連する症状の治療または予防を行うことが可能である。 40

【0081】

本核酸分子、またはそれに相補的な分子やその一部、そのタンパク質やその一部の内の任意のもの、これらの核酸分子やタンパク質を運ぶベクター、及びこれらのリガンドをその他の薬剤と共に投与することが可能である。併用療法に用いる薬剤の選択は、当業者が従来薬学原理に従って行うことができる。薬剤を併用することによって、少量の各薬剤で特定の症状の予防または治療において相乗的な効果をあげることが可能である。

【0082】

核酸を用いる遺伝子発現の調節

遺伝子の発現は、哺乳動物遺伝子の5'または3'調節領域、或いは他の調節領域に対して相補的或いはアンチセンス分子を設計することで調節することが可能である。転写開始 50

部位に対して設計されたオリゴヌクレオチドが好ましい。同様に、ポリメラーゼ、転写因子、または調節分子の結合を阻止する三重螺旋塩基対合で遺伝子発現を阻止することができる (Gee 他 In: Huber and Carr (1994) Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177)。また、リボソームと mRNA との結合を阻止して翻訳が行われなないように、相補的な分子を設計することも可能である。或るいは、核酸分子またはその断片のライブラリをスクリーニングして、翻訳されない調節配列に特異的に結合する核酸分子または断片を同定することも可能である。

【0083】

また、酵素活性をもつ RNA 分子であるリボザイムを用いて、RNA の特異的な切断を触媒してもよい。リボザイム作用のメカニズムは、まずリボザイム分子と相補的な標的 RNA との配列特異的なハイブリダイゼーションが起こり、次に GUA および GUU、GUC などの部位においてヌクレオチド鎖が切断される。このような部位が一旦同定されたら、オリゴヌクレオチドを機能不全に示す二次構造特性について、同じ配列を有するオリゴヌクレオチドを評価することができる。また、候補標的としての適合性は、RNA 分解酵素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を検査して評価することができる。

【0084】

本発明の相補的な核酸およびリボザイムは、固相ホスホラミダイト化学合成法を用いて、in vitro または in vivo での組換え発現によって調製することが可能である。更に、RNA 分子は、その 5' および / または 3' 末端に隣接配列を付加して、或い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-メチルを用いて、細胞内の安定性および半減期が増大するように改変することができる。この改変は PNA の作製に固有であるが、他の核酸分子にも適用することができる。例えばイノシン、queosine、wybutosine などの伝統的でない塩基を含めて、またはアセチル基、メチル基、チオ基でウリジン、アデニン、シチジン、グアニン、およびチミンを修飾して、内在性エンドヌクレアーゼに対する分子の有効性を低くする。

【0085】

スクリーニングアッセイ

本哺乳動物タンパク質をコードする核酸分子を用いて、分子または化合物のライブラリをスクリーニングして特異的な結合親和性を調べることが可能である。このライブラリは、生物系の核酸分子の活性、複製、転写、または翻訳を調節する DNA 分子、RNA 分子、PNA、ペプチド、転写因子などのタンパク質、エンハンサー、リプレッサー、およびその他のリガンドを含み得る。このアッセイは、特異的な結合が許容される条件下で、哺乳動物核酸分子またはその断片を分子のライブラリと結合させるステップと、特異的な結合を検出して、その核酸分子と特異的に結合する少なくとも 1 つの分子を同定するステップとを含む。

【0086】

同様に、本哺乳動物タンパク質またはその一部を用いて、任意の様々なスクリーニングアッセイで分子または化合物のライブラリをスクリーニングすることができる。このようなスクリーニングに用いるタンパク質の一部は、溶液中に遊離している、非生物的若しくは生物学的な基板に固定されている、または細胞内に位置していてもよい。そのタンパク質と分子との特異的な結合を測定することが可能である。スクリーニングするライブラリの種類によって、アッセイでそのタンパク質に特異的に結合する DNA、RNA、PNA 分子、アゴニスト、アンタゴニスト、抗体、免疫グロブリン、インヒビター、ペプチド、タンパク質、薬剤、または他の任意のリガンドを同定することが可能である。極めて少ないアッセイ容量および極少量の試験化合物を用いるハイスループットのスクリーニング方法が、米国特許第 5,876,946 号に記載されており、言及することを以って本明細書の

10

20

30

40

50

一部とする。この方法は、酵素阻害または受容体結合を調べるために多数の分子をスクリーニングする。

【0087】

リガンドの精製

本核酸分子またはその断片を用いてサンプルからリガンドを精製することが可能である。哺乳動物核酸分子またはその断片を用いてリガンドを精製する方法は、特異的な結合が許容される条件下でその核酸分子またはその断片とサンプルとを結合させるステップと、特異的な結合を検出するステップと、結合したタンパク質を回収するステップと、好適な試薬を用いて精製されたリガンドから核酸分子を分離するステップとを含む。

【0088】

同様に、本タンパク質またはその一部を用いてサンプルからリガンドを精製することが可能である。哺乳動物タンパク質またはその一部を用いてリガンドを精製する方法は、特異的な結合が許容される条件下でそのタンパク質またはその一部とサンプルとを結合させるステップと、そのタンパク質をリガンドとの間の特異的な結合を検出するステップと、結合したタンパク質を回収するステップと、好適なカオトロピック剤を用いて精製されたりリガンドからタンパク質を分離するステップとを含む。

【0089】

薬理学

医薬組成物とは、所望の目的を達成するのに効果的な量の活性成分を含んでいる物質である。効果的な薬用量の決定は、当分野の技術者の能力による部分が多い。どんな化合物であっても、初めは細胞培養アッセイ或いは動物モデルの何れかによって治療効果のある薬用量を推定する。また、動物モデルを使って、好適な濃度範囲および投与経路を決定する。次に、このような情報を用いて、ヒトへの効果的な投与経路および薬用量を決定する。

【0090】

治療効果のある薬用量とは、症状または病態を改善するタンパク質またはインヒビターの量である。このような薬剤の薬用効果および毒性は、例えば、ED₅₀（集団の50%に医薬的効果がある投与量）およびLD₅₀（集団の50%に致命的である投与量）などの細胞培養または実験動物における標準的な製薬方法によって決定することができる。或る投与量における毒性効果と治療効果との比率が治療指数となり、LD₅₀/ED₅₀と示すことができる。高い治療指数を示す医薬組成物が好ましい。細胞培養アッセイおよび動物実験から得られたデータを用いて、ヒトへ適用する薬用量の範囲を決定する。

【0091】

モデル系

動物モデルを用いて、ヒトに相当する暴露条件で動物モデルがヒトに類似の毒物反応を示すバイオアッセイを行うことができる。哺乳動物が最も一般的なモデルである。大抵の毒物研究は、低コストで入手が容易であり、かつ参照できる中毒学が豊富であることから、ラットやマウスなどの齧歯類を用いて行われる。齧歯類近交系は、目的の遺伝子の過剰或いは過少な発現の生理学的な原因を調査するのに有用なモデルであり、疾患の診断および治療方法の開発にも有用である。特定の遺伝子を大量に発現する（例えば、乳汁中に分泌される）同系哺乳動物は、その遺伝子によって発現されるタンパク質の便利な供給源となり得る。

【0092】

中毒学

中毒学とは、生物系における物質の影響を研究する学問である。殆どの毒物研究はラットまたはマウスを用いて、物質がヒトの健康に与える影響を推定する。生理機能及び行動、恒常性プロセス、致死率における質的及び量的変化を観察して、毒性プロファイルを作成し、その物質に曝露された後のヒトの健康状態を評価する。

【0093】

遺伝子毒物学は、物質が遺伝子の突然変異を引き起こす能力を同定し分析する。遺伝毒性

10

20

30

40

50

物質は通常、核酸との相互作用を促進する共通の化学的或いは物理的特性を有し、突然変異した染色体が子孫に受け継がれるのが最大の害である。毒物研究によって、受胎前の両親のどちらか一方、または妊娠中の母、発生段階の生物に投与された場合の、子孫における構造的或いは機能的な異常の頻度を増加させる物質を同定することが可能である。マウス及びラットは生殖周期が短く、統計的必要性を満たす多数の子を出産する能力から、これらの試験にはマウス及びラットが用いられる場合が最も多い。

【0094】

急性毒物の検査は、被験体への物質の一回の投与に基づき、その物質による症状または致死率を決定する。この検査では、1) 初めの投与量の範囲を決定する実験と、2) 有効な投与量の範囲を狭める実験と、3) 用量応答曲線を確立する実験の3つの実験が行われる。

10

【0095】

長期に渡る毒性検査では繰り返し物質を投与する。このような検査には、ラットやイヌが一般的に用いられ、分類学上異なった種からデータを収集する。物質を高い投与濃度で3~4ヶ月間、毎日投与することで、発癌を除く、成体動物における殆どの毒性の種類が明らかになるという研究結果が多数報告されている。

【0096】

一年或いはそれ以上の長期に渡る慢性毒性検査は、物質に毒性がないこと、或いは物質の発癌の可能性の何れかを実証するために行われる。ラットで検査が行われる場合、少なくとも3つの検査グループと1つの対照グループが用いられ、最初から最後までである間隔で検査及びモニタリングが行われる。

20

【0097】

遺伝子組換え動物モデル

目的の遺伝子を過剰或いは過小に発現する遺伝子組換え齧歯類を同系交配し、それを用いてヒト疾患モデルを作製したり、治療薬検査や毒物検査を行う(例えば、米国特許第4,736,866号、同第5,175,383号、および同第5,767,337号を参照)。場合によっては、導入遺伝子が、胚発生中若しくは出生後の特定の時期に特定の種類の組織で活性化され得る。導入遺伝子の発現は、実験的薬剤治療を施す前、その最中、またはその後の、遺伝子組換え動物における表現型や組織特異的なmRNAの発現の分析からモニタリングすることができる。

30

【0098】

胚性幹細胞

齧歯類胚から単離された胚性幹細胞(ES細胞)は、胚を形成する可能性を維持している。ES細胞がキャリアとなる胚の中に導入されると、正常な発生が再開され、生まれる動物の全ての組織の一部を担うことになる。ES細胞は、実験用のノックアウトおよびノックイン齧歯類系を作製するのに好適な細胞である。マウス129/SvJ細胞株などのマウスES細胞は、マウスの初期胚から採取されてから当分野で周知の培養条件下で増殖されたものである。ノックアウト系に用いるベクターには、マーカー遺伝子配列を含むように破壊された疾患候補遺伝子が含まれる。このマーカー遺伝子配列は、i n v i v oでその転写および/または翻訳を阻害する。このベクターが、当分野で周知の電気穿孔法およびリポソーム法、マイクロインジェクション等の形質転換方法によってES細胞に導入される。齧歯類の内在性遺伝子が、細胞分裂の際の相同組換えや組み込みによって、破壊された疾患遺伝子に置換される。

40

形質転換されたES細胞が、所定の条件下で選択されて同定され、C57BL/6マウス系などからのマウス細胞胚盤胞に微量注入されるのが好ましい。この胚盤胞を複数の偽妊娠メスに外科的に導入すると、生まれるキメラ子孫がそれぞれ、その遺伝子型を有し、それらを交配してヘテロ接合系またはホモ接合系を作り出す。

【0099】

また、i n v i t r oでの神経細胞および造血系、心筋細胞などの様々な細胞型や組織

50

の分化の研究に、ES細胞が用いられている (Bain 他 (1995) Dev. Biol. 168:342-357; Wiles and Keller (1991) Development 111:259-267; and Klug 他 (1996) J. Clin. Invest. 98:216-224)。近年の研究により、ヒト胚盤胞由来ES細胞を *in vitro* で操作して、内胚葉および中胚葉、外胚葉性の細胞型を含む8つの別の細胞系譜に分化可能であることが実証された (Thomsonら (1998) Science 282:1145-1147)。

【0100】

ノックアウト分析

遺伝子ノックアウト分析では、ヒト疾患遺伝子候補のある領域が、ネオマイシン・ホスホトランスフェラーゼ遺伝子などの非哺乳動物遺伝子を含むように酵素によって改変される (neo; Capecchi (1989) Science 244:1288-1292)。挿入されたコード配列は、標的遺伝子の転写および翻訳を阻害して疾患候補タンパク質の生化学的な合成を阻止する。この改変した遺伝子を培養胚性幹細胞 (上記) に形質転換し、この形質転換細胞を齧歯類胞胚に注入し、この胞胚を偽妊娠メスに移植する。この遺伝子組換え子孫をクロス交配して、ホモ接合近交系を作り出す。

【0101】

ノックイン分析

胚発達の初期段階に現われる分化全能性ES細胞を用いてノックインヒト化動物 (ブタ) または遺伝子組換えヒト疾患動物モデル (マウスまたはラット) を作り出すことができる。ノックイン技術を用いて、ヒト遺伝子のある領域を動物ES細胞に注入し、そのヒト配列が組換えによって動物細胞ゲノムの中に組み込まれるようにする。ヒト遺伝子が組み込まれた全能性ES細胞を上記したように操作する。ヒトの症状に類似の情報を収集するべく、この近交系動物の研究および処置を行う。これらの方法を用いて、幾つかのヒト疾患モデルを作り出す (例えば、Lee 他 (1998) Proc. Natl. Acad. Sci. 95:11371-11376; Baudoin 他 (1998) Genes Dev. 12:1202-1216; and Zhuang 他 (1998) Mol. Cell Biol. 18:3340-3349)。

【0102】

非ヒト霊長類モデル

動物実験の分野は、生理学および遺伝学、化学、薬理学、統計学等の基本科学から得たデータおよび方法論を利用する。このようなデータは、非ヒト霊長類がヒトの健康に関連し得るため、非ヒト霊長類における治療薬の効果の評価に重要である。ヒトの代わりにサルをワクチンおよび薬剤の評価に用いる。サルの反応は、類似の条件下でのヒトへの暴露に相当する。カニクイザル (Macaca fascicularis、Macaca mulatta)、および一般的なマーモセット (Callithrix jacchus) は、このような実験に用いられる最も一般的な非ヒト霊長類 (NHP) である。NHPの群体作りおよび維持には多大な費用が掛かるため、初期の研究および毒物学的検査は、通常は齧歯類モデルで行われる。薬物嗜癖などの行動を調べる研究では、NHPは最良の実験動物である。更に、個々のNHPおよびヒトは、多くの薬剤および毒物に対して異なった感受性を示すため、これらの物質を「高代謝型」および「低代謝型」に分類することができる。

【0103】

更なる実施例では、限定するものではないが、トリプレット遺伝子コードおよび特異的な塩基対相互作用などの特性を含む、現在知られている核酸配列の特性に新しい技術が依存する場合は、本哺乳動物タンパク質をコードする核酸分子を、開発中のあらゆる分子生物学技術に用いることが可能である。

【0104】

【実施例】

本発明は、記載した特定の装置及び物質、方法に限定されるものではないことを理解され

たい。特定の実施例について説明するが、同等の実施例を用いて本発明を具現することも可能である。本発明の範囲は、前記請求の範囲によってのみ限定されるものであって、記載した実施例によって限定されるものではない。また、以下に記載の実施例は、本発明を例示するためのものであって本発明を限定するものではない。例示目的で、ヒト脳 cDNA ライブラリ BRAINON01 の作製方法を記載する。

【0105】

1 代表的な cDNA 配列の調整

BRAINON01 標準化 cDNA ライブラリは、26歳の白人男性が脳の右前頭頭頂部にグレード4のオリゴ星状細胞腫 (oligoastrocytoma) があると診断された後、大脳髄膜の切除術の際に採取された癌性脳組織 (標本番号0003) から作製された。患者は手術の前に、片側麻痺、てんかん、眼瞼の下垂症、および一般的な片頭痛の症状があった。

【0106】

凍結組織を POLYTRON ホモジナイザー (PT-3000; Brinkmann Instruments, Westbury, NY) を用いて、グアニジニウムイソチオシアネート溶液中でホモジナイズして溶解した。この溶解物を RNA 単離プロトコル (Stratagene) に従って酸性フェノール (pH 7.4) で抽出した。この RNA を等倍量の酸性フェノールで抽出し、0.3 M 酢酸ナトリウムおよび 2.5 倍量のエタノールで沈殿させ、DEPC 処理水で再懸濁し、37°C で 25 分間 DNアーゼで処置した。この RNA の抽出および沈殿を初回と同様の方法で行った。この mRNA を OLIGOTEX キット (Qiagen, Chatsworth CA) で単離し、cDNA ライブラリを作製した。

【0107】

この mRNA を、SUPERSCRIPT プラスミドシステム (Life Technologies) を用いて推奨プロトコルに従って処置した。cDNA の合成は、NotI-オリゴ d(T) プライマーを用いて開始した。二本鎖 cDNA を平滑化して SalI アダプターに結合し、NotI で消化してから、SEPHAROSE CL4B カラム (Amersham Pharmacia Biotech) 上で分画し、400 bp を超える cDNA は、PSPORT I プラスミド (Life Technologies) の NotI および SalI 部位に結合させた。この組み換えプラスミドを DH5 コンピテント細胞 (Life Technologies) に形質転換した。

【0108】

2 cDNA ライブラリの標準化

大腸菌系 DH12S コンピテント細胞 (Life Technologies) を電気穿孔法で形質転換した後、カルベニシリン (25 mg/L) およびメチシリン (1 mg/ml) を用いた選択培養液で、BRAINOT01 プラスミドライブラリの独立した 4.9×10^6 個の独立したクローンを増殖させた。ライブラリの存在量に応じて過剰な cDNA の複製を減らすために、次の変更点を除いて Soares ら (1994, Proc Natl Acad Sci 91: 9928-9932) の手順に従って 1 回標準化した。その変更点は、(1) プライマー伸長反応におけるプライマー: 鋳型の割合を 2:1 から 10:1, 2 に下げ、(2) 400 ~ 1000ヌクレオチドの長いプライマー伸長生成物が生成されるように、反応液の各 ddNTP の濃度をそれぞれ 150 μ M に下げ、(3) ハイブリダイゼーションのアニーリング時間を 13 時間から 48 時間に延長した。標準化されたライブラリの一本鎖環状 DNA をヒドロキシアパタイトクロマトグラフィーで精製し、ランダムプライミングにより部分的な二本鎖にしてから、この cDNA を電気穿孔法で DH10B コンピテント細胞 (Life Technologies) に形質転換した。

【0109】

プラスミド DNA を細胞から遊離させ、REAL Prep 96 プラスミドキット (Qiagen) を用いて精製した。以下の変更点を除いて推奨プロトコルに従った。その変

更点は、(1)細菌を、25 mg/Lのカルベニシリンと0.4%のグリセロールを含む1 mlの滅菌TERRIFIC BROTH (Life Technologies)で培養し、(2)接種した後、19時間インキュベートし、インキュベーションの最後に0.3 mlの溶解緩衝液で溶解し、(3)イソプロパノール沈殿の後、プラスミドDNAペレットを0.1 mlの蒸留水に懸濁した。プロトコルのステップを終了した後、サンプルを96ウェルブロックに移し4 で保管した。

【0110】

このcDNAは、MICROLAB 2200 (Hamilton, Reno NV)と共にDNA ENGINEサーマルサイクラー (MJ Research, Watertown, MA)を用いて調製した。cDNAは、ABI PRISM 377 DNAシーケンシングシステム (Applied Biosystems)を用いて、SangerおよびCoulson (1975; J Mol Biol 94: 441f)の方法によってシーケンシングし、リーディングフレームを決定した。

10

【0111】

3 同定及び伸長、構築、解析

コンセンサス配列 (SEQ ID NO: 43) は、ヒトcDNA配列のZOOSEQデータベース (Incyte Genomics) のインサイトクローン番号: 1863920X321F1 (ライブラリ名: PROSNOT19)、2233901H1 (PANC TUT02)、1611565T1 (COLNTUT06)、336541H1 (EOSIHET02)、およびショットガン配列SBAA03902F1から組み立て、これを用いて更なるASP関連配列を同定するためにLIFESEQデータベースおよびZOOSEQデータベース (Incyte Genomics) を検索した。MACDNA SIS PROソフトウェア (Hitachi Software Engineering) を用いたSEQ ID NO: 43及びSEQ ID NO: 44の翻訳によって、翻訳領域であるSEQ ID NO: 46及びSEQ ID NO: 47を解明した。この核酸配列及びアミノ酸配列を、BLASTを用いてGenBankデータベース及びSwiss Prot、BLOCKS、PRINTS、Prosite、PFAMのデータベースに対して問い合わせた。機能解析は、MOTIFS及びHMMアルゴリズムを用いて行い、抗原性指数 (Jameson-Wolf解析) はLASERGENEソフトウェア (DNASTAR) を用いて決定した。

20

30

【0112】

4 配列類似性

配列類似性は、LASERGENEソフトウェアのMEGALIGNプログラム (DNASTAR) のクラスター法を用いて、少なくとも2つの核酸分子間或いはアミノ酸配列間の比較に基づいて同一性 (%) として計算した。このクラスター法は、全ての対間の距離を調べて配列をクラスターに分類するアルゴリズムを用いる。クラスターを対に並べた後、再びグループに分ける。配列A及びBの2つの配列間の配列類似性は、配列AとBの一致する残基の合計数を、配列Aの合計残基数から配列Aのギャップ残基数と配列Bのギャップ残基数とを減じたもので除して求める。2つの配列間の類似性が0或いは0に近いところのギャップは含めていない。

40

【0113】

5 核酸変異体の同定

哺乳動物アスパルチルプロテアーゼをコードする核酸分子 (SEQ ID NO: 43及び44) のスプライスバリエーションである核酸分子をBLAST若しくはBLAST2 (Altschul, 前出; デフォルトパラメータを備えたNCI-BLASTN version 2.0.4) を用いて同定し、SEQ ID NO: 43及び44とアラインメントするクローンをLIFESEQデータベースまたはZOOSEQデータベース (Incyte Genomics) において見出した。哺乳動物核酸分子変異体は、BLASTスコアによって選択した。BLASTスコアは、核酸の一致が高スコア対 (HSP) の各対には+5のスコアを与え、不一致の対には-4のスコアを与えて計算した。BLAS

50

Tアラインメントは視覚的に検査でき、BLASTスコアが100を超えるクローンは、Phrap (Green, 前出)を用いて互いにアラインメントした。タンパク質をコードする核酸配列のsprayバリエーションのspray部位に隣接する核酸配列の例が当分野で周知である。sprayバリエーション核酸分子として同定されたものは、核酸配列の或る部分が100%の同一性であって、その核酸配列の残りの部分が100%未満の同一性であり、かつ推定spray部位に対する開始ヌクレオチド3'がグアニン塩基である。パブリックドメインにおけるESTsの類似の分析により、ヒトアスパルチルプロテアーゼの核酸配列とアラインメントするクローンが同定され、推定spray部位の位置が決定された。

【0114】

10

6 ノーザン分析

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る実験用技術であり、標識したプローブを特定の細胞型或いは組織から抽出したRNAを固定した膜にハイブリダイズさせることが含まれる。

【0115】

BLASTに用いられる類似性コンピュータ技術を用いて、LIFESEQデータベースまたはZOOSEQデータベース(Incyte Genomics)などのヌクレオチドデータベースにおいて、同一或いは関連する分子を検索する。この配列に基づいた解析は、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調節して、任意の特定の一致を、同一または類似と分類することができる。BLAST検索の基準は、(%配列同一性 × %最大BLASTスコア) / 100と定義される積スコアである。積スコアは、2つの配列間の類似度及び配列一致の長さの両方を考慮する。例えば、積スコアが40の場合、その一致は1~2%誤差の範囲内で正確であり、70以上であればその一致は正確であろう。類似する分子或いは関連する分子は通常、8~40の範囲の積スコアを示す分子を選択することにより同定される。

20

【0116】

ノーザン分析の結果は、この哺乳動物タンパク質をコードする転写物が発生したライブラリの分布割合(パーセント)として報告される。分析には、器官/組織及び疾患によるcDNAライブラリの分類も含まれる。器官/組織のカテゴリーには、心血管、皮膚、発生、内分泌、胃腸、造血/免疫、筋骨格、神経、生殖、泌尿器が含まれる。疾患のカテゴリーには、癌、炎症/外傷、細胞増殖、神経が含まれる。カテゴリー別に、目的の配列を発現するライブラリ数をカウントし、この数をこの哺乳動物タンパク質をコードする核酸分子の核酸分子変異体を含むと同定された全てのライブラリ数で除した。

30

【0117】

7 核酸分子の伸長

SEQ ID NO: 43 SEQ ID NO: 44 築に用いた核酸分子の少なくとも1つは、オリゴヌクレオチドプライマーを用いてインサイト社cDNAクローン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)を用いて、約22~約30個のヌクレオチドの長さ、約50%以上のGC含量で、約55~68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る断片は排除した。選択されたヒトcDNAライブラリを用いてこの分子を伸長した。2段階以上の伸長が必要な場合は、追加の或いは組みになった入れ子のプライマー(nested primer)を設計する。

40

【0118】

DNA ENGINE サーマルサイクラー(MJ Research)を用いて96ウェルプレートでPCR法を実施し、高い忠実度の増幅を達成した。反応混合液には、鋳型DNA及び200 nmolの各プライマー、反応緩衝液(Mg²⁺、(NH₄)₂SO₄、β-メルカプトエタノールを含む)、Taq DNAポリメラーゼ(Amersham

50

Pharmacia Biotech)、E LONGASE 酵素 (Life Technologies)、Pfu DNAポリメラーゼ (Stratagene) を含まれている。プラスミドから選択したプライマーの組みに対して以下のパラメーターで増幅を行った。

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

10

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

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。

20

【0119】

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); Molecular Probes)を不透明な蛍光光度計プレート(Corning Science Products, Corning NY)の各ウェルに分注して、DNAがその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、サンプルの蛍光を測定して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電気泳動させて解析し、何れの反応がより長い配列の伸長に成功したかを決定する。

30

【0120】

伸長した配列を脱塩及び濃縮してから384ウェルプレートに移し、CviJIコレラウイルスエンドヌクレアーゼ(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切断した。ショットガンシーケンシングのために、消化した断片を0.6~0.8%の濃度のアガロースゲル上に分離させ、UV光で視認できるように断片を切断し、寒天をAGARACE (Promega)で消化/除去した。T4 DNAリガーゼ(New England Biolabs, Beverly MA)を用いて伸長した断片を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から、コンピテント大腸菌細胞に形質転換した。形質転換細胞が抗生物質を含む培地で選択され、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入った384ウェルプレートの中で、37 で一晩培養した。

40

【0121】

この細胞を溶解して、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を増幅した。

ステップ 1 94 で3分間
 ステップ 2 94 で15秒

50

- 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(dimethyl sulphoxide)(1:2, v/v)で希釈し、DYENAMICエネルギー移動シーケンシングプライマー及び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。

【0122】

同様に上述の手順で、SEQ ID NO: 1及びSEQ ID NO: 10のヌクレオチド配列を利用し、外側に伸長するように設計したオリゴヌクレオチドとゲノムDNAライブラリを用いて調節配列を得た。

【0123】

8 プローブの標識化及びハイブリダイゼーション分析

以下に示す任意の方法によって、核酸を生体サンプルから単離し、標準的なハイブリダイゼーションプロトコル用の基板にその単離した核酸配列を固定する。標的核酸(ゲノムDNAの制限酵素断片)の混合液を、1x TAE(Tris-アセテート-エチレンジアミン4酢酸(EDTA)を含む)緩衝液において、0.7%アガロースゲルで電気泳動して分画し、20xクエン酸ナトリウム(SSC)を用いて毛管輸送によってナイロン膜に移す。別法では、標的核酸を個別にベクターに結合させ、細菌宿主細胞に挿入してライブラリを作製する。標的核酸を以下の任意の方法によって基板上に並べる。第1の方法では、個々のクローンを含む細菌細胞を機械的に選んでナイロン膜上に並べる。細菌増殖培地(カルベニシリンを含むLB寒天)にこのナイロン膜を置き、37 で16時間インキュベートする。細菌コロニーをプロテインキナーゼKで変性及び中和、消化する。STRATALINKER UVクロスリンカー(Stratagene, La Jolla CA)において、ナイロン膜をUV照射してDNAをその膜に架橋結合させる。

【0124】

第2の方法では、挿入物に近接するベクター配列に相補的なプライマーを用いてPCRのサイクルを30回繰り返し、標的核酸を細菌ベクターから増幅する。増幅した標的核酸をSEPHACRYL-400ビーズ(Amersham Pharmacia Biotech)を用いて精製する。精製した標的核酸をガラス製の顕微鏡スライド(Corning Science Products, Corning NY)上に機械的に並べる。このスライドは、0.05%のアミノプロピルシラン(Sigma-Aldrich, St Louis MO)で予めコーティングしてから110 で硬化したものである。ガラススライドのアレイ(マイクロアレイ)を、STRATALINKER UVクロスリンカー(Stratagene)においてUV照射する。

【0125】

cDNAプローブは、鋳型mRNAから作製する。5 µgのmRNAを1 µgのランダムプライマー(Life Technologies)と混合し、70 で10分間インキュベートしてから凍結乾燥する。この凍結乾燥したサンプルを、dNTP混合物及び[³²P]dCTP、ジチオスレイトール、MMLV逆転写酵素(Stratagene)を含む50 µlの1x第1鎖緩衝液(cDNA Synthesis system; Life Technologies)に再懸濁し、42 で1~2時間インキュベートする。インキュベートした後、プローブを42 µlの蒸留水で希釈して95 で3分間加熱し、その後氷上で冷却する。プローブに含まれるmRNAをアルカリ分解によって除去

する。このプローブを中和し、PROBEQUANT G-50マイクロカラム (Amersham Pharmacia Biotech) を用いて、変性した mRNA や組み込まれなかったヌクレオチドを除去する。プローブは、放射性ヌクレオチド [^{32}P] dCTP の代わりに標識マーカ C y 3 - dCTP 或いは C y 5 - dCTP (Amersham Pharmacia Biotech) で標識することもできる。

【0126】

ハイブリダイゼーションは、0.5 M リン酸ナトリウム (pH 7.2) 及び 7% SDS、1 mM EDTA を含むハイブリダイゼーション緩衝液において 65 °C で行う。ハイブリダイゼーション緩衝液において、基板を少なくとも 2 時間、65 °C でインキュベートした後、この緩衝液をプローブ配列を含む新しい緩衝液 10 ml に取り替える。65 °C で 18 時間インキュベートした後、ハイブリダイゼーション緩衝液を除去し、最大で 40 mM リン酸ナトリウム及び 1% SDS、1 mM EDTA、65 °C の条件になるまで徐々にストリンジェンシーを増しながら、連続的に基板を洗浄する。膜にハイブリダイズした放射線標識したプローブによって生成されるシグナルを検出するために、この基板を PHOSPHORIMAGER カセット (Amersham Pharmacia Biotech) に曝露し、そのイメージを IMAGEQUANT データ解析ソフトウェア (Amersham Pharmacia Biotech) を用いて解析する。マイクロアレイにハイブリダイズした蛍光プローブによって生成されるシグナルを検出するために、この基板を共焦点レーザ顕微鏡検査によって検査し、GEMTOOLS 遺伝子発現解析ソフトウェア (Incyte Genomics) を用いてイメージを収集して解析する。

10

20

【0127】

9 相補的な核酸配列

本核酸分子に相補的な配列、或いはその断片を用いて、遺伝子の発現を検出したり、低下させたり、阻害することができる。約 15 ~ 約 30 個の塩基対からなるオリゴヌクレオチドの使用について記載するが、それより小さい或いは大きい配列の断片、またはその誘導体 (PNA) の場合でも同じ方法を用いることができる。オリゴヌクレオチドは、SEQ ID NO: 1-43 において、Oligo 4.06 ソフトウェア (National Biosciences) で設計する。プロモーターの結合を阻害して転写を阻止するために、相補的なヌクレオチドを、最も好ましくはオープンリーディングフレームの開始コドンの前の約 10 個のヌクレオチドである最もユニークな 5' 配列に結合するように設計する。翻訳を阻害するために、相補的なオリゴヌクレオチドを、本哺乳動物タンパク質をコードする mRNA にリボソームが結合するのを阻害するように設計する。

30

【0128】

10 哺乳動物タンパク質の発現

本哺乳動物タンパク質の発現及び精製は、細菌若しくはウイルス系の発現系を用いて行うことができる。細菌で発現させるために、抗生物質耐性遺伝子と cDNA を高レベルで転写させる誘導性プロモーターを含むベクターに 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、以下に限定するものではないが、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。組み換えベクターを、例えば BL21 (DE3) などの好適な細菌宿主に形質転換する。イソプロピル β-D チオガラクトピラノシド (IPTG) で誘導されると、抗生物質耐性細菌がこの哺乳動物タンパク質を発現する。真核細胞での発現は、Autographica californica 核多面性ウイルス (AcMNPV) という組み換えバキュロウイルスを Spodoptera frugiperda (Sf9) 昆虫細胞に感染させて行う。バキュロウイルスのポリヘドリン遺伝子を、相同組み換え或いは転移プラスミド中間体が関係する細菌媒介遺伝子転移の何れかによって、この哺乳動物 cDNA と置換する。ウイルスの感染力は維持され、強力なポリヘドリンプロモーターによって高いレベルで cDNA が転写される。

40

【0129】

大抵の発現系では、この哺乳動物タンパク質がグルタチオン-S-トランスフェラーゼ (

50

GST) またはペプチドエピトープタグ (FLAG) などとの融合タンパク質として合成されるため、未精製の細胞溶解物から親和性を用いた迅速な一段階の組み換え融合タンパク質の精製を行うことができる。GSTによって、タンパク質の活性及び抗原性を維持する条件下で、固定されたグルタチオンでの融合タンパク質の精製が可能となる (Amersham Pharmacia Biotech)。精製の後、GST部分をこの哺乳動物タンパク質から蛋白分解的に遺伝子操作された特定部位で切断できる。アミノ酸8個のペプチドからなるFLAGによって、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) を用いたイムノアフィニティー精製が可能となる。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) での精製が可能となる。タンパク質の発現及び精製の方法は、Ausubel (1995, 前出, ch 10, 16) に記載されている。これらの方法で精製したこの哺乳動物タンパク質を直接用いて以下のアッセイを行うことができる。

【0130】

1.1 機能のアッセイ

ASPのセリンプロテアーゼ活性は、様々なペプチドチオベンジルエーテル (peptide thiobenzylester) 基質の加水分解によって測定する。基質を選択して、異なった種類のSP (チマーゼ、trypase、aspasなど) を表す。アッセイは室温 (~25) で行い、HEPESバッファー (pH 7.5、0.01M CaCl₂、および8%ジメチルスルホキシド) はASPのアリコットおよび好適な基質を含む。反応液はまた、加水分解の際に遊離してチオピリドン (thiopyridone) に変換するチオベンジル基と反応する0.34 mM ジチオピリジン (dithiopyridine) を含む。この反応は蛍光キュベットで行い、生成されたチオピリドン分光光度計を用いて324 nmの波長における吸光度によって測定する。反応で生成されたチオピリドンがASPの活性に比例する。

【0131】

別法では、タンパク質の機能は、哺乳動物細胞培養でASPをコードする配列を生理学的に高レベルで発現させてアッセイする。この核酸配列を、強力なサイトメガロウイルスプロモーターを含むPCMV SPORTベクター (Life Technologies) にサブクローニングし、この組み換えベクター5~10 µgを、当分野で周知の形質転換法方によってヒトの内皮細胞株或いは造血細胞株に形質転換する。更に、蛍光マーカーとなるCD64-GFP (Clontech, Palo Alto CA) をコードする配列を含むプラスミド1~2 µgを同時に形質転換して、フローサイトメトリー (FCM) で形質転換細胞を同定できるようにする。

【0132】

発現における導入された遺伝子の影響は、精製されたこれらの形質転換細胞集団を用いて評価することができる。形質転換細胞の表面で発現されるCD64-GFPが、ヒト免疫グロブリンG (IgG) の保存された領域と結合するため、この形質転換細胞を、ヒトIgGか抗CD64抗体の何れかで被覆された磁気ビーズを用いて分離することができる (DYNAL, Lake Success, NY)。mRNAを細胞から単離して、ハイブリダイゼーション技術で分析することができる。

【0133】

1.2 ASPに特異的な抗体の生産

ポリアクリルアミドゲル電気泳動法でASPを精製し、これを用いてマウスやウサギを免疫する。以下の標準的なプロトコルを用いて抗体を生産する。別法では、ASPのアミノ酸配列をLASERGENEソフトウェア (DNASTAR) を用いて解析し、免疫原性の高い領域を決定する。通常はC末端付近或いは親水性領域内に存在する免疫原性エピトープを選択して合成し、これを用いて抗体を増大させる。通常は15残基程度の長さのエピトープを、Fmoc法でABI 431Aペプチドシンセサイザー (Applied Biosystems) で合成し、これをN-マレイミドベンゾイル-N-ヒドロキシスクシンイミドエステルとの反応でKLH (Sigma-Aldrich) に結合させて免疫

原性を高める。

【0134】

エピトープ - K L H 複合体をフロイントの完全アジュバントと共に投与してウサギを免疫する。その後、間隔を置いてフロイントの不完全アジュバントで免疫化を繰り返す。マウスの場合は少なくとも7週間、ウサギの場合は少なくとも12週間が経過した後、抗血清を採取して抗ペプチド活性を検査する。この検査には、このペプチドをプラスチックに結合し、1%ウシ血清アルブミン(BSA)でブロッキング処理し、ウサギ抗血清と反応させ、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。当分野で周知の方法を用いて抗体力価及び形成された複合体の収量を測定する。

【0135】

10

1.3 特異的な抗体を用いる天然タンパク質の精製

天然或いは組み換えの哺乳動物タンパク質を、このタンパク質に特異的な抗体を用いてイムノアフィニティークロマトグラフィで実質的に精製する。イムノアフィニティークラムは、この抗体をCNBr-活性化SEPHAROSEレジン(Amersham Pharmacia Biotech)に共有結合させて作製する。このタンパク質を含む培養液をイムノアフィニティークラムに通し、このタンパク質を優先的に吸着できるように、界面活性剤の存在下で高イオン強度緩衝液でそのカラムを洗浄する。結合後、緩衝液(pH 2~3)或いは高濃度の尿素やチオシアネートイオンを用いてそのカラムからこのタンパク質を溶離させて抗体とこのタンパク質との結合を切断し、このタンパク質を回収する。

20

【0136】

1.4 本核酸分子若しくはタンパク質と特異的に結合する分子のスクリーニング

本核酸分子やその断片、或いは本タンパク質やその断片を、 ^3H -dCTP、Cy3-dCTP、Cy5-dCTP(Amersham Pharmacia Biotech)、またはBIODIPYやFITC(Molecular Probes)でそれぞれ標識する。予め基板上に配列した候補分子のライブラリを、標識した核酸分子またはタンパク質の存在下でインキュベートする。核酸分子或いはアミノ酸配列の何れかが存在する条件下でインキュベートした後、その基板を洗浄し、特異的な結合或いは複合体形成を示唆する標識が保持されている基板の全ての部分をアッセイし、結合している分子を同定する。様々な濃度の核酸若しくはタンパク質で得られたデータを用いて、標識した核酸またはタンパク質と結合した分子との親和性を計算する。

30

【0137】

1.5 タンパク質の活性の実証

ASP若しくは生物学的活性を有するその断片を、 ^{125}I ボルトンハンター試薬(Boltton 他(1973) Biochem. J. 133: 529-539)で標識する。マルチウェルプレートのウェルにおいて、候補リガンド分子を標識したASPと共にインキュベートし、次に洗浄してから、標識した複合体を有する全てのウェルをアッセイする。様々な濃度のASPで得られたデータを用いて、ASPと結合した候補リガンド分子の関連性、数量、親和性を計算する。

【0138】

40

(表の簡単な説明)

本明細書の開示の一部には、著作権が保護されるべき資料が含まれている。この著作権の所有者は、特許庁のファイルや記録にある本発明の開示に関する任意の明細書部分をファクシミリで複写することについての異議はないが、その他の場合は著作権を留保する。

【0139】

配列表は核酸分子およびタンパク質分子を含む。具体的には本発明の哺乳動物核酸分子であるSEQ ID NO: 1-45と、本発明の哺乳動物タンパク質分子であるSEQ ID NO: 46-49を含む。

【図面の簡単な説明】

【図1A】

50

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 47) をコードするヒト核酸分子 (SEQ ID NO: 44) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。

【図 1 B】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 47) をコードするヒト核酸分子 (SEQ ID NO: 44) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。

【図 1 C】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 47) をコードするヒト核酸分子 (SEQ ID NO: 44) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。

【図 1 D】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 47) をコードするヒト核酸分子 (SEQ ID NO: 44) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。

【図 1 E】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 47) をコードするヒト核酸分子 (SEQ ID NO: 44) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。

【図 1 F】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 47) をコードするヒト核酸分子 (SEQ ID NO: 44) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering, South San Francisco CA) を用いて作成した。

【図 2 A】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 46) をコードするヒト核酸分子 (SEQ ID NO: 43) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering) を用いて作成した。

【図 2 B】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 46) をコードするヒト核酸分子 (SEQ ID NO: 43) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering) を用いて作成した。

【図 2 C】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 46) をコードするヒト核酸分子 (SEQ ID NO: 43) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering) を用いて作成した。

【図 2 D】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 46) をコードするヒト核酸分子 (SEQ ID NO: 43) を示す。このアラインメントは、MACDNASIS PROソフトウェア (Hitachi Software Engineering) を用いて作成した。

【図 2 E】

10

20

30

40

50

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 46) をコードするヒト核酸分子 (SEQ ID NO: 43) を示す。このアラインメントは、MACDNASISS PROソフトウェア (Hitachi Software Engineering) を用いて作成した。

【図2F】

本哺乳動物タンパク質のヒトアミノ酸配列 (SEQ ID NO: 46) をコードするヒト核酸分子 (SEQ ID NO: 43) を示す。このアラインメントは、MACDNASISS PROソフトウェア (Hitachi Software Engineering) を用いて作成した。

【図3A】

LASERGENEソフトウェアのMEGALIGNプログラム (DNASTAR, Madison WI) を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 47)、ヒトASP1、クローンID 1611565 (SEQ ID NO: 46)、ヒトガクトリクシン、GI 1658286 (SEQ ID NO: 48)、Neurospora crassa空胞プロテアーゼ (vacuolar protease)、およびGI 1039445 (SEQ ID NO: 49) との間の化学的および構造的類似性を実証する。保存されたアスパラギン酸プロテアーゼ活性部位 (D - S / T - G - S / T - S / T) には二重のアンダーラインを引いた。

10

【図3B】

LASERGENEソフトウェアのMEGALIGNプログラム (DNASTAR, Madison WI) を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 47)、ヒトASP1、クローンID 1611565 (SEQ ID NO: 46)、ヒトガクトリクシン、GI 1658286 (SEQ ID NO: 48)、Neurospora crassa空胞プロテアーゼ (vacuolar protease)、およびGI 1039445 (SEQ ID NO: 49) との間の化学的および構造的類似性を実証する。保存されたアスパラギン酸プロテアーゼ活性部位 (D - S / T - G - S / T - S / T) には二重のアンダーラインを引いた。

20

【図3C】

LASERGENEソフトウェアのMEGALIGNプログラム (DNASTAR, Madison WI) を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 47)、ヒトASP1、クローンID 1611565 (SEQ ID NO: 46)、ヒトガクトリクシン、GI 1658286 (SEQ ID NO: 48)、Neurospora crassa空胞プロテアーゼ (vacuolar protease)、およびGI 1039445 (SEQ ID NO: 49) との間の化学的および構造的類似性を実証する。保存されたアスパラギン酸プロテアーゼ活性部位 (D - S / T - G - S / T - S / T) には二重のアンダーラインを引いた。

30

【図3D】

LASERGENEソフトウェアのMEGALIGNプログラム (DNASTAR, Madison WI) を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 47)、ヒトASP1、クローンID 1611565 (SEQ ID NO: 46)、ヒトガクトリクシン、GI 1658286 (SEQ ID NO: 48)、Neurospora crassa空胞プロテアーゼ (vacuolar protease)、およびGI 1039445 (SEQ ID NO: 49) との間の化学的および構造的類似性を実証する。保存されたアスパラギン酸プロテアーゼ活性部位 (D - S / T - G - S / T - S / T) には二重のアンダーラインを引いた。

40

【図4A】

50

LASERGENEソフトウェアのSEQMAN IIプログラム(DNA STAR)を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 44)、および核酸配列変異体、1438613F1 (SEQ ID NO: 22)、1438719H1 (SEQ ID NO: 27)、1438613H1 (SEQ ID NO: 28)、5094633H1 (SEQ ID NO: 18)、2900063H1 (SEQ ID NO: 25)、3484819H1 (SEQ ID NO: 23)、2757870H1 (SEQ ID NO: 24)、2123411H1 (SEQ ID NO: 26)、5577513H1 (SEQ ID NO: 17)、2908594H1 (SEQ ID NO: 20)、2007269H1 (SEQ ID NO: 21)、および2293393H1 (SEQ ID NO: 19)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。 10

【図4B】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNA STAR)を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 44)、および核酸配列変異体、1438613F1 (SEQ ID NO: 22)、1438719H1 (SEQ ID NO: 27)、1438613H1 (SEQ ID NO: 28)、5094633H1 (SEQ ID NO: 18)、2900063H1 (SEQ ID NO: 25)、3484819H1 (SEQ ID NO: 23)、2757870H1 (SEQ ID NO: 24)、2123411H1 (SEQ ID NO: 26)、5577513H1 (SEQ ID NO: 17)、2908594H1 (SEQ ID NO: 20)、2007269H1 (SEQ ID NO: 21)、および2293393H1 (SEQ ID NO: 19)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。 20

【図4C】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNA STAR)を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 44)、および核酸配列変異体、1438613F1 (SEQ ID NO: 22)、1438719H1 (SEQ ID NO: 27)、1438613H1 (SEQ ID NO: 28)、5094633H1 (SEQ ID NO: 18)、2900063H1 (SEQ ID NO: 25)、3484819H1 (SEQ ID NO: 23)、2757870H1 (SEQ ID NO: 24)、2123411H1 (SEQ ID NO: 26)、5577513H1 (SEQ ID NO: 17)、2908594H1 (SEQ ID NO: 20)、2007269H1 (SEQ ID NO: 21)、および2293393H1 (SEQ ID NO: 19)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。 30

【図4D】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNA STAR)を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 44)、および核酸配列変異体、1438613F1 (SEQ ID NO: 22)、1438719H1 (SEQ ID NO: 27)、1438613H1 (SEQ ID NO: 28)、5094633H1 (SEQ ID NO: 18)、2900063H1 (SEQ ID NO: 25)、3484819H1 (SEQ ID NO: 23)、2757870H1 (SEQ ID NO: 24)、2123411H1 (SEQ ID NO: 26)、5577513H1 (SEQ ID NO: 17)、2908594H1 (SEQ ID NO: 20)、2007269H1 (SEQ ID NO: 21)、および2293393H1 (SEQ ID NO: 19)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。 40

【図4E】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNA STAR)を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 50

(SEQ ID NO: 44)、および核酸配列変異体、1438613F1 (SEQ ID NO: 22)、1438719H1 (SEQ ID NO: 27)、1438613H1 (SEQ ID NO: 28)、5094633H1 (SEQ ID NO: 18)、2900063H1 (SEQ ID NO: 25)、3484819H1 (SEQ ID NO: 23)、2757870H1 (SEQ ID NO: 24)、2123411H1 (SEQ ID NO: 26)、5577513H1 (SEQ ID NO: 17)、2908594H1 (SEQ ID NO: 20)、2007269H1 (SEQ ID NO: 21)、および2293393H1 (SEQ ID NO: 19)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図4F】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNASTAR)を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 44)、および核酸配列変異体、1438613F1 (SEQ ID NO: 22)、1438719H1 (SEQ ID NO: 27)、1438613H1 (SEQ ID NO: 28)、5094633H1 (SEQ ID NO: 18)、2900063H1 (SEQ ID NO: 25)、3484819H1 (SEQ ID NO: 23)、2757870H1 (SEQ ID NO: 24)、2123411H1 (SEQ ID NO: 26)、5577513H1 (SEQ ID NO: 17)、2908594H1 (SEQ ID NO: 20)、2007269H1 (SEQ ID NO: 21)、および2293393H1 (SEQ ID NO: 19)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図4G】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNASTAR)を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 44)、および核酸配列変異体、1438613F1 (SEQ ID NO: 22)、1438719H1 (SEQ ID NO: 27)、1438613H1 (SEQ ID NO: 28)、5094633H1 (SEQ ID NO: 18)、2900063H1 (SEQ ID NO: 25)、3484819H1 (SEQ ID NO: 23)、2757870H1 (SEQ ID NO: 24)、2123411H1 (SEQ ID NO: 26)、5577513H1 (SEQ ID NO: 17)、2908594H1 (SEQ ID NO: 20)、2007269H1 (SEQ ID NO: 21)、および2293393H1 (SEQ ID NO: 19)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図4H】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNASTAR)を用いて作成したアラインメントであって、ヒトASP2、クローンID 1869869 (SEQ ID NO: 44)、および核酸配列変異体、1438613F1 (SEQ ID NO: 22)、1438719H1 (SEQ ID NO: 27)、1438613H1 (SEQ ID NO: 28)、5094633H1 (SEQ ID NO: 18)、2900063H1 (SEQ ID NO: 25)、3484819H1 (SEQ ID NO: 23)、2757870H1 (SEQ ID NO: 24)、2123411H1 (SEQ ID NO: 26)、5577513H1 (SEQ ID NO: 17)、2908594H1 (SEQ ID NO: 20)、2007269H1 (SEQ ID NO: 21)、および2293393H1 (SEQ ID NO: 19)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図5A】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNASTAR)を用いて作成したアラインメントであって、ヒトASP1、1611565CB1 (SEQ ID NO: 43)、および核酸配列変異体、5157179F6 (SEQ ID NO: 29)、3985758H1 (SEQ ID NO: 30)、2846604H1 (

10

20

30

40

50

SEQ ID NO: 31)、5743028H1 (SEQ ID NO: 32)、4148749H1 (SEQ ID NO: 34)、1855226H1 (SEQ ID NO: 35)、1855420H1 (SEQ ID NO: 36)、5114558H1 (SEQ ID NO: 37)、839538R1 (SEQ ID NO: 38)、4999662H1 (SEQ ID NO: 40)、5588490H1 (SEQ ID NO: 39)、3630054H1 (SEQ ID NO: 41)、4108264H1 (SEQ ID NO: 42)、および3751907H1 (SEQ ID NO: 33)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図5B】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNASTAR)を用いて作成したアラインメントであって、ヒトASP1、1611565CB1 (SEQ ID NO: 43)、および核酸配列変異体、5157179F6 (SEQ ID NO: 29)、3985758H1 (SEQ ID NO: 30)、2846604H1 (SEQ ID NO: 31)、5743028H1 (SEQ ID NO: 32)、4148749H1 (SEQ ID NO: 34)、1855226H1 (SEQ ID NO: 35)、1855420H1 (SEQ ID NO: 36)、5114558H1 (SEQ ID NO: 37)、839538R1 (SEQ ID NO: 38)、4999662H1 (SEQ ID NO: 40)、5588490H1 (SEQ ID NO: 39)、3630054H1 (SEQ ID NO: 41)、4108264H1 (SEQ ID NO: 42)、および3751907H1 (SEQ ID NO: 33)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図5C】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNASTAR)を用いて作成したアラインメントであって、ヒトASP1、1611565CB1 (SEQ ID NO: 43)、および核酸配列変異体、5157179F6 (SEQ ID NO: 29)、3985758H1 (SEQ ID NO: 30)、2846604H1 (SEQ ID NO: 31)、5743028H1 (SEQ ID NO: 32)、4148749H1 (SEQ ID NO: 34)、1855226H1 (SEQ ID NO: 35)、1855420H1 (SEQ ID NO: 36)、5114558H1 (SEQ ID NO: 37)、839538R1 (SEQ ID NO: 38)、4999662H1 (SEQ ID NO: 40)、5588490H1 (SEQ ID NO: 39)、3630054H1 (SEQ ID NO: 41)、4108264H1 (SEQ ID NO: 42)、および3751907H1 (SEQ ID NO: 33)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図5D】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNASTAR)を用いて作成したアラインメントであって、ヒトASP1、1611565CB1 (SEQ ID NO: 43)、および核酸配列変異体、5157179F6 (SEQ ID NO: 29)、3985758H1 (SEQ ID NO: 30)、2846604H1 (SEQ ID NO: 31)、5743028H1 (SEQ ID NO: 32)、4148749H1 (SEQ ID NO: 34)、1855226H1 (SEQ ID NO: 35)、1855420H1 (SEQ ID NO: 36)、5114558H1 (SEQ ID NO: 37)、839538R1 (SEQ ID NO: 38)、4999662H1 (SEQ ID NO: 40)、5588490H1 (SEQ ID NO: 39)、3630054H1 (SEQ ID NO: 41)、4108264H1 (SEQ ID NO: 42)、および3751907H1 (SEQ ID NO: 33)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図5E】

LASERGENEソフトウェアのSEQMAN IIプログラム(DNASTAR)を用いて作成したアラインメントであって、ヒトASP1、1611565CB1 (SEQ ID NO: 43)、および核酸配列変異体、5157179F6 (SEQ ID NO: 29)、3985758H1 (SEQ ID NO: 30)、2846604H1 (SEQ ID NO: 31)、5743028H1 (SEQ ID NO: 32)、4148749H1 (SEQ ID NO: 34)、1855226H1 (SEQ ID NO: 35)、1855420H1 (SEQ ID NO: 36)、5114558H1 (SEQ ID NO: 37)、839538R1 (SEQ ID NO: 38)、4999662H1 (SEQ ID NO: 40)、5588490H1 (SEQ ID NO: 39)、3630054H1 (SEQ ID NO: 41)、4108264H1 (SEQ ID NO: 42)、および3751907H1 (SEQ ID NO: 33)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

Q I D N O : 4 3)、および核酸配列変異体、5 1 5 7 1 7 9 F 6 (S E Q I D N O : 2 9)、3 9 8 5 7 5 8 H 1 (S E Q I D N O : 3 0)、2 8 4 6 6 0 4 H 1 (S E Q I D N O : 3 1)、5 7 4 3 0 2 8 H 1 (S E Q I D N O : 3 2)、4 1 4 8 7 4 9 H 1 (S E Q I D N O : 3 4)、1 8 5 5 2 2 6 H 1 (S E Q I D N O : 3 5)、1 8 5 5 4 2 0 H 1 (S E Q I D N O : 3 6)、5 1 1 4 5 5 8 H 1 (S E Q I D N O : 3 7)、8 3 9 5 3 8 R 1 (S E Q I D N O : 3 8)、4 9 9 9 6 6 2 H 1 (S E Q I D N O : 4 0)、5 5 8 8 4 9 0 H 1 (S E Q I D N O : 3 9)、3 6 3 0 0 5 4 H 1 (S E Q I D N O : 4 1)、4 1 0 8 2 6 4 H 1 (S E Q I D N O : 4 2)、および3 7 5 1 9 0 7 H 1 (S E Q I D N O : 3 3)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

10

【図5F】

L A S E R G E N EソフトウェアのSEQMAN IIプログラム(DNA STAR)を用いて作成したアラインメントであって、ヒトASP1、1 6 1 1 5 6 5 C B 1 (S E Q I D N O : 4 3)、および核酸配列変異体、5 1 5 7 1 7 9 F 6 (S E Q I D N O : 2 9)、3 9 8 5 7 5 8 H 1 (S E Q I D N O : 3 0)、2 8 4 6 6 0 4 H 1 (S E Q I D N O : 3 1)、5 7 4 3 0 2 8 H 1 (S E Q I D N O : 3 2)、4 1 4 8 7 4 9 H 1 (S E Q I D N O : 3 4)、1 8 5 5 2 2 6 H 1 (S E Q I D N O : 3 5)、1 8 5 5 4 2 0 H 1 (S E Q I D N O : 3 6)、5 1 1 4 5 5 8 H 1 (S E Q I D N O : 3 7)、8 3 9 5 3 8 R 1 (S E Q I D N O : 3 8)、4 9 9 9 6 6 2 H 1 (S E Q I D N O : 4 0)、5 5 8 8 4 9 0 H 1 (S E Q I D N O : 3 9)、3 6 3 0 0 5 4 H 1 (S E Q I D N O : 4 1)、4 1 0 8 2 6 4 H 1 (S E Q I D N O : 4 2)、および3 7 5 1 9 0 7 H 1 (S E Q I D N O : 3 3)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

20

【図5G】

L A S E R G E N EソフトウェアのSEQMAN IIプログラム(DNA STAR)を用いて作成したアラインメントであって、ヒトASP1、1 6 1 1 5 6 5 C B 1 (S E Q I D N O : 4 3)、および核酸配列変異体、5 1 5 7 1 7 9 F 6 (S E Q I D N O : 2 9)、3 9 8 5 7 5 8 H 1 (S E Q I D N O : 3 0)、2 8 4 6 6 0 4 H 1 (S E Q I D N O : 3 1)、5 7 4 3 0 2 8 H 1 (S E Q I D N O : 3 2)、4 1 4 8 7 4 9 H 1 (S E Q I D N O : 3 4)、1 8 5 5 2 2 6 H 1 (S E Q I D N O : 3 5)、1 8 5 5 4 2 0 H 1 (S E Q I D N O : 3 6)、5 1 1 4 5 5 8 H 1 (S E Q I D N O : 3 7)、8 3 9 5 3 8 R 1 (S E Q I D N O : 3 8)、4 9 9 9 6 6 2 H 1 (S E Q I D N O : 4 0)、5 5 8 8 4 9 0 H 1 (S E Q I D N O : 3 9)、3 6 3 0 0 5 4 H 1 (S E Q I D N O : 4 1)、4 1 0 8 2 6 4 H 1 (S E Q I D N O : 4 2)、および3 7 5 1 9 0 7 H 1 (S E Q I D N O : 3 3)との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

30

【図6A】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト、1 6 1 1 5 6 5 . c o m p (S E Q I D N O : 4 3) ; およびラット、7 0 1 9 1 8 5 7 5 H 1 (S E Q I D N O : 6) および7 0 0 9 2 1 5 6 6 H 1 (S E Q I D N O : 7)、核酸配列との間の化学的および構造的類似性を実証する。

40

【図6B】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト、1 6 1 1 5 6 5 . c o m p (S E Q I D N O : 4 3) ; およびラット、7 0 1 9 1 8 5 7 5 H 1 (S E Q I D N O : 6) および7 0 0 9 2 1 5 6 6 H 1 (S E Q I D N O : 7)、核酸配列との間の化学的および構造的類似性を実証する。

50

【図6C】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト、
1611565.comp (SEQ ID NO: 43); およびラット、701918575H1 (SEQ ID NO: 6) および700921566H1 (SEQ ID NO: 7)、核酸配列との間の化学的および構造的類似性を実証する。

【図6D】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト 10
、
1611565.comp (SEQ ID NO: 43); およびラット、701918575H1 (SEQ ID NO: 6) および700921566H1 (SEQ ID NO: 7)、核酸配列との間の化学的および構造的類似性を実証する。

【図6E】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト 20
、
1611565.comp (SEQ ID NO: 43); およびラット、701918575H1 (SEQ ID NO: 6) および700921566H1 (SEQ ID NO: 7)、核酸配列との間の化学的および構造的類似性を実証する。

【図7A】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト
、1869569.comp (SEQ ID NO: 44); ラット、701911752H1 (SEQ ID NO: 10)、70190708H1 (SEQ ID NO: 9)、700494841H1 (SEQ ID NO: 8)、701596150H1 (SEQ ID NO: 12)、701031389H1 (SEQ ID NO: 14)、701823687H1 (SEQ ID NO: 11)、701188357H1 (SEQ ID NO: 5)、700769257H1 (SEQ ID NO: 4)、700769871H1 (SEQ ID NO: 3)、700251567H1.comp (SEQ ID NO: 1)、701432228T1.comp (SEQ ID NO: 2)、700542176H1 (SEQ ID NO: 13)、および700376870H1 (SEQ ID NO: 15); およびサル、700709032H1.comp (SEQ ID NO: 16)、核酸配列との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。 30

【図7B】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト 40
、1869569.comp (SEQ ID NO: 44); ラット、701911752H1 (SEQ ID NO: 10)、70190708H1 (SEQ ID NO: 9)、700494841H1 (SEQ ID NO: 8)、701596150H1 (SEQ ID NO: 12)、701031389H1 (SEQ ID NO: 14)、701823687H1 (SEQ ID NO: 11)、701188357H1 (SEQ ID NO: 5)、700769257H1 (SEQ ID NO: 4)、700769871H1 (SEQ ID NO: 3)、700251567H1.comp (SEQ ID NO: 1)、701432228T1.comp (SEQ ID NO: 2)、700542176H1 (SEQ ID NO: 13)、および700376870H1 (SEQ ID NO: 15); およびサル、700709032H1.comp (SEQ ID NO: 16)、核酸配列との間の化学的および構造的類似性を実証する。推定上のスプラ 50

イス部位は()で示した。

【図7C】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト、1869569.comp (SEQ ID NO: 44); ラット、701911752H1 (SEQ ID NO: 10)、70190708H1 (SEQ ID NO: 9)、700494841H1 (SEQ ID NO: 8)、701596150H1 (SEQ ID NO: 12)、701031389H1 (SEQ ID NO: 14)、701823687H1 (SEQ ID NO: 11)、701188357H1 (SEQ ID NO: 5)、700769257H1 (SEQ ID NO: 4)、700769871H1 (SEQ ID NO: 3)、700251567H1.comp (SEQ ID NO: 1)、701432228T1.comp (SEQ ID NO: 2)、700542176H1 (SEQ ID NO: 13)、および700376870H1 (SEQ ID NO: 15); およびサル、700709032H1.comp (SEQ ID NO: 16)、核酸配列との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図7D】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト、1869569.comp (SEQ ID NO: 44); ラット、701911752H1 (SEQ ID NO: 10)、70190708H1 (SEQ ID NO: 9)、700494841H1 (SEQ ID NO: 8)、701596150H1 (SEQ ID NO: 12)、701031389H1 (SEQ ID NO: 14)、701823687H1 (SEQ ID NO: 11)、701188357H1 (SEQ ID NO: 5)、700769257H1 (SEQ ID NO: 4)、700769871H1 (SEQ ID NO: 3)、700251567H1.comp (SEQ ID NO: 1)、701432228T1.comp (SEQ ID NO: 2)、700542176H1 (SEQ ID NO: 13)、および700376870H1 (SEQ ID NO: 15); およびサル、700709032H1.comp (SEQ ID NO: 16)、核酸配列との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図7E】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト、1869569.comp (SEQ ID NO: 44); ラット、701911752H1 (SEQ ID NO: 10)、70190708H1 (SEQ ID NO: 9)、700494841H1 (SEQ ID NO: 8)、701596150H1 (SEQ ID NO: 12)、701031389H1 (SEQ ID NO: 14)、701823687H1 (SEQ ID NO: 11)、701188357H1 (SEQ ID NO: 5)、700769257H1 (SEQ ID NO: 4)、700769871H1 (SEQ ID NO: 3)、700251567H1.comp (SEQ ID NO: 1)、701432228T1.comp (SEQ ID NO: 2)、700542176H1 (SEQ ID NO: 13)、および700376870H1 (SEQ ID NO: 15); およびサル、700709032H1.comp (SEQ ID NO: 16)、核酸配列との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。

【図7F】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト、1869569.comp (SEQ ID NO: 44); ラット、70191175

2H1 (SEQ ID NO: 10)、70190708H1 (SEQ ID NO: 9)、700494841H1 (SEQ ID NO: 8)、701596150H1 (SEQ ID NO: 12)、701031389H1 (SEQ ID NO: 14)、701823687H1 (SEQ ID NO: 11)、701188357H1 (SEQ ID NO: 5)、700769257H1 (SEQ ID NO: 4)、700769871H1 (SEQ ID NO: 3)、700251567H1.comp (SEQ ID NO: 1)、701432228T1.comp (SEQ ID NO: 2)、700542176H1 (SEQ ID NO: 13)、および700376870H1 (SEQ ID NO: 15); およびサル、700709032H1.comp (SEQ ID NO: 16)、核酸配列との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。 10

【図7G】

Phrapソフトウェア(Phil Green, University of Washington, Seattle WA)を用いて作成したアラインメントであって、ヒト、1869569.comp (SEQ ID NO: 44); ラット、701911752H1 (SEQ ID NO: 10)、70190708H1 (SEQ ID NO: 9)、700494841H1 (SEQ ID NO: 8)、701596150H1 (SEQ ID NO: 12)、701031389H1 (SEQ ID NO: 14)、701823687H1 (SEQ ID NO: 11)、701188357H1 (SEQ ID NO: 5)、700769257H1 (SEQ ID NO: 4)、700769871H1 (SEQ ID NO: 3)、700251567H1.comp (SEQ ID NO: 1)、701432228T1.comp (SEQ ID NO: 2)、700542176H1 (SEQ ID NO: 13)、および700376870H1 (SEQ ID NO: 15); およびサル、700709032H1.comp (SEQ ID NO: 16)、核酸配列との間の化学的および構造的類似性を実証する。推定上のスプライス部位は()で示した。 20

【 1 A 】

11 20 25 38 47 56
 5' ACC CTT CAC ATG AAT TTG GCC CTC GAG GCA AAT ATG CCG CAG ACC TCA GTC

 55 74 83 92 101 110
 CTT TAA GAT GGT CTT TCC CAA GCA TCC CCG AAG CCG TCA AAT TTG CCT TCC

 119 128 137 146 155 164
 AAT TTA GAC AAG AAG CTG TTC TCT GCG TTT AAG GCG GTA TCA GGA CTC AAT CCT

 173 182 191 200 209 218
 TGT CAC TTT GTC CCA AAG CCG GAC GAC TCC CTG GAG CCG TTT GAC TGT CAG

 227 236 245 254 263 272
 GCA AAG CAG ACC CAC GTC CCG CAC CTC TTC TCC CTG CAG CTT TGT GGT GAT GAG

 281 290 299 308 317 326
 TTC CCC CTC AAC CAG TCG GAA CTC CTC CCG TCT CTC GGA GCG ACC AATG ATG AAT

 P P T T N Q S F F V T A S Y G G F M I I

【 1 B 】

335 344 353 362 371 380
 GAA GGT ATC CAC CAC TCC CTG TAC ACA GGC AGT CTC TGG GAT ACA CCG ATC CCG

 388 398 407 416 425 434
 CCG CAG TGG TAT TAT CAG CTC AAT ATT GTC CCG GTC GAG ATC ACC GGA CAG GAT

 443 452 461 470 479 488
 CTG AAA ATG GAC TTC AAG GAG TAC AAC TAT GAC PAA GGC AAT GTC GAC ACT GGC

 497 506 515 524 533 542
 ACC ACC AAC CTT CTT TTG CCG AAG AAA GTG TTT AAA GCT GCA CTC AAA TCC AAT

 551 560 569 578 587 596
 AAG GCA GCG CCG TTC ACG GAG AAG TTC CTT GAT TTT TGG CTA GCA GAG CAG

 K A A S S T R K P P L G F M L G E G

【 1 C 】

514 523 611 620 629 638 647 656
 CTG CTG TCC TCG GAA GCA GGC ACC ACC CCG TGG ALC ATG TTT CCA GTC ATC TCA

 659 668 677 686 695 704
 CTC TAC CTA AAG GGT GAG GTT ACC AAC CAG TCC TTC CCG ATC ACC ATC CTT CCG

 713 722 731 740 749 758
 CAG CAA TAC CTG CCG CCA GTG GAA GAT GTG GCC ACC TCC CAA GAC GAC TGT TAC

 767 776 785 794 803 812
 AAG TTT GCC ATC TCA CAG TCA TCC ACC GGC ACT GTT ATG GGA GCT GAT ATC ATG

 821 830 839 848 857 866
 GAG GGC TTC TAC GTT CTC TTT GAT CCG GCT CCA AAA CGA AAT GGC TTT GCT GTTC

 875 884 893 902 911 920
 ACC GCT TGC CMT GTC CAC GAT GAG TTC AAG ACC GCA GCG GTC GAA GGC CCT TTT

 929 938 947 956 965 974
 CTC ACC TTG GAC ATG GAA GAC TGT GTC TAC AAC ATT CCA CAG ACA GAT GAG TCA

 983 992 1001 1010 1019 1028
 ACC CTC ACC AAT GCC TAT GTC ATG GCT GCC ATC TGC GCC CTC TTC ATG CTS

 1037 1046 1055 1064 1073 1082
 CCA CTC TGC CTC ATG GTG TGT CAG TGG CCG TGC CTC CCG TTC CAG CAG CAG

 1091 1100 1109 1118 1127 1136
 GAT GAT GAC TTT GCT GAT GAC AUC TCC CAG CAG AAG TCA GCA GCG CCA TGG GCA

 1145 1154 1163 1172 1181 1190
 GAA GAT AAT TCC CCG GCA CCG CAC CTC GGT GAT TCA CTT TGG TCA GAA GAA

【 1 D 】

875 884 893 902 911 920
 ACC GCT TGC CMT GTC CAC GAT GAG TTC AAG ACC GCA GCG GTC GAA GGC CCT TTT

 929 938 947 956 965 974
 CTC ACC TTG GAC ATG GAA GAC TGT GTC TAC AAC ATT CCA CAG ACA GAT GAG TCA

 983 992 1001 1010 1019 1028
 ACC CTC ACC AAT GCC TAT GTC ATG GCT GCC ATC TGC GCC CTC TTC ATG CTS

 1037 1046 1055 1064 1073 1082
 CCA CTC TGC CTC ATG GTG TGT CAG TGG CCG TGC CTC CCG TTC CAG CAG CAG

 1091 1100 1109 1118 1127 1136
 GAT GAT GAC TTT GCT GAT GAC AUC TCC CAG CAG AAG TCA GCA GCG CCA TGG GCA

 1145 1154 1163 1172 1181 1190
 GAA GAT AAT TCC CCG GCA CCG CAC CTC GGT GAT TCA CTT TGG TCA GAA GAA

【 1 E 】

1199 1208 1217 1226 1235 1244
 GGA GAC ACA GAT GGC ACC TGT GGC CAG AUC AUC TCA GGA CCC TCC CCA CCC ACC
 1263 1262 1271 1280 1289 1298
 AAA TGC CTC TCC CTT GAC GAA GAA AAA GGC TGG GAA GAT GGG TTC CAG GGA
 1307 1316 1325 1334 1343 1352
 CTC TAC CAG TAG GAA GAA AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG
 1361 1370 1379 1388 1387 1406
 ACT CTC GGT GAT CTC AAA TTT AAG TCG GGA AAT TCT GCT TGA AAG TTC AGC
 1415 1424 1433 1442 1451 1460
 CTT GAA CTT TTT TCC ACC AAT CCA TTA AAT TCT CCA ACC CAA AAT ATT CTT CTT
 1469 1478 1487 1496 1505 1514
 TTC TTA GTT TCA GAA GAT CAG TGA CAC GCG GGT TAC CTT GGC GTC TCT CCC
 1523 1532 1541 1550 1559 1568
 TGT GGT ACC CTG CCA GAG AAG AAG CCA ACC ACC TTT CTT CCC TCC TGG CCA AAG TCA

【 1 F 】

577 1586 1595 1604 1613 1622
 GGA GAA GAA GAT CCA CAG TTT GCT AAT TCC TTT AAG AAG ACC AAG TAT ATA AAG
 1631 1640 1649 1658 1667 1676
 AAG CTT AAG AAT GGT GCA AAG AAT GCG TTT TGA ATT AAA AAA AAA AAG TAG ATT
 1686 1694 1703 1712 1721 1730
 GAC TAT TGA TAC AAA TCC GCG CCG CCG GAA AAG AAG AAG AAG AAG AAG AAG AAG
 1739 1748 1757 1766 1775 1784
 AAG CAG GAA ATG GAG GGA TCA AAG CTA GGA AAG GCA GAA GAA CCA CTC ACC
 1793 1802 1811 1820 1829 1838
 ACT CTT AAT TTT ACA CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT CTT
 1847 1856 1865 1874 1883 1892
 TTT GAT TTT CAA TGG TCT CTT TTT CTT GGT GGT GGC ACC CAG AAA AAT GAT AAG
 1901 1910
 GGG AAG GGG CTA TCT AAG C 3'

【 2 A 】

5' AAG CTT AAT CTG GAC TCC AAG GAG TAT AAC GYA GAC AAG ACC ATC GTS GAC AAC 54
 63 62 71 81 90 99 108
 CTT GAG GGG GAC TCT GGC GGC TAC TAC CTG GAG AAG CTG ATC GGG ACC CCC
 M I G I P
 117 126 135 144 153 162
 CCG CAG AAG CTA CAG APT CTC GTT GAC ACT GGA ACC AAT AAC TTT GCC GTG GCA
 P Q R L Q I L V D I G S S N F A V A
 171 180 189 198 207 216
 GSA ACC CCG GAC TCC TAC AAG ACC AAC TTT GAC ACA GAG AAG TCT AAG ACA
 G T P H S Y I D T Y F D T E R S S E
 225 234 243 252 261 270
 TAC CCC TCC AAG GGC TTT GAC GTC ACA GTC AAG GTC ACA CAA GGA AAG TGG AAG
 Y R S X G F D V T V R Y T Q G D H T

【 2 B 】

279 288 297 306 315 324
 GGC TTC GAG GAA GAC CTC GTC ACC ATC CCC AAA GGC TTC AAT ACT TCT TTT
 G F V G E D L V T I P K G F N W S Y
 333 342 351 360 369 378
 CTT GTC AAG ATT GCC AAT ATT TTT GAA TCA GAG AAT TTC TTT CTG GCT GGA APT
 L V N I A T I F E S E N F F L P G I
 387 396 405 414 423 432
 AAA TGG AAT GGA ATA CTT GGC GCA GCT TAT GGC ACA CTT CCC AAG CCA TCA AAG
 K W N G I L G L A Y A T L A K P S S
 441 450 459 468 477 486
 TCT GTS GAG ACC TTC TTC GAC TTC CTC GTS ACA CAA GCA ACC ACC ACC AAC GTT
 S L H H F P D S L V T Q A N I P N V
 495 504 513 522 531 540
 TTC TCC ATG CAG ATG TGT GGA GGC GGC TGG CCC GTT GCT GGA TCT GGC ACC AAC
 F S M Q H C G A G L P V A C S G T N

【 2 C 】

549 558 567 576 585 594
 GGA GGT AGT CTT GTC TTG GAT GGA AAT GAA CTA APT TPG TAT AAA GGA GAC AAT
 C C S L V L G G I E P S L Y R G D I
 603 612 621 630 639 648
 TGG TAT ACC CCT AAT AAG GAA GAG TGG TAC TAC CAG AAG GAA AAT CTTG AAA CTG
 W Y T P I K E E W Y Y Q I E I L K L
 657 666 675 684 693 702
 GAA AAT GGA GGC GAA AAG CTT AAT CTG GAC TCC AGA GAG TAT AAC GCA GAC AAG
 E I G G Q S L N L D C H E Y N A D K
 711 720 729 738 747 756
 GCC ATC GTC GAC AAT GCC ACC ACG CTG CTC GGC CTC CCC CAG AAG CTG TTT GAT
 A I V D S G T L L R L P Q K V F D
 765 774 783 792 801 810
 GCG GTC CTG GAA GCT GTC CCC GCA TCT CAG ATT CCA GAA TTC TCT GAT GAT
 A V V E A V A R A S L I P E F S D G

【 2 D 】

819 828 837 846 855 864
 TTC TGG ACT GCG TCC GAG CTG GCG TGG ACG AAT TCG GAA ACA OCT TGG TCT
 F W T G S Q L A C W T N S E T P W S
 873 882 891 900 909 918
 TAC TTC CCT AAA ATC TCC ATC TAC CTG AGA GAC GAG AAC TCC AAG AGG TCA TTC
 Y P K I S I Y L R D E N S S R S F
 927 936 945 954 963 972
 CCT ATC ACA ATC CTG CCT GAG CTT TAC AAT CAG CCC ATC ATC GGG GGC GGC CTG
 R I T L F Q L Y T Q P M H G A G L
 981 990 999 1008 1017 1026
 AAT TAT GAA TGT TAC GAA TCC GGC ATC TCC CCA TCC AAT GAG AAT GCG CTG GAG AATC
 N Y E C Y R F G I S P S T N A I V I
 1035 1044 1053 1062 1071 1080
 GGT GCC ACG GTC AAG GAG GCG TTC TAC GTC ATC TTC GAC AAG GCG GAG AAG AGG
 G A T V M F G P Y I V I D N A Q K R

【 2 E 】

1089 1098 1107 1116 1125 1134
 GTC GGC TTC GCA GCG AAG CCC TGT GAA GAA AAT GCA GAT CTT CCA GAG TCT GAA
 V G F A A S P C A E I A G A A V S E
 1143 1152 1161 1170 1179 1188
 ATT TCC GGG CCT TTC TCA ACA GAG GAT GTA GCG AAG TAT TAT CAC CCC GCT CAG
 I S G F F S T E D V A S N C V P A Q
 1197 1206 1215 1224 1233 1242
 TCT TCG TGC GAG CCC AAT TCG TGG AAT GTC TCA TAT GCG CTC AAG ACC GTC TCC
 S L S E P I L W I V S Y A L N S V C
 1251 1260 1269 1278 1287 1296
 GGA GGC ATC CTC CTT GTC TTA AUC GTC CTG CTG CCG CCG TTC CCG TGT CAG
 G A I L H V L I V L L L L P F R C Q
 1305 1314 1323 1332 1341 1350
 GAT CTC GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
 R R P R D P E V V N D E S S L V R H

【 2 F 】

1359 1368 1377 1386 1395 1404
 CGC TCG AAA TCA AAT GCA GCG GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
 R W K
 1413 1422 1431 1440 1449 1458
 AGA AAA TCA CTC TTC CAG CCC AGC AAG CCG GAT GCA TGG TGG GCG TTT CTC CTG
 1467 1476 1485 1494 1503 1512
 TGC CCA CCC GTC TTC AAT CTC TGT TGT CCG AGA TGT CTA GAT TCA CTA CTG
 1521 1530 1539 1548 1557 1566
 TCT TGT GAT TCT TCA TTT TCA ACG TTT CAA ATC CTC CCG AAT TCC AAG AAA AAC
 1575 1584 1593
 AAT TAA AAA AAA AAC TTC AAT CTA AAC CAA AAA AA 3'

【 3 A 】

1 G-----1869869
 1 M-----1611565
 1 M K M M V V L V C I Q L L --E A V V V V P L K K --G I 1658286
 1 M K --G A L L T A A L L L G S A Q A G V H T M K L K K V P G I 1039445
 2 -----1869869
 2 -----1611565
 26 --FKS I R E T M K E K G L L G E F --L R T H K Y D G I 1658286
 29 L A D E L E S V P I D V Q V H L G Q K Y T G L R T E S H T G I 1039445
 2 -----1869869
 2 -----1611565
 50 F A W K Y R F G D L S V T -Y E P M A --Y M D A A Y F G G I 1658286
 59 Q A -M F K A T D A Q V S G N H P V P I T N F M N A Q Y F S G I 1039445
 2 --L T P C-----1869869
 2 --L I G T P F O K L O I L V D T G S S N E A V A --G I 1611565
 76 E I S I G T P P O N F L V L F D T G S S N L W V P S V Y C O G I 1658286
 88 E I T I G T P E O T F K V V L D T G S S N L W V P S S O C G I 1039445
 6 T P H S Y I D T Y F D T E R S T Y R S Z G P D V T V K Y T 1611565
 106 S Q A C T H S R E N F S S E S T Y B T N O C T F S L O Y G G I 1658286
 118 S I A C T L H N K R Y E S S E S E S I I K K N Q I S F F K I E L I G G I 1039445

【 3 B 】

6 -----1869869
 56 Q G S W T G F V G E D L V T I F K G F M T S F L V N I A T I 1611565
 136 Q G S L T G F F G Y P T L M Y Q S I Q V P N O E F G L S E N G I 1658286
 148 S G S L T G S F V S Q P R M T L G D I T I N D Q P A B A N S G I 1039445
 6 -----1869869
 6 --L F L P C-----1869869
 146 F E G R N F L P C I K N G I L G A V A T L A K P S S S 1611565
 178 E P G L H A F G --R P D G I L G L G D R L A V D E A T G I 1039445
 15 L R E F T S T V K Q T H V P N -L P G I Q J C G A G P F I 1869869
 116 L E T F E D S L V T O A N I P N -L P S M O K C C G A G L P V 1611565
 194 --T A M O S M V O E G A L T S P V S V Y L S N O Q --G I 1658286
 206 --P P E Y K M V E Q K L V D E P N E S F Y L A D Q D --G I 1039445
 44 N O D L A S V G G S M I I G E T D H S I V T G S L W V T 1869869
 145 A G S --G T N G S L V L G G L E P S L Y K G D I W Y T 1611565
 219 --S S G A V P P S G V D S I L Y P F G O I Y W A G I 1658286
 231 --S E S B --V V P P C V N K D R Y T C R I T T I G I 1039445
 74 P T R E M V Y V Y V Y I V R V N I N G O D K M D C K E Y N 1869869
 172 P I K E S Y Y O I F I L K L B I G G O S L N L D C R R Y N 1611565
 243 P V T O U P L X M O T G I E P L I G G O A S G W -C S E R --G I 1658286
 253 P L R R R A Y W E V D F D A I G Y G K D P A B I --G I 1039445

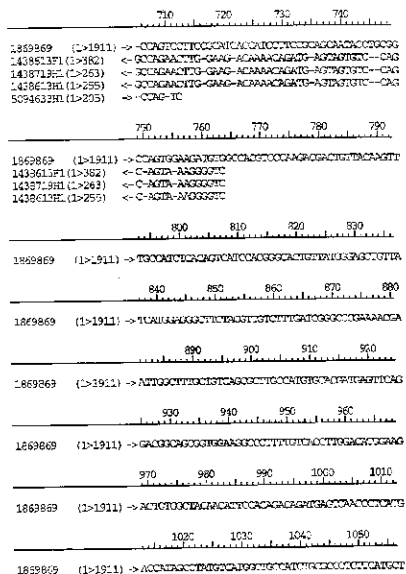
【 3 C 】

104 Y P R S I V D S G T N L R L F K K V F E A A V K S I -K A 1869869
 202 A D K A I V D S G T L L R L F Q K Y F D A V V E A Y -K R 1611565
 276 G C O A I V P F T S L L T V F Q Q Y M S A L L Q A T G A Q G I 1658286
 278 G H G V L L E F G T S L I A L F S Q J A E X L N A Q I C A K G I 1039445
 133 A S I T S K F P D G F W L G E C L V C W Q A G T P P W N I F 1869869
 231 A S L I P E F S D G E W T G S C A C M T K S E T E M S Y F 1611565
 300 P E Y G Q P L --V N C N S I Q N --I G I 1658286
 308 K S E M N G Q F I T --I D C G K K S I S --L G I 1039445
 163 P V I S L V I M G E L V T N Q P R I T I L P Q Q V L R F V E 1869869
 261 P X I S I Y L R D E N S R S F R I T I L F Q L Y I Q E M M 1611565
 317 L S L T P I N G --V F F L P P S S Y I L S N N G I 1658286
 325 E O V T P L A G --Y N F T L G F E D Y I L R A S G I 1039445
 193 D V A T S Q D C V E F A I S I S T G V M G A V I M E G 1869869
 291 G A L L N X E -C Y R F G I S E S T N A L Y I G A T V M E G 1611565
 341 G Y C T V G V E P T Y L S S Q G P L W I L G D V F L R S G I 1658286
 349 G S C H S T F M G M D M P A P V G -P L A I L G I D A P L R K G I 1039445
 223 F Y V V F D R A R K R I G F A M S A C H V H D E F P E T A A V 1869869
 320 K Y V T P D R A O K R V G F A S E C A E I A G A A V S E I 1611565
 374 Y M S V V D L G N N R V G F A T A A G I 1658286
 378 Y M S I V D L G A D T V G I A T A K -----G I 1039445

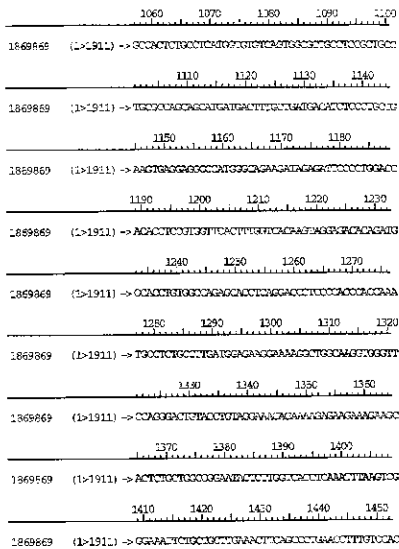
【 3 D 】

293 E G S E V T L D M E D C Q Y N E Q T D F I S I L M T I A V Y 1869869
 350 S G P E T E D V A S N C Y P A S F S R P I L H I V S Y A 1611565
 386 -----G I 1658286
 396 -----G I 1039445
 283 M A A T C A L E M J P L C F M V C Q M R C L R C L R Q Q H 1869869
 388 L M S V C G A T L L V L L L L L L L F R C C R R T R D E I 1611565
 396 -----G I 1658286
 -----G I 1039445
 312 D D F A D D I S I L -----R 1869869
 409 E V V M D R E S S L V R H K H K 1611565
 388 -----G I 1658286
 396 -----G I 1039445

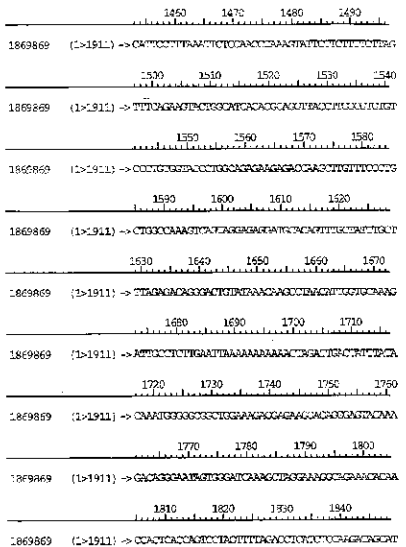
【 図 4 E 】



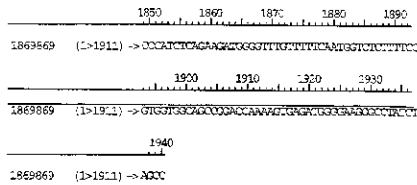
【 図 4 F 】



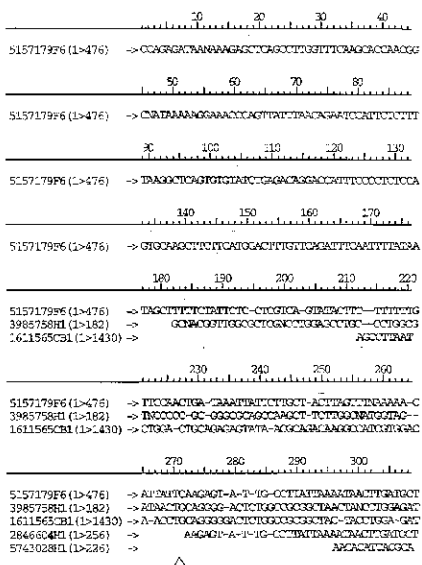
【 図 4 G 】



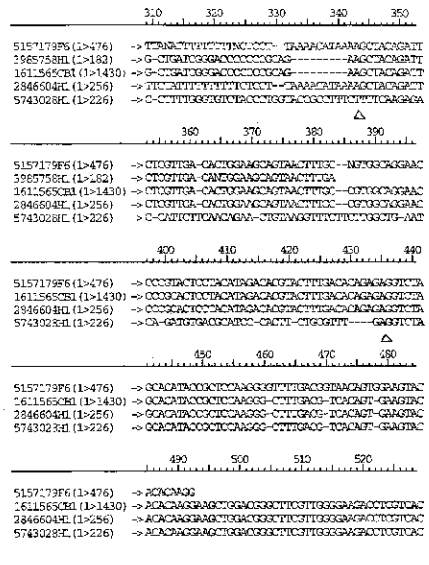
【 図 4 H 】



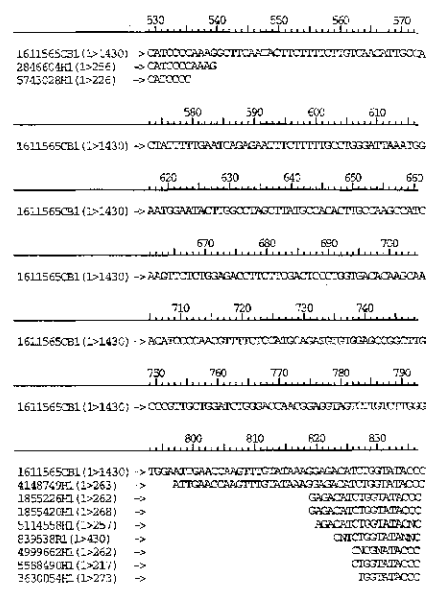
【 図 5 A 】



【 図 5 B 】



【 図 5 C 】



【 図 5 D 】



【 6 B 】

1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1

【 6 C 】

1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1

【 6 D 】

1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1

【 6 E 】

1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1
 1611565CB1
701918579H1
700921566H1

【 7 A 】

1869869.comp 10 20 30 40 50 60 70
 80 90 100 110 120 130 140
 1869869.comp 150 160 170 180 190 200 210
 220 230 240 250 260 270 280
 1869869.comp 290 300 310 320 330 340 350
 360 370 380 390 400 410 420
 1869869.comp 430 440 450 460 470 480 490

【 7 B 】

1869869.comp 500 510 520 530 540 550 560
 570 580 590 600 610 620 630
 1869869.comp 640 650 660 670 680 690 700
 710 720 730 740 750 760 770
 1869869.comp 780 790 800 810 820 830 840
 850 860 870 880 890 900 910
 1869869.comp 920 930 940 950 960 970 980
 704911782HL

【 7 C 】

1869869.comp 990 1000 1010 1020 1030 1040 1050
 1060 1070 1080 1090 1100 1110 1120
 1869869.comp 1130 1140 1150 1160 1170 1180 1190
 1200 1210 1220 1230 1240 1250 1260
 1869869.comp 1270 1280 1290 1300 1310 1320 1330
 1340 1350 1360 1370 1380 1390 1400
 704911782HL

【 7 D 】

700494841HL 1410 1420 1430 1440 1450 1460 1470
 701595150HL 1480 1490 1500 1510 1520 1530 1540
 701031339HL 1550 1560 1570 1580 1590 1600 1610
 701823687HL 1620 1630 1640 1650 1660 1670 1680
 1869869.comp 1690 1700 1710 1720 1730 1740 1750
 70190708HL 1760 1770 1780 1790 1800 1810 1820
 700494841HL 1830 1840 1850 1860 1870 1880 1890
 701031339HL 1900 1910 1920 1930 1940 1950 1960
 701823687HL 1970 1980 1990 2000 2010 2020 2030
 701188947HL 2040 2050 2060 2070 2080 2090 2100
 700769247HL 2110 2120 2130 2140 2150 2160 2170
 700769871HL 2180 2190 2200 2210 2220 2230 2240
 1869869.comp 2250 2260 2270 2280 2290 2300 2310
 70494841HL 2320 2330 2340 2350 2360 2370 2380
 701595150HL 2390 2400 2410 2420 2430 2440 2450
 701031339HL 2460 2470 2480 2490 2500 2510 2520
 701823687HL 2530 2540 2550 2560 2570 2580 2590
 701188947HL 2600 2610 2620 2630 2640 2650 2660
 700769247HL 2670 2680 2690 2700 2710 2720 2730
 700769871HL 2740 2750 2760 2770 2780 2790 2800

【 7 E 】

1869869 comp
 701422287H1 comp
 701188157H1
 701188157H1
 700768701H1
 7002511567H1 comp
 7014322287H1 comp
 1413 1423 1430 1440 1450 1460 1470

1869869 comp
 701422287H1
 701188157H1
 700768701H1
 7002511567H1
 7014322287H1 comp
 1490 1490 1510 1510 1520 1530 1540

【 7 F 】

1869869 comp
 700542176H1
 700768701H1
 700768701H1
 7002511567H1
 7014322287H1 comp
 1550 1560 1570 1590 1600 1610

1869869 comp
 701188157H1
 700768701H1
 7002511567H1
 7014322287H1 comp
 1620 1630 1640 1650 1660 1670 1680

1869869 comp
 7014322287H1 comp
 7014322287H1
 700768701H1
 7002511567H1
 7014322287H1 comp
 1760 1770 1780 1790 1800 1810 1820

【 7 G 】

1869869 comp
 700542176H1
 700768701H1
 700768701H1
 7002511567H1
 7014322287H1 comp
 1890 1900 2000 2010 2020 2030 2040

1869869 comp
 701188157H1
 700768701H1
 7002511567H1
 7014322287H1 comp
 2100 2110 2120 2130 2140 2150 2160

【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 28 June 2001 (28.06.2001)

PCT

(10) International Publication Number WO 01/46398 A2

- (51) International Patent Classification: C12N 9/64
(72) Inventors: and
(75) Inventors/Applicants (for US only): KASER, Matthew, R. [GB/US]; 4793 Ewing Road, Castro Valley, CA 94546 (US); COCKS, Benjamin, G. [AU/US]; 1215 Van Dyke Drive, Sunnyvale, CA 94087 (US); LORING, Jeanne [US/US]; 701 Santa Cruz Lane, Foster City, CA 94404 (US); TANG, X. Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US); YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).
(21) International Application Number: PCT/US00/33743
(22) International Filing Date: 12 December 2000 (12.12.2000)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data: 09/470,954 22 December 1999 (22.12.1999) US
(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application: 09/470,954 (CIP) Filed on 22 December 1999 (22.12.1999)
(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
(74) Agents: TURNER, Christopher et al., Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
(81) Designated States (national): AF, AG, AI, AM, AU, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CL, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GG, GT, HA, HK, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, NZ, NI, NZ, NL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
[continued on next page]

(54) Title: ASPARTYL PROTEASES

5' 11 20 29 38 47 56
ACG CCG CAC ATA GGG AAT TTG GCG CTC GAG GCA AGA ATT CCG CAG AGC TCA GTG
65 74 83 92 101 110
CCG TAA GAG GGT CTE TCC CAA GGA GCA TCC CCC AAG CCT TCA AAC TTG CCT TCC
119 128 137 146 155 164
AGT TTA GAC AGG AAG CTG CCC TCT GCC TTT AGG CCC GTA TGA GGA CTC ACT CCT
173 182 191 200 209 218
TGT CTC TTT CTG CCA AAG CTT GAC GAC TCC CTG GAG CTT TTC TTT GAC TCT CTG
C L F L F K P E D S L R P F F D S I
227 236 245 254 263 272
GTA AAG CAG ACC CAC GTT CCC AAC CTC TTC TCC CTG CAG CTT TGT GGT CCT GGC
V K Q T E V P N H F S L Q L C G A G
281 290 299 308 317 326
TTC CCC CTC AAC CAG TCT GAA CTG CTG GCC TCT GTC GGA GGG AGC ATG ATC ATT
F P L N Q S E V L A S V G G S H I I

(57) Abstract: The invention provide mammalian nucleic acid molecules and fragments thereof. It also provides for the use of the mammalian nucleic acid molecules for the characterization, diagnosis, evaluation, treatment, or prevention of conditions, diseases and disorders associated with Alzheimer's disease and Down syndrome. The invention additionally provides expression vectors and host cells for the production of the protein encoded by the mammalian nucleic acid molecules.



WO 01/46398 A2

WO 01/46398 A2



patent (AM, AZ, BY, KG, KZ, MD, RU, UJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

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

Published:

Without international search report and to be republished upon receipt of that report.

WO 01/46398

PCT/US00/33743

ASPARTYL PROTEASES

This application is filed under the Patent Cooperation Treaty and claims the benefit of U.S. Nonprovisional Application No. 09/470,954, our Docket No. PC-0008 US, filed 22nd December, 1999.

5

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of new mammalian protein, nucleic acid sequence variants thereof, and to the use of these sequences in the characterization, diagnosis, prevention, and treatment of Alzheimer's disease and Down syndrome.

10

BACKGROUND ART

Phylogenetic relationships among organisms have been demonstrated many times, and studies from a diversity of prokaryotic and eukaryotic organisms suggest a more or less gradual evolution of biochemical and physiological mechanisms and metabolic pathways. Despite different evolutionary pressures, proteins that regulate the cell cycle in yeast, nematode, fly, rat, and man have common chemical or structural features and modulate the same general cellular activity. Comparisons of human gene sequences with those from other organisms where the structure and/or function may be known allow researchers to draw analogies and to develop model systems for testing hypotheses. These model systems are of great importance in developing and testing diagnostic and therapeutic agents for human conditions, diseases and disorders.

Senile neuritic plaques, the neuropathological signature of Alzheimer's disease at post-mortem examination, are found in the hippocampus and neocortical regions of the brain. The hippocampus is part of the brain's limbic system and plays an important role in learning and memory. In Alzheimer's disease patients, accumulating plaques damage the neuronal architecture of the hippocampus and eventually cripple the memory process. The disease can be early onset, affecting individuals as young as 40 years of age, and can be familial or sporadic. Familial Alzheimer's disease has been thought to be inherited strictly as an autosomal dominant trait. However, this view is changing as more genetic determinants are isolated. For example, some normal allele variants of apolipoprotein E (ApoE), a protein found in senile plaques, can either protect against, or increase, the risk of developing the disease (Strittmatter *et al.* (1993) Proc. Natl. Acad. Sci. 90:1977-1981). Mutations in three other genes, amyloid precursor protein (APP), presenilin-1, and presenilin-2, have also been shown to predispose an individual to Alzheimer's disease (Selkoe (1999) Nature 399:A23-A31).

β -amyloid protein (A β) is the major component of senile plaques, and it is normally formed when β - and γ -secretases cleave APP. In Alzheimer's disease patients, large quantities of A β are generated and

WO 01/46398

PCT/US00/33743

accumulate extracellularly in these neuropathological plaques. Efforts to understand the mechanism underlying A β deposition have recently focused on the APP-cleaving secretases since two yeast aspartyl proteases have been shown to process APP *in vitro* (Zhang *et al.* (1997) *Biochim Biophys Acta* 1359:110-122). Recent evidence using peptidomimetic probes further confirms that the secretases are intramembrane-cleaving aspartyl proteases (Wolfe *et al.* (1999) *Biochemistry* 38:4720-4727).

The presenilin-1 gene is a candidate for the γ -secretase that cleaves the APP carboxyl terminus. Several lines of evidence support the involvement of presenilins in the disease process. These include coimmunoprecipitation of presenilin and APP and discovery that nonsense point mutations in presenilin genes can result in a particularly aggressive, early onset form of the disease (Haas and De Strooper (1999) *Science* 286:916-919). Presenilins are thought to cleave notch protein, which, like APP, is cleaved by α - and β -secretases in its ectodomain to produce the carboxyl terminal fragment that is a substrate for γ -secretases (Haas and De Strooper, *supra*).

Aspartic proteases are members of the cathepsin family of lysosomal proteases and include pepsin A, gastricsin, chymosin, renin, and cathepsins D and E. Aspartic proteases include two aspartic acid residues in the active site and are most active between pH 2-3, where one of the aspartic acid residues is ionized and the other not. The catalytic aspartic acid residue of the aspartyl protease active site is characterized by the amino acid residue sequence D-T/S-G-I/S-T/S (aspartyl protease signature PS00141; PROSITE PDOC00128). A potent inhibitor of aspartic proteases is the hexapeptide pepstatin which, in the transition state, resembles normal substrates. Aspartic proteases also include bacterial penicillopepsin, mammalian pepsin and chymosin, and certain fungal proteases. Abnormal regulation and expression of cathepsins is evident in various inflammatory disease states such as arthritis.

The discovery of additional nucleic acid sequences encoding aspartyl proteases and nucleic acid sequence variants thereof provides new compositions that are useful in the characterization, diagnosis, prevention, and treatment of Alzheimer's disease and Down syndrome.

25

DISCLOSURE OF INVENTION

The invention is based on the discovery of mammalian nucleic acid molecules encoding aspartyl protease (ASP) and nucleic acid sequence variants thereof, which satisfies a need in the art by providing compositions useful in the characterization, diagnosis, prevention, and treatment of conditions such as Alzheimer's disease and Down syndrome.

The invention provides isolated and purified human, monkey, and rat nucleic acid molecules comprising SEQ ID NOs:1-43, and fragments thereof, encoding the mammalian protein comprising amino acid sequences of SEQ ID NOs:46 and 47 or portions thereof, a biologically active fragment of SEQ ID NO:46, and an immunologically active fragment of SEQ ID NO:46.

WO 01/46398

PCT/US00/33743

The invention further provides a probe that hybridizes under high stringency conditions to the mammalian nucleic acid molecule or fragments thereof. The invention also provides isolated and purified nucleic acid molecules that are complementary to the nucleic acid molecules of SEQ ID NOs:1-43. In one aspect, the probe is a single stranded complementary RNA or DNA molecule.

5 The invention further provides a method for detecting a nucleic acid molecule in a sample, the method comprising the steps of hybridizing a probe to at least one nucleic acid molecule of a sample, forming a hybridization complex; and detecting the hybridization complex, wherein the presence of the hybridization complex indicates the presence of the nucleic acid molecule in the sample. In one aspect, the method further comprises amplifying the nucleic acid molecule prior to hybridization. The nucleic acid
10 molecule or a fragment thereof may comprise either an element or a target on a microarray.

The invention also provides a method for using a nucleic acid molecule or a fragment thereof to screen a library of molecules to identify at least one ligand that specifically binds the nucleic acid molecule, the method comprising combining the nucleic acid molecule with a library of molecules under conditions allowing specific binding, and detecting specific binding, thereby identifying a ligand that specifically binds
15 the nucleic acid molecule. Such libraries include DNA and RNA molecules, peptides, PNAs, proteins, and the like. In an analogous method, the nucleic acid molecule or a fragment thereof is used to purify a ligand.

The invention also provides an expression vector containing at least a fragment of the nucleic acid molecule. In another aspect, the expression vector is contained within a host cell. The invention further provides a method for producing a protein, the method comprising the steps of culturing the host cell under
20 conditions for the expression of the protein and recovering the protein from the host cell culture.

The invention also provides substantially purified mammalian ASP or a portion thereof. The invention further provides isolated and purified proteins having the amino acid sequence of SEQ ID NO:45, a biologically active fragment of SEQ ID NO:46, and an immunologically active fragment of SEQ ID NO:46. Additionally, the invention provides a pharmaceutical composition comprising a substantially
25 purified mammalian protein or a portion thereof in conjunction with a pharmaceutical carrier.

The invention further provides a method for using at least a portion of the mammalian protein to produce antibodies. The invention also provides a method for using a mammalian protein or a portion thereof to screen a library of molecules to identify at least one ligand that specifically binds the protein, the method comprising combining the protein with the library of molecules under conditions allowing specific
30 binding, and detecting specific binding, thereby identifying a ligand that specifically binds the protein. Such libraries include DNA and RNA molecules, peptides, agonists, antagonists, antibodies, immunoglobulins, drug compounds, pharmaceutical agents, and other ligands. In one aspect, the ligand identified using the method modulates the activity of the mammalian protein. In an analogous method, the protein or a portion thereof is used to purify a ligand. The method involves combining the mammalian protein or a portion

WO 01/46398

PCT/US00/33743

thereof with a sample under conditions to allow specific binding, detecting specific binding between the protein and ligand, recovering the bound protein, and separating the protein from the ligand to obtain purified ligand.

The invention further provides a method for inserting a marker gene into the genomic DNA of a mammal to disrupt the expression of the natural mammalian nucleic acid molecule. The invention also provides a method for using the mammalian nucleic acid molecule to produce a mammalian model system, the method comprising constructing a vector containing the mammalian nucleic acid molecule; introducing the vector into a totipotent mammalian embryonic stem cell; selecting an embryonic stem cell with the vector integrated into genomic DNA; microinjecting the selected cell into a mammalian blastocyst, thereby forming a chimeric blastocyst; transferring the chimeric blastocyst into a pseudopregnant dam, wherein the dam gives birth to a chimeric mammal containing at least one additional copy of mammalian nucleic acid molecule in its germ line; and breeding the chimeric mammal to generate a homozygous mammalian model system.

15 BRIEF DESCRIPTION OF THE SEQUENCE LISTING

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

20 The Sequence Listing contains the nucleic acid sequence of exemplary mammalian nucleic acid molecules of the invention, SEQ ID NOs:1-45; the protein sequence of exemplary mammalian protein molecules of the invention, SEQ ID NOs:46-49.

BRIEF DESCRIPTION OF THE FIGURES AND TABLE

25 Figures 1A, 1B, 1C, 1D, 1E, and 1F show the human nucleic acid molecule (SEQ ID NO:44) encoding the human amino acid sequence (SEQ ID NO:47) of the mammalian protein. The alignment was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figures 2A, 2B, 2C, 2D, 2E, and 2F show the human nucleic acid molecule (SEQ ID NO:43) encoding the human amino acid sequence (SEQ ID NO:46) of the mammalian protein. The alignment was 30 produced using MACDNASIS PRO software (Hitachi Software Engineering)

Figures 3A, 3B, 3C, and 3D demonstrate the chemical and structural similarity among human ASP2, clone ID 1869869 (SEQ ID NO:47), human ASP1, clone ID 1611565 (SEQ ID NO:46), human gastricsin, GI 1658286 (SEQ ID NO:48), *Neurospora crassa* vacuolar protease, GI 1039445 (SEQ ID NO:49) proteins, produced using the MEGALIGN program of LASERGENE software (DNASTAR,

WO 01/46398

PCT/US00/33743

Madison WI). Conserved aspartyl protease active sites (D-S/T)-G-S/T-S/T) are double underlined.

Figures 4A, 4B, 4C, 4D, 4E, 4F, 4G, and 4H demonstrate the chemical and structural similarity among the nucleic acids encoding human ASP2, clone ID 1869869 (SEQ ID NO:44) and nucleic acid variants, 1438613F1 (SEQ ID NO:22), 1438719H1 (SEQ ID NO:27), 1438613H1 (SEQ ID NO:28), 5 5094633H1 (SEQ ID NO:18), 2900063H1 (SEQ ID NO:25), 3484819H1 (SEQ ID NO:23), 2757870H1 (SEQ ID NO:24), 2123411H1 (SEQ ID NO:26), 5577513H1 (SEQ ID NO:17), 2908594H1 (SEQ ID NO:20), 2007269H1 (SEQ ID NO:21), and 2293393H1 (SEQ ID NO:19), produced using the SEQMAN II program of LASERGENE software (DNASTAR). Putative splice junction sites are indicated by (Δ).

Figures 5A, 5B, 5C, 5D, 5E, 5F and 5G demonstrate the chemical and structural similarity among 10 the nucleic acids encoding human ASP1, 1611565CB1 (SEQ ID NO:43) and nucleic acid variants, 5157179F6 (SEQ ID NO:29), 3985758H1 (SEQ ID NO:30), 2846604H1 (SEQ ID NO:31), 5743028H1 (SEQ ID NO:32), 4148749H1 (SEQ ID NO:34), 1855226H1 (SEQ ID NO:35), 1855420H1 (SEQ ID NO:36), 5114558H1 (SEQ ID NO:37), 839538R1 (SEQ ID NO:38), 4999662H1 (SEQ ID NO:40), 5588490H1 (SEQ ID NO:39), 3630054H1 (SEQ ID NO:41), 4108264H1 (SEQ ID NO:42), and 15 3751907H1 (SEQ ID NO:33), produced using the SEQMAN II program of LASERGENE software (DNASTAR). Putative splice junction sites are indicated by (Δ).

Figures 6A, 6B, 6C, 6D, and 6E demonstrate the chemical and structural similarity among human, 1611565.comp (SEQ ID NO:43); and rat, 701918575H1 (SEQ ID NO:6) and 700921566H1 (SEQ ID NO:7), nucleic acid sequences, produced using Phrap software (Phil Green, University of Washington, 20 Seattle WA).

Figures 7A, 7B, 7C, 7D, 7E, 7F, and 7G demonstrate the chemical and structural similarity among human, 1869569.comp (SEQ ID NO:44); rat, 701911752H1 (SEQ ID NO:10), 70190708H1 (SEQ ID NO:9), 700494841H1 (SEQ ID NO:8), 701596150H1 (SEQ ID NO:12), 701031389H1 (SEQ ID NO:14) 701823687H1 (SEQ ID NO:11), 701188357H1 (SEQ ID NO:5), 700769257H1 (SEQ ID NO:4), 25 700769871H1 (SEQ ID NO:3), 700251567H1.comp (SEQ ID NO:1), 701432228T1.comp (SEQ ID NO:2), 700542176H1 (SEQ ID NO:13), and 700376870H1 (SEQ ID NO:15); and monkey, 700709032H1.comp (SEQ ID NO:16), nucleic acid sequences, produced using Phrap software (Green, supra). Putative splice junction sites are indicated by (Δ).

30 MODES FOR CARRYING OUT THE INVENTION

It is understood that this invention is not limited to the particular machines, materials and methods described. 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. As used herein, the singular forms "a," "an," and "the" include plural

WO 01/46398

PCT/US00/33743

reference unless the context clearly dictates otherwise. For example, a reference to "a host cell" includes a plurality of such host cells known to those skilled in the art.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. 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.

10 Definitions

"Aspartyl protease" or "ASP", refers to a substantially purified protein obtained from any mammalian species, including murine, bovine, ovine, porcine, rodent, canine, simian, and preferably the human species, and from any source, whether natural, synthetic, semi-synthetic, or recombinant. "ASPs" refers to more than one ASP.

15 "Altered gene expression" and "altered expression" refers to the increase or decrease of gene expression and the presence or absence of transcribed messenger RNA transcribed in a sample.

"Biologically active" refers to a protein having structural, immunological, regulatory, or chemical functions of a naturally occurring, recombinant or synthetic molecule.

"Complementary" refer to the natural hydrogen bonding by base pairing between purines and 20 pyrimidines. For example, the sequence A-C-G-T forms hydrogen bonds with its complements T-G-C-A or U-G-C-A. Two single-stranded molecules may be considered partially complementary, if only some of the nucleotides bond, or completely complementary, if nearly all of the nucleotides bond. The degree of complementarity between nucleic acid strands affects the efficiency and strength of the hybridization and amplification reactions.

25 "Derivative" refers to the chemical modification of a nucleic acid molecule or amino acid sequence. Chemical modifications can include replacement of hydrogen by an alkyl, acyl, or amino group or glycosylation, pegylation, or any similar process that retains or enhances biological activity or lifespan of the molecule or sequence.

"Fragment" refers to an Inocyte clone or any part of a nucleic acid molecule that retains a usable, 30 functional characteristic. Useful fragments include oligonucleotides which may be used in hybridization or amplification technologies or in regulation of replication, transcription or translation.

"Hybridization complex" refers to a complex between two nucleic acid molecules by virtue of the formation of hydrogen bonds between purines and pyrimidines.

"Ligand" refers to any molecule, agent, or compound that will bind specifically to a

WO 01/46398

PCT/US00/33743

complementary site on a nucleic acid molecule or protein. Such ligands stabilize or modulate the activity of nucleic acid molecules or proteins of the invention and may be composed of at least one of the following: inorganic and organic substances including nucleic acids, proteins, carbohydrates, fats, and lipids.

5 "Modulates" refers to a change in activity (biological, chemical, or immunological) or lifespan resulting from specific binding between a molecule and either a nucleic acid molecule or a protein.

"Nucleic acid molecule" refers to a nucleic acid, oligonucleotide, nucleotide, polynucleotide or any fragment thereof. It may be DNA or RNA of genomic or synthetic origin, double-stranded or single-stranded, and combined with carbohydrate, lipid, protein or other materials to perform a particular activity such as transformation or form a useful composition such as a peptide nucleic acid (PNA).

10 "Oligonucleotide" is substantially equivalent to the terms amplicon, primer, oligomer, element, target, and probe and is preferably single stranded. The nucleic acid molecules may be splice variants of another nucleic acid molecule.

"Protein" refers to an amino acid sequence, oligopeptide, peptide, polypeptide or portions thereof whether naturally occurring or synthetic.

"Portion", as used herein, refers to any part of a protein used for any purpose, but especially for the screening of a library of molecules that specifically bind to that portion, or for the production of antibodies. An exemplary portion is the first 40 contiguous amino acids of SEQ ID NO:46.

"Similarity" as applied to polynucleotide sequences, refers to the quantified 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.

A suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI), Bethesda MD. BLAST, Basic Local Alignment Search Tool (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410; Altschul (1993) *J. Mol. Evol.* 36:290-300), is available from NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) and several other sources. 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. 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 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

WO 01/46398

PCT/US00/33743

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; and Filter: on.

Similarity 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.

10 "Sample" is used in its broadest sense. A sample containing nucleic acid molecules may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; genomic DNA, RNA, or cDNA in solution or bound to a substrate; a cell; a tissue; a tissue print or fingerprint; and the like.

"Substantially purified" refers to nucleic acid molecules or proteins that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free, from other components with which they are naturally associated.

"Substrate" refers to any rigid or semi-rigid support to which nucleic acid molecules or proteins are bound and includes membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, capillaries or other tubing, plates, polymers, and microparticles with a variety of surface forms including wells, trenches, pins, channels and pores.

"Variant" refers to nucleic acid molecules that are splice variants of a nucleic acid molecule that encodes ASP and which may be determined by BLAST score, wherein the score is at least 100, and most preferably at least 400.

25 The Invention

The invention is based on the discovery of new mammalian nucleic acid molecules that encode ASPs and nucleic acid molecule variants thereof, and on the use of the nucleic acid molecules, or fragments thereof, and protein, or portions thereof, as compositions in the characterization, diagnosis, treatment, or prevention of conditions such as Alzheimer's disease and Down syndrome.

30 Nucleic acids encoding ASP1 of the present invention were first identified in Incyte Clone 1611565H1 from the colon tissue cDNA library (COLNTUT06) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:43, was derived from Incyte Clones 1863920X321F1 (PROSNOT19), 2233901H1 (PANCTUT02), 1611565T1 (COLNTUT06), 336541H1 (GCSJHET02), and the shotgun sequence SBAA03902F1.

WO 01/46398

PCT/US00/33743

Nucleic acids encoding the ASP2 were first identified in Incyte Clone 1869869H1 from the skin tissue cDNA library (SKINBIT01) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:44, was derived from Incyte Clones 1869869F6 (SKINBIT01), 2549562F6 (LUNGNOT06), 3338603H1 (SPLNNOT10), and the shotgun sequences 70317925D1, 70317570D1, and 70319214D1.

Nucleic acid molecules encoding the mammalian aspartyl protease and nucleic acid molecule variants of the present invention were first identified by comparing nucleic acid sequences and amino acid sequences encoding ASP1 of the present invention and ASP2 with nucleic acid sequences in the LIFESEQ and ZOOSEQ databases (Incyte Genomics, Palo Alto CA).

ASP comprises the amino acid sequences of SEQ ID NO:46, 423 amino acids in length, and of SEQ ID NO:47, 322 amino acids in length. The following chemical and structural characterization of the ASP1 (SEQ ID NO:46) will be based on absolute positions (30 residues per line) shown in Figures 3A, 3B, 3C, and 3D.

The human ASP1 protein (SEQ ID NO:46) has two potential N-glycosylation sites at N75 and N271; a leucine zipper site at residue K102; nine casein kinase II phosphorylation sites at residues A29, H33, L84, L114, S118, Q251, L315, I354, and C367; and five protein kinase C phosphorylation sites at residues D37, T42, K51, L272, and E275. In addition, SEQ ID NO:46 has two eukaryotic and viral aspartyl protease sites at residues P12 and Y205. SEQ ID NO:46 shares 43.5% similarity with SEQ ID NO:47, 18.8% similarity with human gastricsin (SEQ ID NO:48; GI 1658286), and 14.1% similarity with *Neurospora crassa* vacuolar protease (SEQ ID NO:49; GI 1039445). Percent similarity was calculated using the MEGALIGN program of LASERGENE software (DNASTAR). The amino acids of SEQ ID NO:46 from residue T35 to residue G47 are useful for antibody production and were selected using the PROTEAN program of LASERGENE software (DNASTAR). BLOCKS analysis ($p < 10^{-12}$) confirmed the identity of SEQ ID NO:46 as an aspartyl protease based on amino acid sequence identity from L40 through G55, G129 through T157, G187 through G196, A235 through L244, V341 through A364 of SEQ ID NO:46. In addition, PRINTS analysis ($p < 10^{-3}$) identifies two aspartyl protease signatures in SEQ ID NO:46 from residues I33 through V53 and G182 through K195, A235 through L244, V240 through D334 of SEQ ID NO:46.

SEQ ID NO:44 was used to identify human ASP2 variant nucleic acid molecules in the Incyte LIFESEQ database (Incyte Genomics). Incyte clones (libraries) 1499218H1 (SINTBST01), 1438719T6 (PANCNOT08), 3642853T6 (LUNGNOT34), 1499218F6 (SINTBST01), 1377178H1 and 1377178F1 (LUNGNOT10), 3642853F6 (LUNGNOT34), 5577513H1 (BRAPNOT04), 5094633H1 (EPIMNOT05), 2293393H1 (BRAINNOT1), 2908594H1 (THYMNOT05), 2007269H1 (TESTNOT03), 1438613F1 (PANCNOT08), 3484819H1 (KIDNNOT31), 2757870H1 (TIP1AZS08), 2900063H1 (DRGCNOT01),

WO 01/46398

PCT/US00/33743

2123411H1 (BRSTNOT07), 1438719H1 (PANCNOT08), and 1438613H1 (PANCNOT08), SEQ ID NOs:17-28, respectively; are variants of the consensus sequence, SEQ ID NO:44 (Figures 4A, 4B, 4C, 4D, 4E, 4F, 4G, and 4H). Putative splice junction sites are indicated on the figures by (Δ). These clones were expressed in libraries with the following tissue, organ, or system associations: reproductive (3/15),
 5 cardiovascular (2/15), digestive system (4/15), neurological (3/15), hematopoietic (2/15), and urologic (1/15) tissues.

SEQ ID NO:43 was used to identify human ASP1 variant nucleic acid molecules in the Incyte LIFEBEQ database (Incyte Genomics). Incyte clones (libraries) 5157179F6 (BRSTMT02), 3985758H1 (UTRST05), 2846604H1 (DRGLN01), 5743028H1 (LUNGNON03), 3751907H1 (UTRSNOT18),
 10 4148749H1 (SINITUT04), 1855226H1 and 1855420H1 (HNT3AZT01), 5114558H1 (ENDITXT01), 839538R1 (PROSTUT05), 5588490H1 (ENDINOT02), 4999662H1 (MYEPTXT02), 3630054H1 (COLNNOT38), and 4108264H1 (PROSBPT07), SEQ ID NOs:29-42, respectively; are variants of the consensus sequence, SEQ ID NO:43 (Figures 5A, 5B, 5C, 5D, 5E, 5F and 5G). Putative splice junction sites are indicated on the figures by (Δ). These clones were expressed in libraries with the following cell,
 15 tissue, organ, or system associations: reproductive (5/15), cardiovascular (4/15), digestive system (2/15), thoracic ganglion (1/15), hematopoietic (2/15) tissues, and teratocarcinoma cell line (2/15).

Additional nucleic acid molecules encoding the mammalian aspartyl protease of the present invention were identified by using BLAST or BLAST2 (Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402) against the ZOOSEQ database (Incyte Genomics) to align rat and monkey clones with the human
 20 aspartyl protease (SEQ ID NOs:43 and 44). The clones (from various libraries): 700251567H1 (RABMNOT01), 701432228T1 (RALITXT06), 700769871H1 (RAHYNOT02), 700769257H1 (RAHYNOT02), 701188357H1 (RACONON02), 701918575H1 (RABYUNS03), 700921566H1 (RANPNOT02), 700494841H1 (RABCNOT01), 700885432H1 (RAVANOT01), 701901708H1 (RABYUNS09), 701911752H1 (RABYUNS03), 701823687H1 (RAKITXT08), 701596150H1
 25 (RALITXT26), 700542176H1 (RACONOT01), 701031389H1 (RABMNON02), 700376870H1 (RALANOT01), and 700709032H1 (MNBFNOT01) presented in the Sequence Listing as SEQ ID NOs:1-16, respectively; are non-human variants of the nucleic acid molecules, SEQ ID NOs:43 and 44, which encode ASPs (Figures 6A, 6B, 6C, 6D, 6E and Figures 7A, 7B, 7C, 7D, 7E, 7F, 7G, respectively). Putative splice junction sites are indicated on the figures by (Δ). These clones were expressed in libraries
 30 with the following tissue, organ, or system associations: neurological (10/16), digestive system (4/16), cardiovascular (1/16), and urologic (1/16) tissues.

The nucleic acid sequences, SEQ ID NOs:17-42 may be used in hybridization and amplification technologies to identify and distinguish among SEQ ID NOs:43 and 44 and similar molecules in a sample. The molecules may be used to mimic human conditions, diseases, or disorders, produce transgenic animal

WO 01/46398

PCT/US00/33743

models for these conditions, or to monitor animal toxicology studies, clinical trials, and subject/patient treatment profiles.

Characterization and Use of the Invention

5 cDNA libraries

In a particular embodiment disclosed herein, mRNA was isolated from mammalian cells and tissues using methods that are well known to those skilled in the art and used to prepare the cDNA libraries. The Incyte clones listed above were isolated from mammalian cDNA libraries. At least one library preparation representative of the invention is described in the EXAMPLES below. The consensus mammalian
10 sequences were chemically and/or electronically assembled from fragments including Incyte clones and extension and/or shotgun sequences using computer programs such as Phrap (Green, supra), GELVIEW Fragment Assembly system (Genetics Computer Group, Madison WI), and AUTOASSEMBLER application (Applied Biosystems, Foster City CA). Clones, extension and/or shotgun sequences are electronically assembled into clusters and/or master clusters.

15 Sequencing

Methods for sequencing nucleic acids are well known in the art and may be used to practice any of the embodiments of the invention. These methods employ enzymes such as the Klenow fragment of DNA polymerase I, SEQUENASE, Taq DNA polymerase and thermostable T7 DNA polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such
20 as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD).

Preferably, sequence preparation is automated with machines such as the HYDRA microdispenser (Robbins Scientific, Sunnyvale CA), MICROLAB 2200 system (Hamilton, Reno NV), and the DNA ENGINE thermal cycler (PTC200; MJ Research, Watertown MA). Machines used for sequencing include the ABI PRISM 3700, 377 or 373 DNA sequencing systems (Applied Biosystems), the MEGABACE 1000
25 DNA sequencing system (Amersham Pharmacia Biotech), and the like. The sequences may be analyzed using a variety of algorithms which are well known in the art and described in Ausubel et al. (1997; Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7) and Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

Shotgun sequencing is used to generate more sequence from cloned inserts derived from multiple
30 sources. Shotgun sequencing methods are well known in the art and use thermostable DNA polymerases, heat-labile DNA polymerases, and primers chosen from representative regions flanking the nucleic acid molecules of interest. Incomplete assembled sequences are inspected for identity using various algorithms or programs such as CONSED (Gordon (1998) Genome Res. 8:195-202) which are well known in the art. Contaminating sequences including vector or chimeric sequences or deleted sequences can be removed or

WO 01/46398

PCT/US00/33743

restored, respectively, or organizing the incomplete assembled sequences into finished sequences.

Extension of a Nucleic Acid Sequence

The sequences of the invention may be extended using various PCR-based methods known in the art. For example, the XL-PCR kit (Applied Biosystems), nested primers, and commercially available cDNA or genomic DNA libraries may be used to extend the nucleic acid sequence. 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) to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to a target molecule at temperatures from about 55°C to about 68°C. When extending a sequence to recover regulatory elements, it is preferable to use genomic, rather than cDNA libraries.

Use of The Mammalian Nucleic Acid Molecule

Hybridization

The mammalian nucleic acid molecule and fragments thereof can be used in hybridization technologies for various purposes. A probe may be designed or derived from unique regions such as the 5' regulatory region or from a conserved motif such as the aspartyl protease signature and used in protocols to identify naturally occurring molecules encoding the mammalian protein, allelic variants, or related molecules. The probe may be DNA or RNA, is usually single stranded and should have at least 50% sequence identity to any of the nucleic acid sequences. The probe may comprise at least 16 contiguous nucleotides of a nucleic acid molecule. Hybridization probes may be produced using oligolabeling, nick translation, end-labeling, or PCR amplification in the presence of labeled nucleotide. A vector containing the nucleic acid molecule or a fragment thereof may be used to produce an mRNA probe *in vitro* by addition of an RNA polymerase and labeled nucleotides. These procedures may be conducted using commercially available kits such as those provided by Amersham Pharmacia Biotech.

The stringency of hybridization is determined by G+C content of the probe, salt concentration, and temperature. In particular, stringency can be increased by reducing the concentration of salt or raising the hybridization temperature. In solutions used for some membrane based hybridizations, addition of an organic solvent such as formamide allows the reaction to occur at a lower temperature. Hybridization can be performed at low stringency with buffers, such as 5xSSC with 1% sodium dodecyl sulfate (SDS) at 60°C, which permits the formation of a hybridization complex between nucleic acid sequences that contain some mismatches. Subsequent washes are performed at higher stringency with buffers such as 0.2xSSC with 0.1% SDS at either 45°C (medium stringency) or 68°C (high stringency). At high stringency, hybridization complexes will remain stable only where the nucleic acid molecules are completely complementary. In some membrane-based hybridizations, preferably 35% or most preferably 50%,

WO 01/46398

PCT/US00/33743

formamide can be added to the hybridization solution to reduce the temperature at which hybridization is performed, and background signals can be reduced by the use of other detergents such as Sarkosyl or Triton X-100 (Sigma Aldrich, St. Louis MO) and a blocking agent such as denatured salmon sperm DNA.

Selection of components and conditions for hybridization are well known to those skilled in the art and are reviewed in Ausubel (*supra*) and Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY.

Microarrays may be prepared and analyzed using methods known in the art. Oligonucleotides may be used as either probes or targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and single nucleotide polymorphisms. Such information may be used to determine gene function; to understand the genetic basis of a condition, disease, or disorder; to diagnose a condition, disease, or disorder; and to develop and monitor the activities of therapeutic agents. (See, e.g., Brennan *et al.* (1995) USPN 5,474,796; Schana *et al.* (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler *et al.* (1995) PCT application WO95/251116; Shalon *et al.* (1995) PCT application WO95/35505; Heller *et al.* (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller *et al.* (1997) USPN 5,605,662.)

Hybridization probes are also useful in mapping the naturally occurring genomic sequence. The probes may be hybridized to: 1) a particular chromosome, 2) a specific region of a chromosome, 3) artificial chromosome constructions such as human artificial chromosomes (HAC), yeast artificial chromosomes (YAC), bacterial artificial chromosomes (BAC), bacterial P1 constructions, or single chromosomes, or cDNA libraries made from any of these sources.

Expression

A multitude of nucleic acid molecules encoding ASP may be cloned into a vector and used to express the protein, or portions thereof, in host cells. The nucleic acid sequence can be engineered by such methods as DNA shuffling (Stemmer and Cramer (1996) USPN 5,830,721) and site-directed mutagenesis to create new restriction sites, alter glycosylation patterns, change codon preference to increase expression in a particular host, produce splice variants, extend half-life, and the like. The expression vector may contain transcriptional and translational control elements (promoters, enhancers, specific initiation signals, and polyadenylated 3' sequence) from various sources that have been selected for their efficiency in a particular host. The vector, nucleic acid molecule, and regulatory elements are combined using *in vitro* recombinant DNA techniques, synthetic techniques, and/or *in vivo* genetic recombination techniques well known in the art and described in Sambrook (*supra*, ch. 4, 8, 16 and 17).

A variety of host systems may be transformed with an expression vector. These include, but are not limited to, bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems transformed with baculovirus

WO 01/46398

PCT/US00/33743

expression vectors; plant cell systems transformed with expression vectors containing viral and/or bacterial elements, or animal cell systems (Ausubel et al., unit 16). For example, an adenovirus transcription/translation complex may be utilized in mammalian cells. After sequences are ligated into the E1 or E3 region of the viral genome, the infective virus is used to transform and express the protein in host cells. The Rous sarcoma virus enhancer or SV40 or EBV-based vectors may also be used for high-level protein expression.

Routine cloning, subcloning, and propagation of nucleic acid sequences can be achieved using the multifunctional pBLUESCRIPT vector (Stratagene, La Jolla CA) or pSPORT1 plasmid (Life Technologies). Introduction of a nucleic acid sequence into the multiple cloning site of these vectors disrupts the *lacZ* gene and allows colorimetric screening for transformed bacteria. 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.

For long term production of recombinant proteins, the vector can be stably transformed into cell lines along with a selectable or visible marker gene on the same or on a separate vector. After transformation, cells are allowed to grow for about 1 to 2 days in enriched media and then are transferred to selective media. Selectable markers, antimetabolite, antibiotic, or herbicide resistance genes, confer resistance to the relevant selective agent and allow growth and recovery of cells that successfully express the introduced sequences. Resistant clones identified either by survival on selective media or by the expression of visible markers, such as anthocyanins, green fluorescent protein (GFP), β glucuronidase, luciferase, and the like, may be propagated using culture techniques. Visible markers are also used to quantify the amount of protein expressed by the introduced genes. Verification that the host cell contains the desired mammalian nucleic acid molecule is based on DNA-DNA or DNA-RNA hybridizations or PCR amplification techniques.

The host cell may be chosen for its ability to modify a recombinant protein in a desired fashion. Such modifications include acetylation, carboxylation, glycosylation, phosphorylation, lipidation, acylation and the like. Post-translational processing that cleaves a "prepro" form may also be used to specify protein targeting, folding, and/or activity. Different host cells available from the ATCC (Manassas MD) which have specific cellular machinery and characteristic mechanisms for post-translational activities may be chosen to ensure the correct modification and processing of the recombinant protein.

30 Recovery of Proteins from Cell Culture

Heterologous moieties engineered into a vector for ease of purification include glutathione S-transferase (GST), calmodulin binding peptide (CBP), 6-His, FLAG, MYC, and the like. GST, CBP, and 6-His are purified using commercially available affinity matrices such as immobilized glutathione, calmodulin, and metal-chelate resins, respectively. FLAG and MYC are purified using commercially

WO 01/46398

PCT/US00/33743

available monoclonal and polyclonal antibodies. A proteolytic cleavage site may be located between the desired protein sequence and the heterologous moiety for ease of separation following purification. Methods for recombinant protein expression and purification are discussed in Ausubel (*supra*, unit 16) and are commercially available.

5 Chemical Synthesis of Peptides

Proteins or portions thereof may be produced not only by recombinant methods, but also by using chemical methods well known in the art. Solid phase peptide synthesis may be carried out in a batchwise or continuous flow process which sequentially adds α -amino- and side chain-protected amino acid residues to an insoluble polymeric support via a linker group. A linker group such as methylamine-derivatized poly(ethylene glycol) is attached to poly(styrene-co-divinylbenzene) to form the support resin. The amino acid residues are *N*- α -protected by acid labile Boc (*t*-butyloxycarbonyl) or base-labile Fmoc (9-fluorenylmethoxycarbonyl). The carboxyl group of the protected amino acid is coupled to the amine of the linker group to anchor the residue to the solid phase support resin. Trifluoroacetic acid or piperidine are used to remove the protecting group in the case of Boc or Fmoc, respectively. Each additional amino acid is added to the anchored residue using a coupling agent or pre-activated amino acid derivative, and the resin is washed. The full length peptide is synthesized by sequential deprotection, coupling of derivatized amino acids, and washing with dichloromethane and/or *N,N*-dimethylformamide. The peptide is cleaved between the peptide carboxy terminus and the linker group to yield a peptide acid or amide. (Novabiochem 1997/98 Catalog and Peptide Synthesis Handbook, San Diego CA pp. 51-520). Automated synthesis may also be carried out on machines such as the ABI 431A peptide synthesizer (Applied Biosystems). A protein or portion thereof may be substantially purified by preparative high performance liquid chromatography and its composition confirmed by amino acid analysis or by sequencing (Creighton (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY).

Preparation and Screening of Antibodies

Various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with ASP or any portion thereof. Adjuvants such as Freund's, mineral gels, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemacyanin (KLH), and dinitrophenol may be used to increase immunological response. The oligopeptide, peptide, or portion of protein used to induce antibodies should consist of at least about five amino acids, more preferably ten amino acids, which are identical to a portion of the natural protein. Oligopeptides may be fused with proteins such as KLH in order to produce the chimeric molecule.

Monoclonal antibodies may be prepared using any technique that provides for the production of antibodies by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kofler *et*

WO 01/46398

PCT/US00/33743

al. (1975) *Nature* 256:495-497; Kozbor et al. (1985) *J. Immunol. Methods* 81:31-42; Cote et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole et al. (1984) *Mol. Cell. Biol.* 62:109-120.)

Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce epitope specific single chain antibodies. Antibody fragments 5 which contain specific binding sites for epitopes of the mammalian protein may also be generated. For example, such fragments include, but are not limited to, F(ab)₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab)₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse et al. (1989) *Science* 246:1275-1281).

10 The mammalian protein or a portion thereof may be used in screening assays of phagemid or B-lymphocyte immunoglobulin libraries to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoassays 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 the protein and its specific antibody. A two-site, monoclonal-based 15 immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may also be employed (Fould (1998) *Immunochemical Protocols*, Humana Press, Totowa NJ).

Labeling of Molecules for Assay

A wide variety of labels and conjugation techniques are known by those skilled in the art and may 20 be used in various nucleic acid, amino acid, and antibody assays. Synthesis of labeled molecules may be achieved using Promega (Madison WI) or Amersham Pharmacia Biotech kits for incorporation of a labeled nucleotide such as ³²P-dCTP, Cy3-dCTP or Cy5-dCTP or amino acid such as ³⁵S-methionine. Nucleotides and amino acids may be directly labeled with a variety of substances including fluorescent, chemiluminescent, or chromogenic agents, and the like, by chemical conjugation to amines, thiols and other 25 groups present in the molecules using reagents such as BIODIPY or FITC (Molecular Probes, Eugene OR).

Diagnostics

The nucleic acid molecules, fragments, oligonucleotides, complementary RNA and DNA molecules, and PNAs may be used to detect and quantify altered gene expression, absence/presence vs. 30 excess, expression of mRNAs or to monitor mRNA levels during therapeutic intervention. Conditions, diseases or disorders associated with altered expression include epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple

WO 01/46398

PCT/US00/33743

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-Strausler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia, a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infectious of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas. The diagnostic assay may use hybridization or amplification technology to compare gene expression in a biological sample from a patient to standard samples in order to detect altered gene expression. Qualitative or quantitative

WO 01/46398

PCT/US00/33743

methods for this comparison are well known in the art.

For example, the nucleic acid molecule or probe may be labeled by standard methods and added to a biological sample from a patient under conditions for the formation of hybridization complexes. After an incubation period, the sample is washed and the amount of label (or signal) associated with hybridization complexes, is quantified and compared with a standard value. If the amount of label in the patient sample is significantly altered in comparison to the standard value, then the presence of the associated condition, disease or disorder is indicated.

In order to provide a basis for the diagnosis of a condition, disease or disorder associated with gene expression, a normal or standard expression profile is established. This may be accomplished by combining a biological sample taken from normal subjects, either animal or human, with a probe under conditions for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a substantially purified target sequence is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular condition is used to diagnose that condition.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies and in clinical trial or to monitor the treatment of an individual patient. Once the presence of a condition is established and a treatment protocol is initiated, diagnostic 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 a normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

Immunological Methods

Detection and quantification of a protein 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 is preferred, but a competitive binding assay may be employed (Coligan et al. (1997) Current Protocols in Immunology, Wiley-Interscience, New York NY; Pound, supra).

30

Therapeutics

Chemical and structural similarity, in the context of sequences and motifs, exists between regions of SEQ ID NOs:46 and 47 and other ASPs, human gastricsin (g1658286) and *Neurospora crassa* vacuolar protease A (g1039445). In addition, gene expression is closely associated with digestive system,

WO 01/46398

PCT/US00/33743

reproductive and brain tissues, and in the pons from Alzheimer's disease and appears to play a role in conditions such as Alzheimer's disease and Down syndrome. In the treatment of conditions associated with increased expression or activity, it is desirable to decrease expression or protein activity. In the treatment of conditions associated with decreased expression or activity, it is desirable to increase expression or protein activity.

Chemical and structural similarity, in the context of sequences, exists between regions of variant nucleic acid molecules SEQ ID NOs:1-42 and SEQ ID NOs:46 and 47. In addition, gene expression of the variant nucleic acid molecules is closely associated with digestive, reproductive, brain, and cardiovascular tissues, and appears to play a role in conditions such as Alzheimer's disease and Down syndrome.

In one embodiment, the mammalian protein or a portion or derivative thereof may be administered to a subject to treat or prevent a condition associated with altered expression or activity of the mammalian protein. Examples of such conditions include, but are not limited to, epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, cataplexy, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and

WO 01/46398

PCT/US00/33743

- gynecoanastia, dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma,
- 5 biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis,
- 10 hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas.
- 15 In another embodiment, a pharmaceutical composition comprising a substantially purified mammalian protein in conjunction with a pharmaceutical carrier may be administered to a subject to treat or prevent a condition associated with altered expression or activity of the endogenous protein including, but not limited to, those provided above.
- In a further embodiment, a ligand which modulates the activity of the mammalian protein may be
- 20 administered to a subject to treat or prevent a condition associated with altered lifespan, expression, or activity of the protein including, but not limited to, those listed above. In one aspect, an antibody which specifically binds the mammalian protein may be used as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue that express the mammalian protein.
- In an additional embodiment, a vector capable of expressing the mammalian protein or a portion or
- 25 derivative thereof may be administered to a subject to treat or prevent a condition associated with altered lifespan, expression, or activity of protein including, but not limited to, those described above.
- In a still further embodiment, a vector expressing the complement of the nucleic acid molecule or fragments thereof may be administered to a subject to treat or prevent a condition associated with altered lifespan, expression, or activity of the protein including, but not limited to, those described above.
- 30 Any of the nucleic acid molecules, complementary molecules and fragments thereof, proteins or portions thereof, vectors delivering these nucleic acid molecules or proteins, and their ligands may be administered in combination with other therapeutic agents. Selection of the agents for use in combination therapy may be made by one of ordinary skill in the art according to conventional pharmaceutical principles. A combination of therapeutic agents may act synergistically to affect prevention or treatment of
- 35 a particular condition at a lower dosage of each agent.

WO 01/46398

PCT/US00/33743

Modification of Gene Expression Using Nucleic Acids

Gene expression may be modified by designing complementary or antisense molecules (DNA, RNA, or PNA) to the control, 5', 3', or other regulatory regions of the mammalian gene. Oligonucleotides designed with reference to the transcription initiation site are preferred. Similarly, inhibition can be achieved using triple helix base-pairing that inhibits the binding of polymerases, transcription factors, or regulatory molecules (Geo *et al.* In: Huber and Carr (1994) *Molecular and Immunologic Approaches*, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary molecule may also be designed to block translation by preventing binding between ribosomes and mRNA. In one alternative, a library of nucleic acid molecules or fragments thereof may be screened to identify those which specifically bind a regulatory, nontranslated sequence.

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 at sites such as GUA, GUU, and GUC. Once such sites are identified, an oligonucleotide with the same sequence may be evaluated for secondary structural features that would render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing their hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary nucleic acids and ribozymes of the invention may be prepared via recombinant expression, *in vitro* or *in vivo*, or using solid phase phosphoramidite chemical synthesis. In addition, RNA molecules may be modified to increase intracellular stability and half-life by addition of flanking sequences at the 5' and/or 3' ends of the molecule or by the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. Modification is inherent in the production of PNAs and can be extended to other nucleic acid molecules. Either the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, and/or the modification of adenine, cytidine, guanine, thymine, and uridine with acetyl-, methyl-, thio- groups renders the molecule less available to endogenous endonucleases.

Screening Assays

The nucleic acid molecule encoding the mammalian protein may be used to screen a library of molecules or compounds for specific binding affinity. The libraries may be DNA molecules, RNA molecules, PNAs, peptides, proteins such as transcription factors, enhancers, repressors, and other ligands that regulate the activity, replication, transcription, or translation of the nucleic acid molecule in the biological system. The assay involves combining the mammalian nucleic acid molecule or a fragment thereof with the library of molecules under conditions allowing specific binding, and detecting specific binding to identify at least one molecule that specifically binds the nucleic acid molecule.

WO 01/46398

PCT/US00/33743

Similarly the mammalian protein or a portion thereof may be used to screen libraries of molecules or compounds in any of a variety of screening assays. The portion of the protein employed in such screening may be free in solution, affixed to an abiotic or biotic substrate (e.g. borne on a cell surface), or located intracellularly. Specific binding between the protein and molecule may be measured. Depending on the kind of library being screened, the assay may be used to identify DNA, RNA, or PNA molecules, agonists, antagonists, antibodies, immunoglobulins, inhibitors, peptides, proteins, drugs, or any other ligand, which specifically binds the protein. One method for high throughput screening using very small assay volumes and very small amounts of test compound is described in USPN 5,876,946, incorporated herein by reference, which screens large numbers of molecules for enzyme inhibition or receptor binding.

10 Purification of Ligand

The nucleic acid molecule or a fragment thereof may be used to purify a ligand from a sample. A method for using a mammalian nucleic acid molecule or a fragment thereof to purify a ligand would involve combining the nucleic acid molecule or a fragment thereof with a sample under conditions to allow specific binding, detecting specific binding, recovering the bound protein, and using an appropriate agent to separate the nucleic acid molecule from the purified ligand.

Similarly, the protein or a portion thereof may be used to purify a ligand from a sample. A method for using a mammalian protein or a portion thereof to purify a ligand would involve combining the protein or a portion thereof with a sample under conditions to allow specific binding, detecting specific binding between the protein and ligand, recovering the bound protein, and using an appropriate chaotropic agent to separate the protein from the purified ligand.

Pharmacology

Pharmaceutical compositions are those substances wherein the active ingredients are contained in an effective amount to achieve a desired and intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compound, the therapeutically effective dose may be estimated initially either in cell culture assays or in animal models. The animal model is also used to achieve a desirable concentration range and route of administration. Such information may then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or inhibitor that ameliorates the symptoms or condition. Therapeutic efficacy and toxicity of such agents may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED₅₀ (the dose therapeutically effective in 50% of the population) and LD₅₀ (the dose lethal to 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and it may be expressed as the ratio, LD₅₀/ED₅₀. Pharmaceutical compositions that exhibit large therapeutic indexes are preferred. The data obtained from cell culture assays and animal studies are used in formulating a range of dosage for human

WO 01/46398

PCT/US00/33743

use.

Model Systems

Animal models may be used as bioassays where they exhibit a toxic response similar to that of humans and where exposure conditions are relevant to human exposures. Mammals are the most common models, and most toxicity studies are performed on rodents such as rats or mice because of low cost, availability, and abundant reference toxicology. Inbred rodent strains provide a convenient model for investigation of the physiological consequences of under- or over-expression of genes of interest and for the development of methods for diagnosis and treatment of diseases. A mammal inbred to over-express a particular gene (for example, secreted in milk) may also serve as a convenient source of the protein expressed by that gene.

Toxicology

Toxicology is the study of the effects of agents on living systems. The majority of toxicity studies are performed on rats or mice to help predict the effects of these agents on human health. Observation of qualitative and quantitative changes in physiology, behavior, homeostatic processes, and lethality are used to generate a toxicity profile and to assess the consequences on human health following exposure to the agent.

Genetic toxicology identifies and analyzes the ability of an agent to produce genetic mutations. Genotoxic agents usually have common chemical or physical properties that facilitate interaction with nucleic acids and are most harmful when chromosomal aberrations are passed along to progeny. Toxicological studies may identify agents that increase the frequency of structural or functional abnormalities in progeny if administered to either parent before conception, to the mother during pregnancy, or to the developing organism. Mice and rats are most frequently used in these tests because of their short reproductive cycle which produces the number of organisms needed to satisfy statistical requirements.

Acute toxicity tests are based on a single administration of the agent to the subject to determine the symptomatology or lethality of the agent. Three experiments are conducted: 1) an initial dose-range-finding experiment, 2) an experiment to narrow the range of effective doses, and 3) a final experiment for establishing the dose-response curve.

Prolonged toxicity tests are based on the repeated administration of the agent. Rat and dog are commonly used in these studies to provide data from species in different families. With the exception of carcinogenesis, there is considerable evidence that daily administration of an agent at high-dose concentrations for periods of three to four months will reveal most forms of toxicity in adult animals.

Chronic toxicity tests, with a duration of a year or more, are used to demonstrate either the absence of toxicity or the carcinogenic potential of an agent. When studies are conducted on rats, a

WO 01/46398

PCT/US00/33743

minimum of three test groups plus one control group are used, and animals are examined and monitored at the outset and at intervals throughout the experiment.

Transgenic Animal Models

Transgenic rodents that over-express or under-express a gene of interest may be inbred and used to model human diseases or to test therapeutic or toxic agents. (See USPN 4,736,866; USPN 5,175,383; and USPN 5,767,337). In some cases, the introduced gene may be activated at a specific time in a specific tissue type during fetal development or postnatally. Expression of the transgene is monitored by analysis of phenotype or tissue-specific mRNA expression in transgenic animals before, during, and after challenge with experimental drug therapies.

10 Embryonic Stem Cells

Embryonic stem cells (ES) isolated from rodent embryos retain the potential to form an embryo. When ES cells are placed inside a carrier embryo, they resume normal development and contribute to all tissues of the live-born animal. ES cells are the preferred cells used in the creation of experimental knockout and knockin rodent strains. Mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and are grown under culture conditions well known in the art. Vectors for knockout strains contain a disease gene candidate modified to include a marker gene that disrupts transcription and/or translation *in vivo*. The vector is introduced into ES cells by transformation methods such as electroporation, liposome delivery, microinjection, and the like which are well known in the art. The endogenous rodent gene is replaced by the disrupted disease gene through homologous recombination and integration during cell division. Transformed ES cells are identified, and preferably 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.

ES cells are also used to study the differentiation of various cell types and tissues *in vitro*, such as neural cells, hematopoietic lineages, and cardiomyocytes (Bain *et al.* (1995) *Dev. Biol.* 168:342-357; Wiles and Keller (1991) *Development* 111:259-267; and Klug *et al.* (1996) *J. Clin. Invest.* 98:216-224). Recent developments demonstrate that ES cells derived from human blastocysts may also be manipulated *in vitro* to differentiate into eight separate cell lineages, including endoderm, mesoderm, and ectodermal cell types (Thomson (1998) *Science* 282:1145-1147).

30 Knockout Analysis

In gene knockout analysis, a region of a human disease gene candidate is enzymatically modified to include a non-mammalian gene such as the neomycin phosphotransferase gene (*neo*; Capecchi (1989) *Science* 244:1288-1292). The inserted coding sequence disrupts transcription and translation of the targeted gene and prevents biochemical synthesis of the disease candidate protein. The modified gene is

WO 01/46398

PCT/US00/33743

transformed into cultured embryonic stem cells (described above), the transformed cells are injected into rodent blastulae, and the blastulae are implanted into pseudopregnant dams. Transgenic progeny are crossbred to obtain homozygous inbred lines.

Knockin Analysis

5 Totipotent ES cells, present in the early stages of embryonic development, can be used to create knockin humanized animals (pigs) or transgenic animal models (mice or rats) of human diseases. With knockin technology, a region of a human gene is injected into animal ES cells, and the human sequence integrates into the animal cell genome by recombination. Totipotent ES cells that contain the integrated human gene are handled as described above. Inbred animals are studied and treated to obtain information
10 on the analogous human condition. These methods have been used to model several human diseases. (See, e.g., Lee *et al.* (1998) *Proc. Natl. Acad. Sci.* 95:11371-11376; Baudoin *et al.* (1998) *Genes Dev.* 12:1202-1216; and Zhuang *et al.* (1998) *Mol. Cell Biol.* 18:3340-3349.)

Non-Human Primate Model

The field of animal testing deals with data and methodology from basic sciences such as
15 physiology, genetics, chemistry, pharmacology and statistics. These data are paramount in evaluating the effects of therapeutic agents on non-human primates as they can be related to human health. Monkeys are used as human surrogates in vaccine and drug evaluations, and their responses are relevant to human exposures under similar conditions. Cynomolgus monkeys (*Macaca fascicularis*, *Macaca mulatta*) and common marmosets (*Callithrix jacchus*) are the most common non-human primates (NHPS) used in these
20 investigations. Since great cost is associated with developing and maintaining a colony of NHPS, early research and toxicological studies are usually carried out in rodent models. In studies using behavioral measures such as drug addiction, NHPS are the first choice test animal. In addition, NHPS and individual humans exhibit differential sensitivities to many drugs and toxins and can be classified as "extensive metabolizers" and "poor metabolizers" of these agents.

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

30

EXAMPLES

It is to be understood that this invention is not limited to the particular machines, materials and methods described. Although particular embodiments are described, equivalent embodiments may be used to practice the invention. The described embodiments are not intended to limit the scope of the invention which is limited only by the appended claims. The examples below are provided to illustrate the subject

WO 01/46398

PCT/US00/33743

invention and are not included for the purpose of limiting the invention. For purposes of example, the preparation of the human brain cDNA library, BRAINON01, is described.

I Representative cDNA sequence preparation

5 The BRAINON01 normalized cDNA library was constructed from cancerous brain tissue obtained from a 26-year-old Caucasian male (specimen # 0003) during cerebral meningeal excision following diagnosis of grade 4 oligoastrocytoma localized in the right fronto-parietal part of the brain. Prior to surgery the patient was also diagnosed with hemiplegia, epilepsy, ptosis of eyelid, and common migraine. The frozen tissue was homogenized and lysed using a POLYTRON homogenizer (PT-3000; Brinkmann
10 Instruments, Westbury NY) in guanidinium isothiocyanate solution. The lysate was extracted with acid phenol, pH 4.7, per the RNA isolation protocol (Stratagene). The RNA was extracted with an equal volume of acid phenol, reprecipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in DEPC-treated water, and treated with DNase for 25 min at 37°C. RNA extraction and precipitation were repeated as before. The mRNA was isolated with the OLIGOTEX kit (Qiagen, Chatsworth CA) and
15 used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SUPERSKRIP[™] plasmid system (Life Technologies). cDNA synthesis was initiated with a NotI-oligo d(T) primer. Double-stranded cDNA was blunted, ligated to SalI adaptors, digested with NotI, and fractionated on a SEPHAROSE CL4B column (Amersham Pharmacia Biotech), and those cDNAs exceeding 400 bp were ligated into the NotI and
20 SalI sites of the PSPORT1 plasmid (Life Technologies). The plasmid was transformed into DH5 α competent cells (Life Technologies).

II Normalization, Isolation, and Sequencing of cDNA clones

4.9 x 10⁶ independent clones of the BRAINON01 plasmid library in *E. coli* strain DH12S (Life
25 Technologies) were grown in liquid culture under carbenicillin (25 mg/l) and methicillin (1 mg/ml) selection following transformation by electroporation. To reduce the number of excess cDNA copies according to their abundance levels in the library, the cDNA library was then normalized in a single round according to the procedure of Soares et al. (1994, Proc. Natl. Acad. Sci. 91:9228-9232), with the following modifications. The primer to template ratio in the primer extension reaction was increased from 2:1 to
30 10:1. The ddNTP concentration in this reaction was reduced to 150 μ M each ddNTP, allowing the generation of longer (400-1000 nt) primer extension products. The reannealing hybridization was extended from 13 to 48 hours. The single stranded DNA circles of the normalized library were purified by hydroxyapatite chromatography and converted to partially double-stranded by random priming, followed by electroporation into DH10B competent bacteria (Life Technologies).

WO 01/46398

PCT/US00/33743

Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Qiagen). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Life Technologies) with carbenicillin at 25 mg/l and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were prepared using a MICROLAB 2200 (Hamilton, Reno NV) in combination with DNAENGINE thermal cyclers (MJ Research, Watertown, MA). The cDNAs were sequenced by the method of Sanger and Coulson (1975, J Mol Biol 94:441-448) using ABI PRISM 377 DNA Sequencing systems (Applied Biosystems), and reading frame was determined.

III Identification, Extension, Assembly, and Analyses

The consensus sequence (SEQ ID NO:43) was assembled from Incyte clones (library) 1863920X321F1 (PROSNOT19), 2233901H1 (PANCYTUT02), I611565T1 (COLNTUT06), 336541H1 (EOSIHET02), and the shotgun sequence, SBAA03902F1, from the LIFESEQ database (Incyte Genomics) of human cDNA sequences and used to identify additional sequences in the LIFESEQ and ZOOSEQ databases (Incyte Genomics) related to ASPs. Translation of SEQ ID NO:43 and 44 using MACDNASIS PRO software (Hitachi Software Engineering) elucidated the coding regions, SEQ ID NO:46 and 47. The nucleic acid and amino acid sequences were queried against databases such as the GenBank databases, SwissProt, BLOCKS, PRINTS, Prosite, and PFAM using BLAST. Motifs and HMM algorithms were used to perform functional analyses, and the antigenic index (Jameson-Wolf analysis) was determined using LASERGENE software (DNASTAR).

25 IV Sequence Similarity

Sequence similarity was calculated as percent identity based on comparisons between at least two nucleic acid molecules or amino acid sequences using the clustal method of the MEGALIGN program of LASERGENE software (DNASTAR). The clustal method uses an algorithm that groups sequences into clusters by examining the distances between all pairs. After the clusters are aligned pairwise, they are realigned in groups. Percent similarity between two sequences, sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of very low or zero similarity between the two sequences are not included.

WO 01/46398

PCT/US00/33743

V Identification of Nucleic Acid Variants

Nucleic acid molecules which are splice variants of the nucleic acid molecules encoding the mammalian aspartyl protease (SEQ ID NOs:43 and 44) were identified by using BLAST or BLAST2 (Altschul, *supra*; NCI-BLASTN version 2.0.4 with default parameters), to identify clones in the LIFEBSEQ or ZOOSEQ database (Incyte Genomics) which aligned with SEQ ID NOs:43 and 44. Mammalian nucleic acid molecule variants were selected by BLAST score. The BLAST score is calculated by scoring +5 for every base that matches in a nucleic acid High Scoring Pair (HSP) and -4 for every mismatch. The BLAST alignments were visually inspected and those clones with BLAST scores greater than 100 were aligned together using Phrap (Green, *supra*). Examples of nucleic acid sequences that flank splice junction sites of splice variants of nucleic acid sequences encoding proteins are well known in the art. Splice variant nucleic acid molecules were identified as those which aligned with 100% identity in a part of the nucleic acid sequence, and with less than 100% identity in the remaining part of the nucleic acid sequence and in which the initial nucleotide 3' to the putative splice junction was a guanine base. A similar analysis of ESTs in the public domain identified clones that aligned with the nucleic acid sequences of human aspartyl protease and confirmed the location of the putative splice junctions.

VI Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled probe to a membrane on which RNAs from a particular cell type or tissue have been bound.

Analogous computer techniques applying BLAST2 (NCI-BLASTN version 2.0.4 with default parameters) were used to search for identical or related molecules in nucleotide databases such as the LIFEBSEQ database or ZOOSEQ database (both Incyte Genomics). Sequence-based analysis is much faster than membrane-based hybridization, and 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 BLAST search is the product score which is defined as: (percent sequence identity x percent maximum BLAST score) divided by 100. The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and with a product score of at least 70, the match will be exact. Similar or related molecules are usually identified by selecting those which show product scores between 8 and 40.

The results of northern analyses are reported as a fraction or a percentage distribution of libraries in which the transcript encoding the mammalian protein occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous,

WO 01/46398

PCT/US00/33743

reproductive, and urologic. The disease categories included cancer, inflammation/trauma, cell proliferation, and neurological. For each category, the number of libraries expressing the sequence was counted and divided by the total number of libraries identified containing nucleic acid molecule variants of the nucleic acid molecule encoding the mammalian protein.

5

VII Extension of Nucleic Acid Molecules

At least one of the nucleic acid molecules used to assemble SEQ ID NOs 43 and 44 was produced by extension of an Inocyte cDNA clone using oligonucleotide primers. One primer was synthesized to initiate 5' extension of the known fragment, and the other, to initiate 3' extension. The initial primers were
10 designed using OLIGO 4.06 software (National Biosciences) to be about 22 to 30 nucleotides in length, to have a GC content of about 50%, and to anneal to the target sequence at temperatures of about 55 °C to about 68 °C. Any fragment which would result in hairpin structures and primer-primer dimerizations was avoided. Selected human cDNA libraries were used to extend the molecule. If more than one extension is needed, additional or nested sets of primers are designed.

15 High fidelity amplification was obtained by performing PCR in 96-well plates using the DNA ENGINE thermal cycler (MJ Research). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair selected from the plasmid: Step 1: 94 °C, 3 min;
20 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, parameters for the primer pair, T7 and SK+ (Stratagene), 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.

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

The extended sequences were desalted, 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

WO 01/46398

PCT/US00/33743

digested fragments were separated on about 0.6-0.8% agarose gels, fragments were excised as visualized under UV light, and agar removed/digested with AGARACE (Promega). Extended fragments were religated using T4 DNA 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 transformed 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 carbenicillin liquid media.

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

In like manner, a nucleic acid molecule of SEQ ID NOs:1-44 is used to obtain regulatory sequences using the procedure above, oligonucleotides designed for outward extension, and a genomic DNA library.

20

VIII Labeling of Probes and Hybridization Analyses

Nucleic acids are isolated from a biological source and applied to a substrate for standard hybridization protocols by one of the following methods. A mixture of target nucleic acids, a restriction digest of genomic DNA, is fractionated by electrophoresis through an 0.7% agarose gel in 1xTAE [Tris-acetate-ethylenediamine tetraacetic acid (EDTA)] running buffer and transferred to a nylon membrane by capillary transfer using 20x saline sodium citrate (SSC). Alternatively, the target nucleic acids are individually ligated to a vector and inserted into bacterial host cells to form a library. Target nucleic acids are arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane. The membrane is placed on bacterial growth medium, LB agar containing carbenicillin, and incubated at 37°C for 16 hours. Bacterial colonies are denatured, neutralized, and digested with proteinase K. Nylon membranes are exposed to UV irradiation in a STRATALINKER UV-crosslinker (Stratagene) to cross-link DNA to the membrane.

30

In the second method, target nucleic acids are amplified from bacterial vectors by thirty cycles of PCR using primers complementary to vector sequences flanking the insert. Amplified target nucleic acids

WO 01/46398

PCT/US00/33743

are purified using SEPHACRYL-400 beads (Amersham Pharmacia Biotech). Purified target nucleic acids are robotically arrayed onto a glass microscope slide (Corning Science Products, Corning NY). The slide is previously coated with 0.05% aminopropyl silane (Sigma-Aldrich) and cured at 110°C. The arrayed glass slide (microarray) is exposed to UV irradiation in a STRATALINKER UV-crosslinker (Stratagene).

5 cDNA probes are made from mRNA templates. Five micrograms of mRNA is mixed with 1 µg random primer (Life Technologies), incubated at 70°C for 10 minutes, and lyophilized. The lyophilized sample is resuspended in 50 µl of 1x first strand buffer (cDNA Synthesis systems; Life Technologies) containing a dNTP mix, [α -³²P]dCTP, dithiothreitol, and MMLV reverse transcriptase (Stratagene), and incubated at 42°C for 1-2 hours. After incubation, the probe is diluted with 42 µl dH₂O, heated to 95°C for
10 3 minutes, and cooled on ice. mRNA in the probe is removed by alkaline degradation. The probe is neutralized, and degraded mRNA and unincorporated nucleotides are removed using a PROBEQUANT G-50 MicroColumn (Amersham Pharmacia Biotech). Probes can be labeled with fluorescent markers, Cy3-dCTP or Cy5-dCTP (Amersham Pharmacia Biotech), in place of the radionucleotide, [α -³²P]dCTP.

Hybridization is carried out at 65°C in a hybridization buffer containing 0.5 M sodium phosphate
15 (pH 7.2), 7% SDS, and 1 mM EDTA. After the substrate is incubated in hybridization buffer at 65°C for at least 2 hours, the buffer is replaced with 10 ml of fresh buffer containing the probes. After incubation at 65°C for 18 hours, the hybridization buffer is removed, and the substrate is washed sequentially under increasingly stringent conditions, up to 40 mM sodium phosphate, 1% SDS, 1 mM EDTA at 65°C. To detect signal produced by a radiolabeled probe hybridized on a membrane, the substrate is exposed to a
20 PHOSPHORIMAGER cassette (Amersham Pharmacia Biotech), and the image is analyzed using IMAGEQUANT data analysis software (Amersham Pharmacia Biotech). To detect signals produced by a fluorescent probe hybridized on a microarray, the substrate is examined by confocal laser microscopy, and images are collected and analyzed using GEMTOOLS gene expression analysis software (Incyte Genomics).

25

IX Complementary Nucleic Acid Molecules

Molecules complementary to the nucleic acid molecule, or a fragment thereof, are used to detect, decrease, or inhibit gene expression. Although use of oligonucleotides comprising from about 15 to about 30 base pairs is described, the same procedure is used with larger or smaller fragments or their derivatives
30 (PNAs). Oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and SEQ ID NOs:1-43. To inhibit transcription by preventing promoter binding, a complementary oligonucleotide is designed to bind to the most unique 5' sequence, most preferably about 10 nucleotides before the initiation codon of the open reading frame. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the mRNA encoding the mammalian protein.

WO 01/46398

PCT/US00/33743

X Expression of the Mammalian Protein

Expression and purification of the mammalian protein are achieved using bacterial or virus-based expression systems. For expression in bacteria, cDNA is subcloned into a 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 bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express the mammalian protein upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression in eukaryotic cells is achieved by infecting *Spodoptera frugiperda* (Sf9) insect cells with recombinant baculovirus, *Autographica californica* nuclear polyhedrosis virus. The polyhedrin gene of baculovirus is replaced with the mammalian cDNA 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.

In most expression systems, the mammalian protein is synthesized as a fusion protein with, e.g., GST or a peptide epitope tag, such as FLAG, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST 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 the mammalian protein at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunofluorescence 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 (supra, unit 16). Purified mammalian protein obtained by these methods can be used directly in the following activity assay.

25

XI Functional Assays

Serine protease activity of ASP is measured by the hydrolysis of various peptide thiobenzyl ester substrates. The substrates are chosen to represent the different SP types (chymase, trypase, aspartase, etc.). Assays are performed at ambient temperature (~25°C) and contain an aliquot of ASP and the appropriate substrate in HEPES buffer, pH 7.5 containing 0.01M CaCl₂ and 8% dimethylsulfoxide. The reaction also contains 0.34 mM dithiopyridine which reacts with the thiobenzyl group that is released during hydrolysis and converts it to thiopyridone. The reaction is carried out in an optical cuvette, and the generation of thiopyridone is measured in a spectrophotometer by the absorption produced at 324 nm. The amount of thiopyridone produced in the reaction is proportional to the activity of ASP.

WO 01/46398

PCT/US00/33743

In the alternative, protein function is assessed by expressing the sequences encoding ASP at physiologically elevated levels in mammalian cell culture. The nucleic acid molecule is subcloned into PCMV SPORT vector (Life Technologies), which contains the strong cytomegalovirus promoter, and 5-10 μ g of the vector is transformed into an endothelial or hematopoietic human cell line using transformation methods well known in the art. An additional 1-2 μ g of a plasmid containing sequence encoding CD64-GFP (Clontech, Palo Alto CA) is co-transformed to provide a fluorescent marker to identify transformed cells using flow cytometry (FCM).

The influence of the introduced genes on expression can be assessed using purified populations of these transformed cells. Since CD64-GFP, which is expressed on the surface of transformed cells, binds to conserved regions of human immunoglobulin G (IgG), the transformed cells are separated using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA is purified from the cells and analyzed by hybridization techniques.

XII Production of ASP Specific Antibodies

ASP is purified using polyacrylamide gel electrophoresis and used to immunize mice or rabbits. Antibodies are produced using the protocols below. Alternatively, the amino acid sequence of ASP is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity. An immunogenic epitope, usually found near the C-terminus or in a hydrophilic region is selected, synthesized, and used to raise antibodies. Typically, epitopes of about 15 residues in length are produced using an ABI Model 431A peptide synthesizer (Applied Biosystems) using Fmoc-chemistry and coupled to KLH (Sigma-Aldrich) by reaction with N-malimidobenzoyl-N-hydroxysuccinimide ester to increase immunogenicity.

Rabbits are immunized with the epitope-KLH complex in complete Freund's adjuvant. Immunizations are repeated at intervals thereafter in incomplete Freund's adjuvant. After a minimum of seven weeks for mouse or twelve weeks for rabbit, antisera are drawn and tested for anti-peptide activity. Testing involves binding the peptide to plastic, blocking with 1% bovine serum albumin, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG. Methods well known in the art are used to determine antibody titer and the amount of complex formation.

XIII Purification of Naturally Occurring Protein Using Specific Antibodies

Naturally occurring or recombinant mammalian protein is substantially purified by immunoaffinity chromatography using antibodies specific for the protein. An immunoaffinity column is constructed by covalently coupling the antibody to CNBr-activated SEPHAROSE resin (Amersham Pharmacia Biotech). Media containing the protein is passed over the immunoaffinity column, and the column is washed using high ionic strength buffers in the presence of detergent to allow preferential

WO 01/46398

PCT/US00/33743

absorbance of the protein. After coupling, the protein is eluted from the column using a buffer of pH 2-3 or a high concentration of urea or thiocyanate ion to disrupt antibody/protein binding, and the protein is collected.

5 **XIV Screening Molecules for Specific Binding with the Nucleic Acid Molecule or Protein**

The nucleic acid molecule, or fragments thereof, or the protein, or portions thereof, are labeled with ³²P-dCTP, Cy3-dCTP, or Cy5-dCTP (Amersham Pharmacia Biotech), or with BIODIPY or FITC (Molecular Probes, Eugene OR), respectively. Libraries of candidate molecules previously arranged on a substrate are incubated in the presence of labeled nucleic acid molecule or protein. After incubation under 10 conditions for either a nucleic acid or amino acid sequence, the substrate is washed, and any position on the substrate retaining label, which indicates specific binding or complex formation, is assayed, and the binding molecule is identified. Data obtained using different concentrations of the nucleic acid or protein are used to calculate affinity between the labeled nucleic acid or protein and the bound molecule.

15 **XV Demonstration of Protein Activity**

ASP, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent (Bolton et al. (1973) Biochem. J. 133:529-539). Candidate ligand molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled ASP, washed, and any wells with labeled ASP complex are assayed. Data obtained using different concentrations of ASP are used to calculate values for the number, 20 affinity, and association of ASP with the candidate ligand molecules.

WO 01/46398

PCT/US00/33743

CLAIMS

What is claimed is:

1. A substantially purified mammalian nucleic acid molecule and fragments thereof encoding ASPs, variants, and portions thereof, selected from:
 - 5 a) an amino acid sequence of SEQ ID NO:46;
 - b) a naturally occurring amino acid sequence having a BLAST score of at least 250 bits when compared to the sequence of SEQ ID NO:46, wherein the BLAST score is calculated using NCI-BLASTX version 2.0.4;
 - c) a biologically active fragment of the amino acid sequence of SEQ ID NO:46; and
 - 10 d) an immunologically active fragment of the amino acid sequence of SEQ ID NO:46.
2. An isolated and purified mammalian nucleic acid molecule or the complement thereof selected from:
 - a) a nucleic acid sequence of SEQ ID NOS:1-5, 7-29, 31-33, 37, and 43;
 - b) a naturally occurring nucleic acid sequence having a BLAST score of at least 200 bits when
 - 15 compared to a sequence of SEQ ID NOS:1-43, wherein the BLAST score is calculated using NCI-BLASTN version 2.0.4; and
 - c) a naturally occurring nucleic acid sequence having a BLAST score of between 200 and 400 bits when compared to a sequence of SEQ ID NO:44, wherein the BLAST score is calculated using NCI-BLASTN version 2.0.4.
- 20 3. A fragment of the mammalian nucleic acid sequence of claim 2(a).
4. A probe comprising at least 16 contiguous nucleic acids which hybridizes under high stringency conditions to the mammalian nucleic acid molecule of claim 2 or a complement or a fragment thereof.
5. A recombinant nucleic acid molecule comprising a promoter operably linked to the mammalian nucleic acid molecule of claim 2.
- 25 6. A cell transformed with the recombinant nucleic acid molecule of claim 5.
7. A method of producing a polypeptide, the method comprising:
 - a) culturing the cell of claim 6 under conditions for expression of the polypeptide, and
 - b) recovering the polypeptide so expressed.
8. A transgenic organism comprising the recombinant nucleic acid molecule of claim 5.
- 30 9. A method for detecting the mammalian nucleic acid molecule in a sample, the method comprising:
 - a) hybridizing the sample with the probe of claim 2 to form a hybridization complex; and
 - b) detecting the hybridization complex, wherein the hybridization complex indicates the presence of the mammalian nucleic acid molecule in the sample.
10. The method of claim 9 further comprising amplifying the nucleic acid molecule or a fragment

WO 01/46398

PCT/US00/33743

thereof prior to hybridization.

11. A method of using the mammalian nucleic acid molecule or a fragment thereof to screen a library of molecules to identify at least one ligand that specifically binds the mammalian nucleic acid molecule, the method comprising:
- 5 a) combining the mammalian nucleic acid molecule of claim 2 with the library of molecules under conditions to allow specific binding, and
- b) detecting specific binding, thereby identifying a ligand that specifically binds the mammalian nucleic acid molecule.
12. The method of claim 11 wherein the library is selected from DNA molecules, RNA molecules,
10 PNAs, peptides, and proteins.
13. A method of using the mammalian nucleic acid molecule or a fragment thereof to purify a ligand that specifically binds the mammalian nucleic acid molecule from a sample, the method comprising:
- a) combining the mammalian nucleic acid molecule or a fragment thereof of claim 2 with a sample under conditions to allow specific binding;
- 15 b) recovering the bound mammalian nucleic acid molecule; and
- c) separating the mammalian nucleic acid molecule from the ligand, thereby obtaining purified ligand.
14. A substantially purified protein or a portion thereof selected from:
- a) an amino acid sequence of SEQ ID NO:46;
- 20 b) a naturally occurring amino acid sequence having a BLAST score of at least 250 bits when compared to a sequence of SEQ ID NO:46, wherein the BLAST score is calculated using NCI-BLASTX version 2.0.4;
- c) a biologically active fragment of the amino acid sequence of SEQ ID NO:46; and
- d) an immunologically active fragment of the amino acid sequence of SEQ ID NO:46.
- 25 15. A method for using a protein or a portion thereof to screen a library of molecules to identify at least one ligand that specifically binds the protein, the method comprising:
- a) combining the protein or a portion thereof of claim 14 with the library of molecules under conditions to allow specific binding, and
- b) detecting specific binding, thereby identifying a ligand that specifically binds the protein.
- 30 16. The method of claim 15 wherein the library is selected from DNA molecules, RNA molecules, PNAs, peptides, proteins, agonists, antagonists, antibodies, immunoglobulins, inhibitors, drug compounds and pharmaceutical agents.
17. A method of using a protein or a portion thereof to purify a ligand that specifically binds the protein from a sample, the method comprising:

WO 01/46398

PCT/US00/33743

- a) combining the protein or a portion thereof of claim 14 with a sample under conditions to allow specific binding,
 - b) recovering the bound protein, and
 - c) separating the protein from the ligand, thereby obtaining purified ligand.
- 5 15. A pharmaceutical composition comprising an effective amount of the protein of claim 14 and a pharmaceutically acceptable excipient.
19. A method for treating a disease or condition associated with altered expression of ASP, comprising administering to a patient in need of such treatment the pharmaceutical composition of claim 18.
20. The method of claim 19, wherein the disease or condition is selected from Alzheimer's disease and
- 10 Down syndrome.

WO 01/46398

143

PCT/US00/33743

```

11      20      29      38      47      56
5'  ACG CCT CAC ATA GGG AAT TTG GCC CTC GAG GCA AGA AAT CGG CAG AGC TCA GTG
-----
65      74      83      92     101     110
CCT TTA GAT GGT CTT TCC DMA GGA TCC DCC AAG CCT TCA AAT TTG CCT TCC
-----
119     128     137     146     155     164
AGT TTA GAC AGG AAG CTG GCC TCT GCC TTT AAG GCC GTA TGA GSA CTC ACT CCT
-----
173     182     191     200     209     218
TGT CTC TTT CTG CCA AAG CCT GAC GAC TCC CTG GAG CCT TTC TTT GAC TCT CTG
-----
227     236     245     254     263     272
GTA AAG CAG ACC CAC GTT CCC AAC CTC TTC TCC CTG CTG CAG CTT TGT GGT GCT GGC
-----
281     290     299     308     317     326
TTC CCC CTC AAC CAG TCT GAA GTG CTG GGC TCT GTC GSA GGG AGC ATG ATC ATT
-----
F P L N Q S E V L A S V G G S M I I
-----
G L T P
-----

```

Figure 1A

WO 01/46398

243

PCT/US00/33743

```

335 344 353 362 371 380
GGA GGT ATC GAC CAC TCG CTC TAC ACA GGC AGT CTC TGG TAT ACA CCC ATC CGG
G G I D H S L Y T G S L W Y T P I R
-----
389 398 407 416 425 434
CGG GAG TGG TAT TAT GAG GTC ATC ATT GAG CGG GTG GAG ATC AAT GGA CAG GAT
R E W Y Y E V I I V R V E I N G Q D
-----
443 452 461 470 479 488
CTG AAA ATG GAC TGC AAG GAG TAC AAC TAT GAC AAG AGC ATT GTG GAC AGT GGC
L K M D C K E Y N Y D K S I V D S G
-----
497 506 515 524 533 542
ACC ACC AAC CTT CGT TTG CCC AAG AAA GTG TTT GAA GCT GCA GTC AAA TCC ATC
T T N L R L P K K V F E A A V K S I
-----
551 560 569 578 587 596
AAG GCA GCC TCC TCC ACG GAG AAG TTC OCT GAT GGT TTC TGG CTA GGA GAG CAG
K A A S S T E K F P D G F W L G E Q

```

Figure 1B

WO 01/46398

343

PCT/US00/33743

```

605 614 623 632 641 650
CTG CTG TGG CAA GCA GGC ACC CCT TGG AAC ATT TTC CCA GTC ATC TCA
L V C W Q A G T P P W N I F P V I S
659 668 677 686 695 704
CTC TAC CTA ATG GGT GAG GTT ACC AAC CAG TCC TTC CGC ATC ACC ATC CTT CCG
L Y L M G E V T N Q S F R I T I L P
713 722 731 740 749 758
CAG CAA TAC CTG CCG CCA GTG GAA GAT GTG GCC ACG TCC CAA GAC GAC TGT TAC
Q Q Y L R P V E D V A T S Q D D C Y
767 776 785 794 803 812
AAG TTT GCC ATC TCA CAG TCA TCC ACG GGC ACT GTT ATG GGA GCT GTT ATC ATG
K F A I S Q S S T G T V M G A V I M
821 830 839 848 857 866
GAG GGC TTC TAC GTT GTC TTT GAT CCG GCC CGA AAA CGA ATT GCC TTT GCT GTC
E G F Y V V F D R A R K R I G F A V

```

Figure 1C

WO 01/46398

443

PCT/US00/33743

```

875      884      893      902      911      920
AGC GCT TGC CAG GTG CAC GAT GAG TTC AGG ACC GCA GGG GTG GAA GGC CCT TTT
-----
S   A   C   H   V   H   D   E   F   R   T   A   A   V   E   G   P   F
-----
929      938      947      956      965      974
GTC ACC TTG GAC ATG GAA GAC TGT GGC TAC AAC ATT CCA CAG ACA GAT GAG TCA
-----
V   T   L   D   M   E   D   C   G   Y   N   I   P   Q   T   D   E   S
-----
983      992      1001      1010      1019      1028
ACC CTC ATG ACC ATA GCC TAT GTC ATG GCT GCC ATC TGC GCC CTC TTC ATG CTG
-----
T   L   M   T   I   A   Y   V   M   A   A   I   C   A   L   F   M   L
-----
1037      1046      1055      1064      1073      1082
CCA CTC TGC CTC ATG GTG TGT CAG TGG CCG TGC CTC CCG TGC CTG CCG CAG CAG
-----
P   L   C   L   M   V   C   Q   W   R   C   L   R   C   L   R   Q   Q
-----
1091      1100      1109      1118      1127      1136
CAT GAT GAC TTT GCT GAT GAC ATC TCC CTG AAG TGA GGA GGC CCA TGG GCA
-----
H   D   D   F   A   D   D   I   S   L   L   K
-----
1145      1154      1163      1172      1181      1190
GAA GAT AGA GAT TCC CCT GGA CCA CAC CTC GGT GGT TCA CTT TGG TCA CAA GTA
-----

```

Figure 1D

WO 01/46398

543

PCT/US00/33743

```

1199      1208      1217      1226      1235      1244
GGA GAC ACA GAT GCC ACC TGT GGC CAG AGC ACC TCA GGA CCC TCC CCA CCC ACC
-----
1253      1262      1271      1280      1289      1298
AAA TGC CTC TGC CTT GAT GGA GAA GGA AAA GCC TGG CAA GGT GGG TTC CAG GGA
-----
1307      1316      1325      1334      1343      1352
CTG TAC CTG TAG GAA ACA GAA AAG AGA AGA AAG AAG CAC TCT GCT GGC GGG AAT
-----
1361      1370      1379      1388      1397      1406
ACT CTT GGT CAC CTC AAA TTT AAG TCG GGA AAT TCT GCT TGA AAC TTC ACC
-----
1415      1424      1433      1442      1451      1460
CCT GAA CCT TTG TCC ACC ATT CCT TTA AAT TCT CCA ACC CAA AGT ATT CTT CTT
-----
1469      1478      1487      1496      1505      1514
TTC TTA GTT TCA GAA GTA CTG GCA TCA CAC GCA GGT TAC CTT GCC GTG TGT CCC
-----
1523      1532      1541      1550      1559      1568
TGT GGT ACC CTG GCA GAG AAG AGA CCA AGC TTG TTT CCC TCC TGG CCA AAG TCA
-----

```

Figure 1E

WO 01/46398

643

PCT/US80/33743

```

1577 1586 1595 1604 1613 1622
CTA GGA GAG GAT GCA CAG TTT GCT ATT TGC TTT AGA GAC AGG GAC TGT ATA AAC
-----
1631 1640 1649 1658 1667 1676
AAG CCT AAC ATT GCT GCA AAG ATT GCC TCT TCA ATT AAA AAA AAC TAG ATT
-----
1685 1694 1703 1712 1721 1730
GAC TAT TTA TAC AAA TGC GGG CTG GAA AGA GGA GAA GGA GAG GGA GTA CAA
-----
1739 1748 1757 1766 1775 1784
AGA CAG GGA ATA GTG GGA TCA AAG CTA GGA AAG GCA GAA ACA CAA CCA CTC ACC
-----
1793 1802 1811 1820 1829 1838
AGT CCT AGT TTT AGA CCT CAT CTC CAA GAT AGC ATC CCA TCT CAG AAG ATG GGG
-----
1847 1856 1865 1874 1883 1892
TTT GTT TTT CAA TGG TCT CTT TTC CGT GGT GGC AGC CGG ACC AAA AGT GAG ATG
-----
1901 1910
GGG AAG GGC CTA TCT AGC C 3'
-----

```

Figure 1F

WO 01/46398

743

PCT/US00/33743

```

5'  AGC CTT AAT CTG GAC TGC AGA GAG TAT AAC GCA GAC AAG GCC ATC GTG GAC AAC      54
-----
          9          18          27          36          45          54
          AAT CTG GAC TGC AGA GAG TAT AAC GCA GAC AAG GCC ATC GTG GAC AAC
          63          72          81          90          99          108
          GGG GAC TCT GGC CGC GGC TAC TAC CTG GAG ATG CTG ATC GGG ACC CCC
          -----
          M L I G T P
          117          126          135          144          153          162
          CCG CAG AAG CTA CAG ATT CTC GTT GAC ACT GGA AGC AGT AAC TTT GCC GTG GCA
          P Q K L Q I L V D T G S S N F A V A
          171          180          189          198          207          216
          GGA ACC CCG CAC TCC TAC ATA GAC ACG TAC TTT GAC ACA GAG AGG TCT AGC ACA
          G T P H S Y I D T Y F D T E R S S T
          225          234          243          252          261          270
          TAC CGC TCC AAG GGC TTT GAC GTC ACA GTG AAG TRC ACA CAA GGA ACC TGG ACG
          Y R S K G F D V T V K Y T Q G S W T
  
```

Figure 2A

WO 01/46398

843

PCT/US00/33743

```

279  GGC TTC GTT GGG GAA GAC CTC CTC ACC ATC CCC AAA GGC TTC AAT ACT TCT TTT 324
-----
G F V G E D L V T I P K G F N T S F
-----
333  CTT GTC AAC ATT GCC ACT AAT TTT GAA TCA GAG AAT TTC TTT TTG CCT GGG ATT 378
-----
L V N I A T I F E S E N F F L P G I
-----
387  AAA TGG AAT GGA ATA CTT GGC CTA GCT TAT GCC ACA CTT GCC AAG CGA TCA AST 432
-----
K W N G I L G L A Y A T L A K P S S
-----
441  TCT CTG GAG ACC TTC TTC GAC TCC CTG GTG ACA CAA GCA AAC ATC CCC AAC GTT 486
-----
S L E T F F D S L V T Q A N I P N V
-----
495  TTC TCC ATG CAG ATG TGT GGA GCC GGC TTG CCC GTT GCT GGA TCT GGG ACC AAC 540
-----
F S M Q M C G A G L P V A G S G T N

```

Figure 2B

WO 01/46398

9/43

PCT/US00/33743

```

549 558 567 576 585 594
GGA GGT AGT CTT GTC TTG GGT GGA ATT GAA CCA AGT TTG TAT TAT GGA GAC ATC
-----
G G S L V L G G I E P S L Y K G D I
-----
603 612 621 630 639 648
TGG TAT ACC CCT AAT AAG GAA GAG TGG TAC TAC CAG ATA GAA ATT CTG AAA TTG
-----
W Y T P I K E E W Y Y Q I E I L K L
-----
657 666 675 684 693 702
GAA ATT GGA GGC CAA AGC CTT AAT CTG GAC TGC AGA GAG TAT AAC GCA GAC AAG
-----
E I G G Q S L N L D C R E Y N A D K
-----
711 720 729 738 747 756
GCC ATC GTG GAC AGT GGC ACC ACG CTG CTG CGC CTG CCC CAG AAG GTG TTT GAT
-----
A I V D S G T T L L R L P Q K V F D
-----
765 774 783 792 801 810
GGC GTG GTG GAA GCT GTG GCC CGC CCA TCT CTG ATT CCA GAA TTC TCT GAT GGT
-----
A V V E A V A R A S L I P E F S D G

```

Figure 2C

WO 01/46398

10/43

PCT/US00/33743

```

819      828      837      846      855      864
TTC TGG ACT GGG TCC CAG CTG GCG TGC TGG ACG AAT TCG GAA ACA CCT TGG TCT
F W T G S Q L A C W T N S E T P W S
-----
873      882      891      900      909      918
TAC TTC CCT AAA ATC TCC ATC TAC CTG AGA GAC GAG AAC TCC ACC AGG TCA TTC
Y F P K I S I Y L R D E N S S R S F
-----
927      936      945      954      963      972
CGT ATC ACA ATC CTG CCT CAG CTT TAC ATT CAG CCC ATG ATG GGG GCC GGC CTG
R I T I L P Q L Y I Q P M M G A G L
-----
981      990      999      1008      1017      1026
AAT TAT GAA TGT TAC CGA TTC GGC ATT TCC CCA TCC ACA AAT GCG CTG GTG ATC
N Y E C Y R F G I S P S T N A L V I
-----
1035     1044     1053     1062     1071     1080
GGT GCC ACG GTG ATG GAG GGC TTC TAC GTC ATC TTC GAC ACA GCC CAG AAG AGG
G A T V M E G F Y V I F D R A Q K R
-----

```

Figure 2D

WO 01/46398

11/43

PCT/US00/33743

```

1089      1098      1107      1116      1125      1134
GTG GGC TTC GCA GCG AGC CCC TGT GCA GAA ATT GCA GGT GCT GCA GTG TCT GAA
-----
V G F A A S P C A E I A G A A V S E
-----
1143      1152      1161      1170      1179      1188
ATT TCC GGG CCT TTC TCA ACA GAG GAT GTA GCC AGC AAC TGT GTC CCC GCT CAG
-----
I S G P F S T E D V A S N C V P A Q
-----
1197      1206      1215      1224      1233      1242
TCT TTG AGC GAG CCC ATT TTG TGG ATT GTG TCC TAT GCG CTC ATG AGC GTC TGT
-----
S L S E P I L W I V S Y A L M S V C
-----
1251      1260      1269      1278      1287      1296
GGA GCC ATC CTC CTT GTC TTA ATC GTC CTG CTG CTG CCG TTC CCG TGT CAG
-----
G A I L L V L I V L L L L L P F R C Q
-----
1305      1314      1323      1332      1341      1350
CGT CGC CCC CGT GAC CCT GAG GTC GTC AAT GAT GAG TCC TCT CTG GTC AGA CAT
-----
R R P R D P E V V N D E S S L V R H

```

Figure 2E

WO 01/46398

12/43

PCT/US00/33743

```

1359      1368      1377      1386      1395      1404
CGC TGG AAA TGA ATA GCC AGG CCT GAC CTC AAG CAA CCA TGA ACT CAG CTA TTA
R   W   K
-----
1413      1422      1431      1440      1449      1458
AGA AAA TCA CAT TTC CAG GGC AGC AGC GGG GAT CGA TGG TGG CGC TTT CTC CTG
-----
1467      1476      1485      1494      1503      1512
TGC CCA CCC GTC TTC AAT CTC TGT TCT GCT CCC AGA TGC CTT CTA GAT TCA CTG
-----
1521      1530      1539      1548      1557      1566
TCT TTT GAT TCT TGA TTT TCA AGC TTT CAA ATC CTC CCT ACT TCC AAG AAA AAT
-----
1575      1584      1593
AAT TAA AAA AAA AAC TTC ATT CTA AAC CAA AAA AA 3'
-----

```

Figure 2F

WO 01/46398

13/43

PCT/US80/33743

```

1  G  - - - - - 1869869
1  M  - - - - - 1611565
1  M  K W M V V V L V C L Q L L - - - E A A V V K V P L L K K - - - G I 1658286
1  M  K - - G A L L T A A M L L G S A Q A G V H T M K L K K V P G I 1039445
2  - - - - - 1869869
2  - - - - - 1611565
26  - - - - - F K S I R E T M K E K G L L G E F F - - - L R T H K Y D G I 1658286
29  L A D E L E S V P I D V Q V Q H L G Q K Y T G L R T E S H T G I 1039445
2  - - - - - 1869869
2  - - - - - 1611565
50  P A M K Y R F G D L S V T - Y E P M A - - - Y M D A A Y F G G I 1658286
59  Q A - M F K A T D A Q V S G N H P V P I T N F M N A Q Y F S G I 1039445
2  - - - - - 1869869
2  - - - - - 1611565
76  E I S I G T P P Q N F L V L F D T G S S N L W V P S V Y C Q G I 1658286
88  E I T I G T P P Q T F K V V L D T G S S N L W V P S S Q C G G I 1039445
6  - - - - - 1869869
26  T P H S Y I D T Y F D T E R S S T Y R S K G F D V T V K Y T 1611565
106  S Q A C T S H S R F N F S E S S T Y S T N G Q T F S L Q Y G G I 1658286
118  S I A C Y L H N K Y E S S E S S T Y K K N G T S F K I E Y G G I 1039445

```

Figure 3/A

WO 01/46398

14/43

PCT/US00/33743

6 - - - - - L F L P - - - - - K P D D S 1869869
56 Q G S M T G F V G E D L V T I P K G F N T S F L V N I A T I 1611565
136 S G S L T G F F G Y D T L T V Q S I Q V P N Q E F G L S E N G I 1658286
148 S G S L S G F V S Q D R M T I G D I T I N D Q L F A E A T S G I 1039445

6 - - - - - L F L P - - - - - K P D D S 1869869
86 F E S E N F F L P G I K W N G I L G L A Y A T L A K P S S S 1611565
166 E P G T N F V Y A - - Q F D G I M G L A Y P A L S V D E A T G I 1658286
178 E P G L A E A F G - - R F D G I L G L G Y D R L A V E G I T G I 1039445

15 L E L F F D S L V K Q T H V P N - L F S L Q L C G A G E P L 1869869
116 L E T F F D S L V T Q A N I P N - V F S M Q M C G A G L P Y 1611565
194 - - F A M Q G M V Q E G A L T S P V F S V Y L S N Q Q - - - G I 1658286
206 - - P P F Y K M V E Q K L V D E P P V F S F Y L A D Q D - - - G I 1039445

44 N Q S E V L A S V G G S M I I G G I D H S L Y T G S L W Y T 1869869
145 A G S - - - G T N G G S L V L G G I E P S L Y K G D I W Y T 1611565
219 - - - - - G S S G G A V V F G G V D S S L Y T T G Q I Y W A G I 1658286
231 - - - - - G E S E - - V V F G G V N K D R Y T T G K I T T I G I 1039445

74 P I R R E W Y Y E V I I I V R V E I N G Q D L K M D C K E Y N 1869869
172 P I K E E M Y Y Q I E I L K L E I G G Q S L N L D C R E Y N 1611565
243 P V T Q E L Y W Q I G I E E F L I G G Q A S G W - C S E - - G I 1658286
253 P L R R K A Y W E V D F D A I G Y G K D F A E L - - - E - - - G I 1039445

Figure 3B

WO 01/46398

15/43

PCT/IS00/33743

104 YDKSIVDSGTTNLRLPKKVFEEAVKSI-KA 1869869
202 ADKAIIVDSGTTLLRLPQKVFDAVVEAV-AR 1611565
270 GCQAIVDTGTTSLLTVPQYMSALLQAATGAQ GI 1658286
278 GHGVIIDTGTSLIALPSQLAEMLNAAQIGAK GI 1039445

133 ASSTEKFPDGGFWLGEQLVCWQAAGTTPWNI F 1869869
251 ASLIEFSDGFWTGSQLA CWTNS ETPWSY F 1611565
300 EDEYGGFFL-----VNCNSIQN-----L GI 1658286
308 KSNWNGQFT-----IDCGKKSS-----L GI 1039445

163 FVLSLYLMGEVTNQSFRIITLLPQQLYLR[P]VE 1869869
261 PKISLYLRDENSSRSFRITLLPQLYIQPMM 1611565
317 ESLTFIYNG-----VEFPLPSSYI LSNNGI 1658286
325 EDVTFTLAG-----YNFTLGEEDYILEAS GI 1039445

193 DVATSQDDCYKFAISQSSTGTVMGAVIMEG 1869869
291 GAGLNYE-CYRFGISPS TNALVI GATVMEG 1611565
34L GYCTVGV EPTYLSSQNGQPLWILDVFLRS GI 1658286
349 GSCLS TFMGM DMPAPVG-PLAILGD AFLRK GI 1039445

223 FVVVFDRARKRIGFAVSA[CHVHDEFR]TAAV 1869869
320 FVVFDRARQRVGF AASPCEIAGAAVSEI 1611565
37L YYSVYDILGNR RVGFATAA GI 1658286
378 YYSIYDILGADTVG IATAK-----GI 1039445

Figure 3C

WO 01/46398

16/43

PCT/US00/33743

253 E G P F V T L D M E D C G Y N I P Q T D E S T L M T I A Y V 1869869
 350 S G P F S T F D V A S N C V P A Q S L S E P I L W I V S Y A 1611565
 388 ----- GI 1658286
 396 ----- GI 1039445

 283 M A A I C A L F M L P L C - L M V C Q W R C L R C L R Q Q H 1869869
 380 L M S Y C G A I L L V L I V L L L L P F R C Q R R P R D P - 1611565
 388 ----- GI 1658286
 396 ----- GI 1039445

 312 D D F A D D I S L L - - - - K 1869869
 409 E V V N D E S S L V R H R W K 1611565
 388 ----- GI 1658286
 396 ----- R ----- GI 1039445

Figure 3D

WO 01/46398

1743

PCT/US00/33743

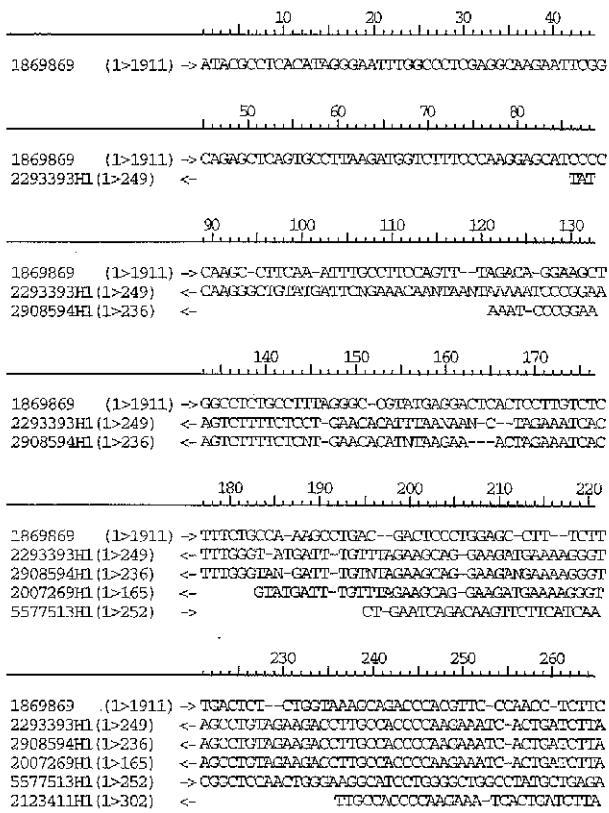


Figure 4A

WO 01/46398

18/43

PCT/US00/33743

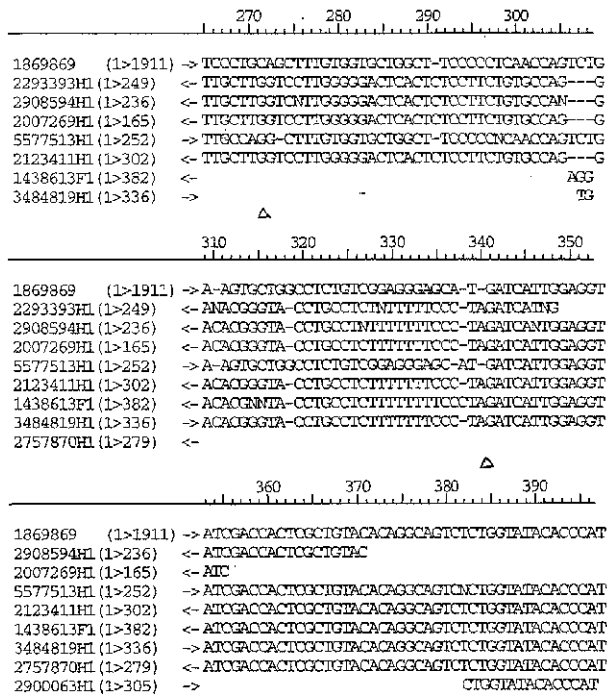


Figure 4B

WO 01/46398

19/43

PCT/US00/33743

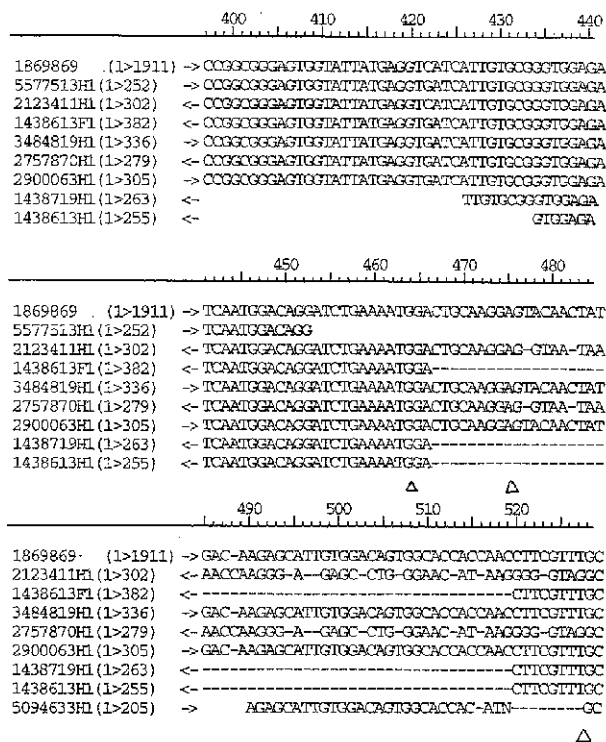


Figure 4C

WO 01/46398

20/43

PCT/US00/33743

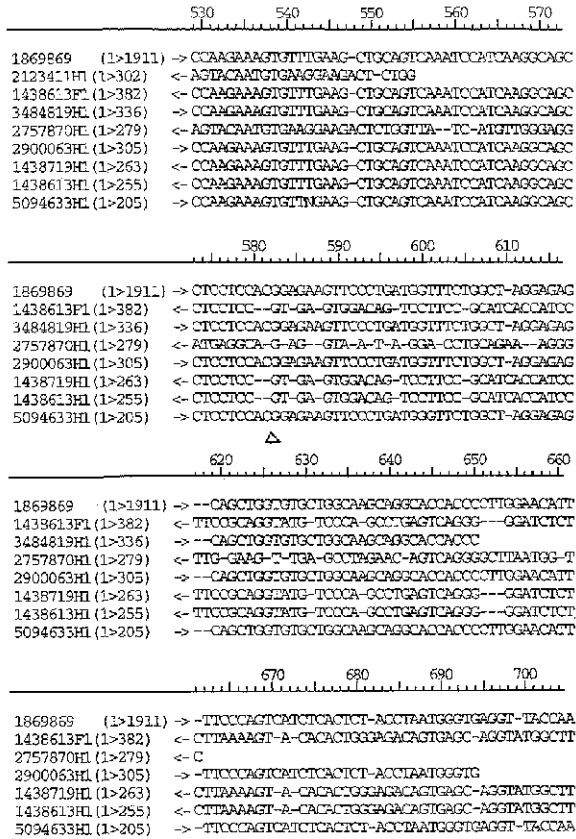


Figure 4D

WO 01/46398

21/43

PCT/US00/33743

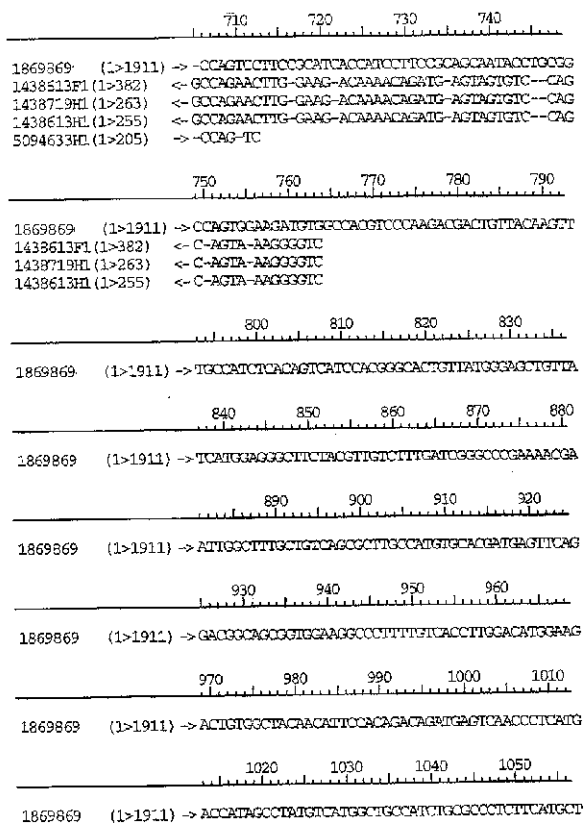


Figure 4E

WO 01/46398

22/43

PCT/US00/33743

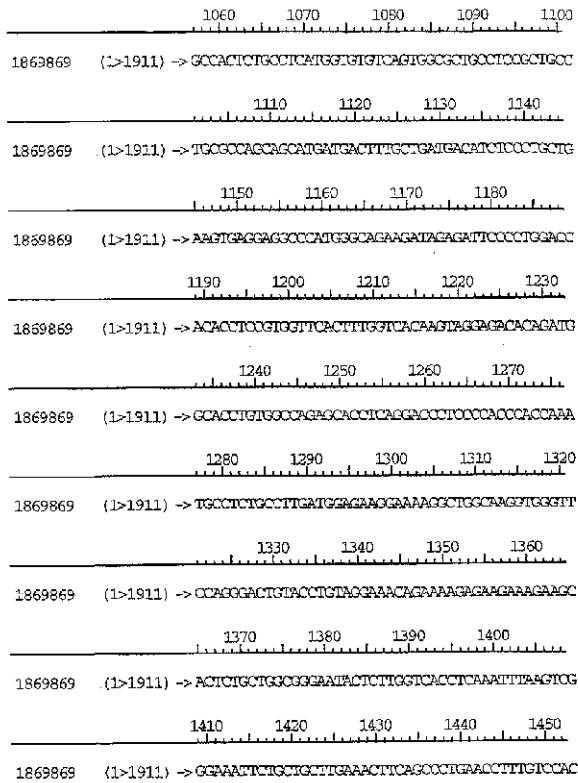


Figure 4F

WO 01/46398

23/43

PCT/US00/33743

```

      1460   1470   1480   1490
-----
1869869 (1>1911) -> CATTCCCTTAACTTCTCCACCCCAAAGTATTCTCTTPTCTTAG
      1500   1510   1520   1530   1540
-----
1869869 (1>1911) -> TTTCAGAGTACTGGCCATCACCCGAGTTACCTTGGGCTGTCT
      1550   1560   1570   1580
-----
1869869 (1>1911) -> CXCCTGTGGTACCTTGGCAGAGAGAGACCACACTTGTPTCCCTG
      1590   1600   1610   1620
-----
1869869 (1>1911) -> CTGGCCAAACTCAGTAGGAGAGGATGCCACAGTTGCCATTTGCT
      1630   1640   1650   1660   1670
-----
1869869 (1>1911) -> TTAGGACACGGGCTGTGTTTAAACAGCCCTACACTTGGTGCRAAG
      1680   1690   1700   1710
-----
1869869 (1>1911) -> ATTGCCCTCTTGAATTAATAAAAAAACTAGACTTACTATTTATA
      1720   1730   1740   1750   1760
-----
1869869 (1>1911) -> CAAATGGGGGGGCTGGAAAGAGGAGGACAGAGGAGTACAAA
      1770   1780   1790   1800
-----
1869869 (1>1911) -> GACAGGGAATAGTGGATCAAGCTAGGAAAGCCACAACTCAA
      1810   1820   1830   1840
-----
1869869 (1>1911) -> CCACTCACCAGTCTAGTGTTTAGACCTTCATCTCCAGATACCAT

```

Figure 4G

WO 01/46398

24/43

PCT/US00/33743

```

      1850      1860      1870      1880      1890
-----
1869869 (1>1911) -> CCATTCACAGAGATGGGGTTTGTTHTTTCAATGGTCTCTTTTCC
      1900      1910      1920      1930
-----
1869869 (1>1911) -> GTGGTGGCAGTCGGACCAAAAGTGAGATGGCGAAGGCGCTATCT
      1940
-----
1869869 (1>1911) -> ACC
```

Figure 4H

WO 01/46398

25/43

PCT/US00/33743

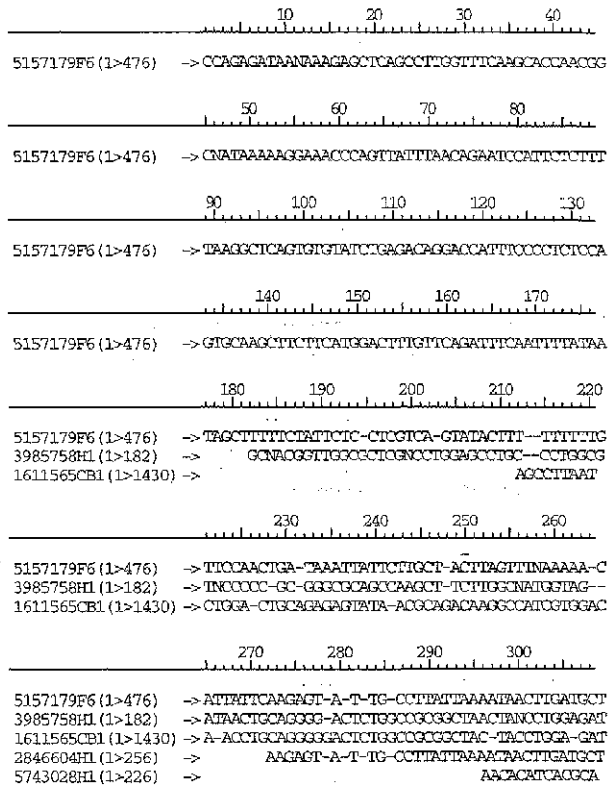


Figure 5A

WO 01/46398

26/43

PCT/US00/33743

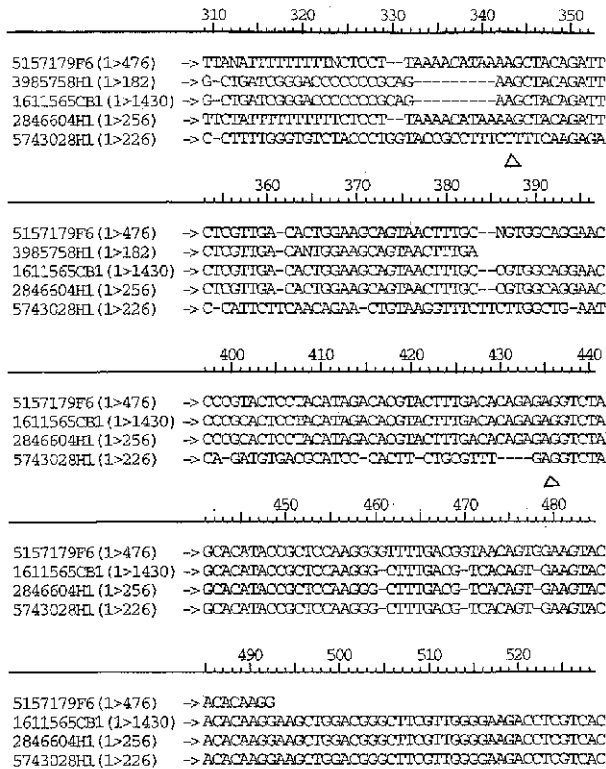


Figure 5B

WO 01/46398

27/43

PCT/US00/33743

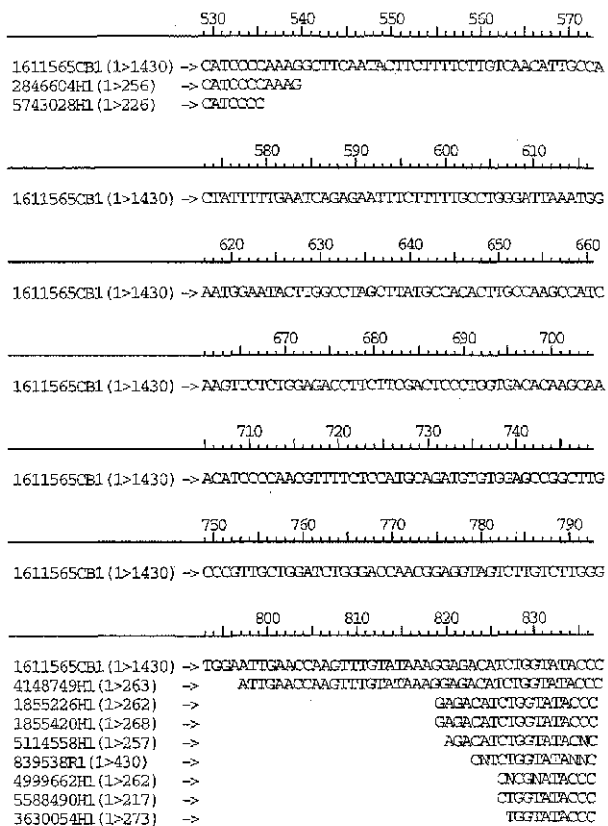


Figure 5C

WO 01/46398

28/43

PCT/US00/33743

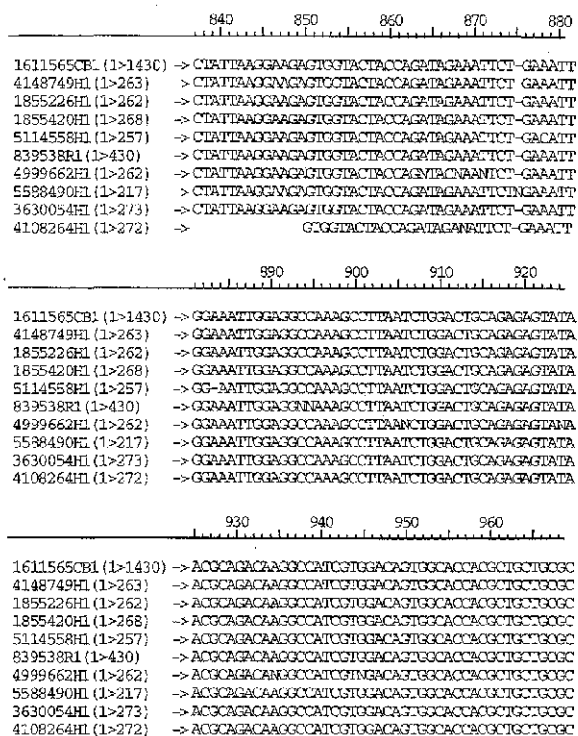


Figure 5D

WO 01/46398

29/43

PCT/US00/33743

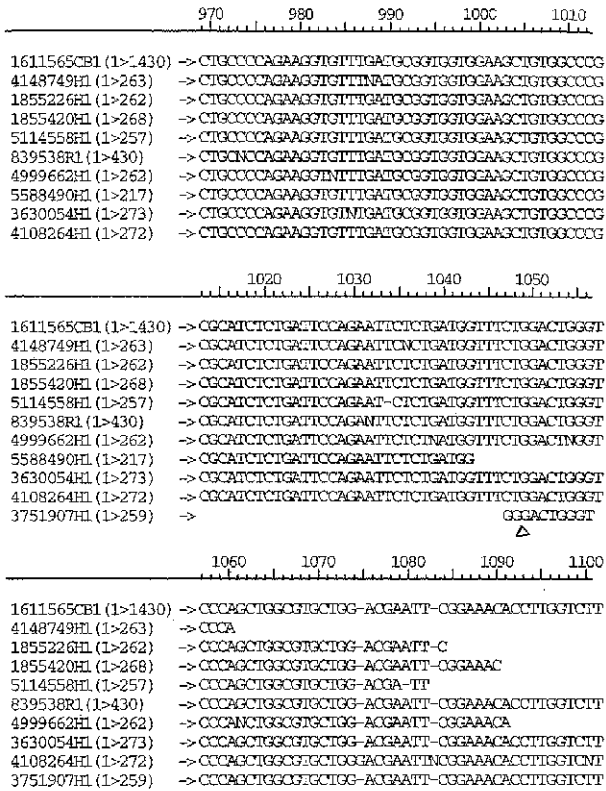


Figure 5E

WO 01/46398

30/43

PCT/US00/33743

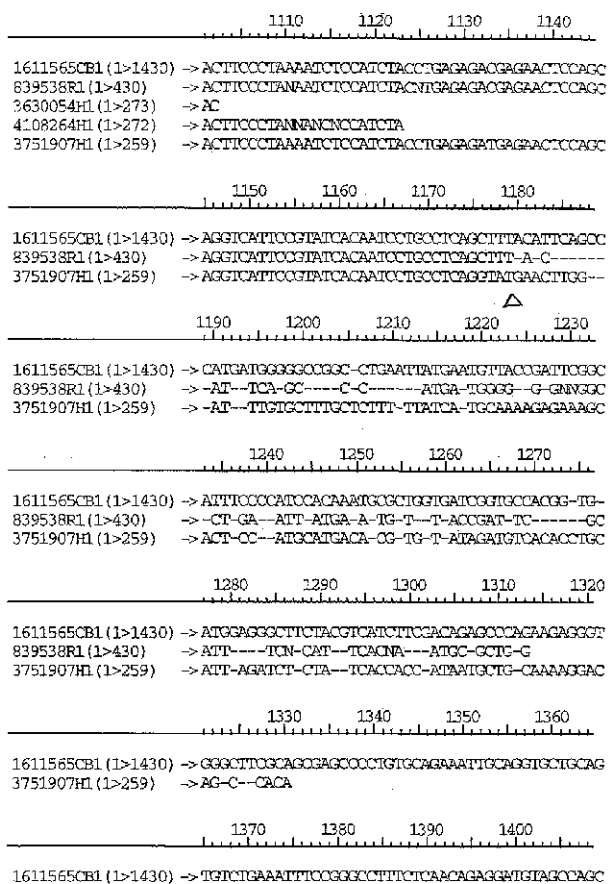


Figure 5F

WO 01/46398

31/43

PCT/US00/33743

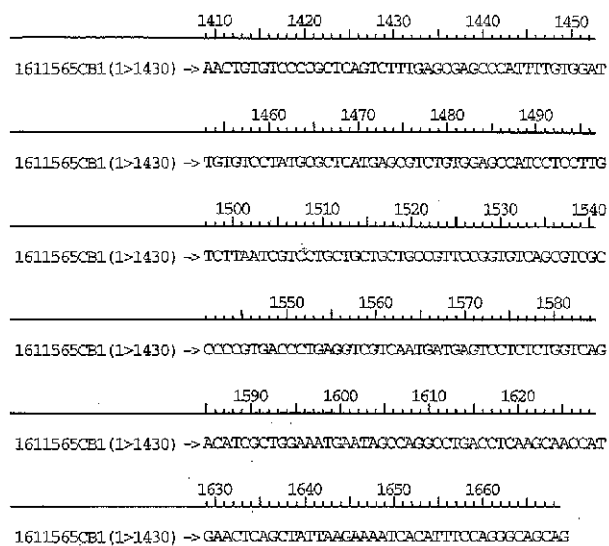


Figure 5G

WO 01/46398

32/43

PCT/US00/33743

```

1611565CB1      agccttaatctggactgcagagagataaacgcagacaagccatcgtggacaacctggcaggggactctg
701918575HL      .....
700921566HL      10 20 30 40 50 60 70
1611565CB1      gccggctactacctgggatgctgatcgggaccocccggcagcagctacagattctcgttgacactgg
701918575HL      .....
700921566HL      80 90 100 110 120 130 140
1611565CB1      aagcagtaactttgccgtggcaggaccocccgactcctacatagacagctacttgacacagagggctct
701918575HL      .....
700921566HL      150 160 170 180 190 200 210
1611565CB1      agcacataccgctcccagggtttgacgtccacagtgaagtacacacacaagaagctggcagggcttcggttg
701918575HL      .....
700921566HL      220 230 240 250 260 270 280
1611565CB1      gggaaacctcgtccaccatcccacaaggcttcaactctttcttctgtcaacattggccactatctttga
701918575HL      .....
700921566HL      290 300 310 320 330 340 350

```

Figure 6A

WO 01/46398

33/43

PCT/US00/33743

```

1611565CB1
701918575H1
700921566H1
atcagagaattctctttgacctgggatkaaatggaatggaatacttggccttagcttlatgccacacttggc
.....
360      370      380      390      400      410      420
.....

1611565CB1
701918575H1
700921566H1
aagccatcaagttctctggagacctctctgactctccctggtagacacaagcaaacatcccccaagcttctct
.....
430      440      450      460      470      480      490
.....

1611565CB1
701918575H1
700921566H1
ccatgcagatggtggagccggcttgcctggcttgcctggatctgggacccaagggaggtgcttcttggg
.....
500      510      520      530      540      550      560
.....

1611565CB1
701918575H1
700921566H1
tggaaatgaaccaagtttgrataaaggagacatctggtatccccctataaggaaaggtggtactaccag
.....
570      580      590      600      610      620      630
.....

1611565CB1
701918575H1
700921566H1
atagaaattctgaaattggaattggaggccaagccttaattctggactgcagagagataaacgcagaca
.....
640      650      660      670      680      690      700
.....

```

Figure 6B

WO 01/46398

34/43

PCT/US00/33743

```

1611565CB1
701918575H1
700921566H1
    aggccatcgtggagacagtgccaccacgctgctgcgctgccccagaagtgcttggatgcccgtggtggaagc
    .....
    710      720      730      740      750      760      770
1611565CB1
701918575H1
700921566H1
    tgtggcccgccatctctgatccagaattctctgattgggtttctggactgggtcccagctggcctgctg
    .....
    780      790      800      810      820      830      840
1611565CB1
701918575H1
700921566H1
    acgaaattcgggaacacacacctgcttactccctaaactctctatctacacctgagagacgagaactccagca
    .....
    850      860      870      880      890      900      910
1611565CB1
701918575H1
700921566H1
    ggtcattccggtatcaccaatcctgctcagctttacattcagcccatgatggggccgctgaattatga
    .....
    920      930      940      950      960      970      980
1611565CB1
701918575H1
700921566H1
    atgtaccgattcggcattccccatccacaatggcgtggctggtgatcgggtcccacgctaatggaggcttc
    .....
    990      1000     1010     1020     1030     1040     1050

```

Figure 6C

WO 01/46398

35/43

PCT/US00/33743

```

1611565CB1      tacgtcattctgcacagagcccagaagagggtgggtcttcgcagcggcccctgtgcgaaaattgcagggtg
701918575H1      taactgtctctttgacagagc.....
700921566E1      .....1060      1070      1080      1090      1100      1110      1120

1611565CB1      ctgcagtctgaaatttcggggcctttctcaacagagagatgagccagcaactgtgtcccgcctcagtc
701918575H1      .....
700921566E1      .....1130      1140      1150      1160      1170      1180      1190

1611565CB1      tttagcagcccattttgggattgtcctatgctcctcatgagcgtctgtggagccatcctcctgttc
701918575H1      .....
700921566E1      .....1200      1210      1220      1230      1240      1250      1260

1611565CB1      ttaatcgtcctgtctgtcgtccgttcctcgggtgtcagcgtcggcccctgacctgacctgaggtcgtcaatgatg
701918575H1      .....
700921566E1      .....1270      1280      1290      1300      1310      1320      1330

1611565CB1      agtcctctctgtgtcagacatogctggaatgaaatgaaatgacagccagcctgacctcaagcaaccatgaactcagct
701918575H1      .....
700921566E1      .....1340      1350      1360      1370      1380      1390      1400

```

Figure 6D

WO 01/46398

PCT/US00/33743

36/43

```

1611565CB1
701918575H1
700921566H1
    attagaaatcacattccaggccagcagccggatccgatgtggcgctttctcctgtgccaccgctc
    .....
    1410      1420      1430      1440      1450      1460      1470
1611565CB1
701918575H1
700921566H1
    ttcaatctctgttctgtccccagatgccttctcagattccactgtcttttgattcttgatttccaagcttcc
    .....
    1480      1490      1500      1510      1520      1530      1540
1611565CB1
701918575H1
700921566H1
    aaatccctacttccaaagaaataataaaaaaaacttcaattcctaaacccaaaaa.....
    .....
    1550      1560      1570      1580      1590      1600      1610

```

Figure 6E

WO 01/46398

37/43

PCT/US00/33743

```

1869869.comp      .....10 20 30 40 50 60 70
                  .....ggctagataggccct

1869869.comp      tccccatcacttttgggtccggctgccaccacgaaagagaccattgaaaaaaacccccatctctg
                  80 90 100 110 120 130 140

1869869.comp      agatgggatgctatcttggagaatgagatctaaactgagactgggtgaggggtgtgttctctgcttctct
                  150 160 170 180 190 200 210

1869869.comp      agctttgatccccactatccccctgtcttttctactcctctctctctctctctctctctctctctctctct
                  220 230 240 250 260 270 280

1869869.comp      ataaatagctcaatctagttttttttttttttttttttttttttttttttttttttttttttttttttt
                  290 300 310 320 330 340 350

1869869.comp      acagtcctgtctctaaagcaaatagcaactgtgcatctctctctctctctctctctctctctctctctct
                  360 370 380 390 400 410 420

1869869.comp      aagcttggctctctctctccagggtaccacacagggacacacacacacacacacacacacacacacacac
                  430 440 450 460 470 480 490

```

Figure 7A

WO 01/46398

38/43

PCT/US00/33743

1869869.comp cttctgaacttaagaagaataacttttgggttgagaatttaaaaggaaatggtggaacaaagggtcaggt 500 510 520 530 540 550 560

1869869.comp gctgaagtttcaagcagcagaatttcccgacttaaatttgaggtgacccaaggtatttccgcgcagcagag 570 580 590 600 610 620 630

1869869.comp tgcctcttttctcttcttcttcttctcaaggtacagtcacctggaacaccaccttgcacgaccttctctt 640 650 660 670 680 690 700

1869869.comp ctccatcagcagcagcagcatttgggtgggtgggaggtctctgagggtgctctggccacaggtgacctctgt 710 720 730 740 750 760 770

1869869.comp gttctctacttctgacccaagtgaaacacagcgggtgtgtgtccaggggaatctctatcttctgacctggg 780 790 800 810 820 830 840

1869869.comp cctctcacttccagcagcagcagatgtcaccagcaagtcacctgtctgtgctggtgccaagcagcagcagc 850 860 870 880 890 900 910

1869869.comp gcaactgacacaccatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 920 930 940 950 960 970 980
701911752H1

Figure 7B

WO 01/46398

39/43

PCT/US00/33743

1869869.comp
 701911752H1
 701901708H1
 990 1000 1010 1020 1030 1040 1050
 ggtcatgagggttgaactcatctgtgtggaatgttgaaccacagatctccatgtcccaagtgaacaaa
 acatggtgattagccagangctactacctcttgaacagacagatctgggaattagctgctgycagcaag
ctggcagcaaga

1869869.comp
 701911752H1
 701901708H1
 70049484.LH1
 70159615.0H1
 701031389H1
 1060 1070 1080 1090 1100 1110 1120
 gggccttcaccgctgcctcctgaactcatctgtccacatggcagcctgcacagcctgacaaagccaaattcggtt
 acatgancgactgancaccacaccctcctcaccatggcagcctgcacagcctgacaaagcctgacaaagccttggg
 calygaagactgaccacacaccctcctcaccatggcagcctgacaaagcctgacaaagccttggg
cttattctgtctctgcttt
gtataagaactaaacacltctgtacatcagcttcgacccctcaggt
agt
 1060 1070 1080 1090 1100 1110 1120

1869869.comp
 701911752H1
 701901708H1
 70049484.LH1
 70159615.0H1
 701031389H1
 1130 1140 1150 1160 1170 1180 1190
 ttggggcccgatcaagatacauglagaagccctccatgatacaagctcccaatacaagctgcccgtggatga
 gctcgatcaagagaccacataagaagcttccatgtagccctccatccatcaacgctgctggatgactgtg
 ctcgatcaagagaccacataagaagcttccatgtagccctccatcaacgctgctggatgactgtga
 ctctgctctttagtcaccatacactngcttctcccaagccctggatgttgaactttccgcaacccctt
 ctctgaaaaagtgtgtgcttcccttaacctgttttagatgagggatgtggcactgcaagcagggagcc
 ctctgaaaaagtgtgtgcttcccttaacctgttttagatgagggatgtggcactgcaagcagggagcc
 1130 1140 1150 1160 1170 1180 1190

1869869.comp
 701911752H1
 701901708H1
 ctgtgagatggcaaaccttgttaacagctgctgtgggacgtggccacacatctccatggccgggtattgc
 agacggcgaattgttaacagctgctgtgggacgtggccacacatctccatggccgggtattgcggggga
 gacggcgaacttgaacagctgctgtgggacgtggccacacatctccatggccgggtattgtctggggga

Figure 7C

WO 01/46398

42/43

PCT/US00/33743

```

1869869.comp      agactgacctgtgt*acagcgagtggttcgatacactccaatgatcatctccctccgacagagggccagcaact
700251567H1.comp agactgccaagt*atagggaaatggtcg.....
701432228T1.comp agactgccaagt*atagggaaatggttcgatacactccaatgatctaaagaataaagagacagacactatgt
700709032H1.comp agactgacctgtgttacagcgagtggttcgatacactccaatgatcatctccctccgacagagggcc.....
1620      1630      1640      1650      1660      1670      1680
          Δ
1869869.comp      tcagactggttgaggggagcccgaccccaaaagctccaggggagagaggttggaaagctgggtctgctgct
701432228T1.comp cctagcacagagagagaggttaccccaaaagctccaggggagagaggttggaaagctgggtctgctgct
1690      1700      1710      1720      1730      1740      1750
1869869.comp      ttaccagaggtcaagaaaggctccagggagctcgtcagggctttggcagaagaagacaaaggagtgatcc
701432228T1.comp tctctaatcttccatcagcccatgcatgtaagtttagtttctttaaagttcttcaagacagAACACMI
700542176H1      .....tcatgccccatgcatgtaagtttagtttctttaaagttcttcaagacagAACACACMI
700376670H1      .....tttagtTTCCTAAATGTTTCAGAGAAACACAI
1760      1770      1780      1790      1800      1810      1820
1869869.comp      tcatacggccctaaagcagagccagcttccctctctaaactggaaagcaaatTTgaaggcttgggggat
701432228T1.comp CAGATTATTGTATCAAAATCTATAATCCCTTCAGAGAGTtattta.....
700542176H1      CAGATTATTGTATCAAAATCTATAATCCCTTCAGAGAGTtatttaaaccttaagccactacaataaacg
700376670H1      CAGATTATTGTATCAAAATCTATAATCCCTTCAGAGAGTtatttaaaccttaagccactacaataaacg
1830      1840      1850      1860      1870      1880      1890

```

Figure 7F

WO 01/46398

43/43

PCT/US00/33743

```

1869869.comp      . . . . . gctccttgggaagaccatcttaagggaactgaggetctgcccgaattcttgcctcgaggccaaaattcccta
700542176H1      . . . . . tggaaatgacaaaattcaaaaaaaangg.....
700376870H1      . . . . . tggaaatgacaaaattctatgctgtcttcataatcgttttgtttgtttgttgcctgttggagaaccactcattactg
1900             1910   1920   1930   1940   1950   1960
1869869.comp      . . . . . tgtgaggcgtat.....
700376870H1      . . . . . tccctgcctaatcaataaatgggttgactg.....
1970             1980   1990   2000   2010   2020   2030

```

Figure 7G

WO 01/46398

PCT/US00/33743

```

<110> INCYTE GENOMICS, INC.
      KASER, Matthew R.
      COCKS, Benjamin G.
      LORINE, Jeanne F.
      TANG, Y. Tom
      YUE, Henry

<120> ASPARTYL PROTEASES

<130> PC-0008 PCT

<140> To Be Assigned
<141> Herewith

<150> 09/470,954
<151> 1999-12-22

<160> 49
<170> PERL Program

<210> 1
<211> 215
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No.: 700251567K1

<220>
<221> unsure
<222> 64, 94, 102, 106, 132, 143, 148, 186, 191-195, 205, 214
<223> a, t, c, g, or other

<400> 1
cgaccattcc ctatacaactg gcagttctctg gtacacaacc atccggcggg agtggfatta 60
tggngtgc atgttaactg tagaaatcaa tggcaagat cngcactgg actgcacaga 120
gtacaactat gcaagagca tcttggngt ggaccacaa ctctgtttgc ccaagaagt 180
atttggacc nnnnatcct tcagnagcct cctng 215

<210> 2
<211> 423
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No.: 701432228T1

<220>
<221> unsure
<222> 346, 350, 400, 405
<223> a, t, c, g, or other

<400> 2
taaatnactt cctgtcnagg attatagatt ttgatadcaa taactctgat tgtttctctt 60
gaacacattt agaaactaa acattaccat gcaaggggca tgatggaga tttagagaga 120
tcttgacacc ccaagaaatg gctgatctta tcaattgggt ttgggtacc tctctctctg 180
tgctaggaca tagtctctgt ctctttttct tttagatcat ttgggtatc gaccattccc 240
tatcaactgg cagttctctg tacaaaccaa tccggcggga gtgggtatc gaagtatca 300
ttgtactgtt agaaatcaat ggtcaagatc tgaatgggg actgcaagg aggtaatnaa 360
ctggaggggg acccagtga agtaagagg taaatcccau tgtngagaa ccatgagggt 420

```

WO 01/46398 PCT/US00/33743
 423

lgy

<210> 3
 <211> 275
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No.: 700769871H1

<400> 3
 gttctgtgat ggcctctgcag gccctgccct tcaacttagg gagtgetgtg tgattgcaa 60
 cttctgcct tggcaacagcc ctaacctcag ccaggtcct gtgcctcac cctctctcc 120
 accctcagg ttctctcaca ctgggattaa cctcttact tcactgggtc cctctccagt 180
 ttattacctc ctgcagtcct attttcagat ctgaccalt gattttcaca cgtacaatga 240
 tcaactcata ataccactcc cgcctgatgg gtgtg 275

<210> 4
 <211> 271
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No.: 700769257H1

<220>
 <221> unsure
 <222> 201, 209, 251
 <223> a, t, c, g, or other

<400> 4
 gttctgtgat ggcctctgcag gccctgccct tcaacttagg gagtgetgtg tgattgcaa 60
 cttctgcct tggcaacagcc ctaacctcag ccaggtcct gtgcctcac cctctctcc 120
 accctcagg ttctctcaca ctgggattaa cctcttact tcactgggtc cctctccagt 180
 ttattacctc ctgcagtcct nttttcagnt ctgaccalt gattttcaca cgtacaatga 240
 tcaactcata ataccactcc cgcctgatgg g 271

<210> 5
 <211> 276
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No.: 701188357H1

<220>
 <221> unsure
 <222> 270
 <223> a, t, c, g, or other

<400> 5
 gaactcttct ggttttcat tactgtact aagctctatg ctcaagttct gtgatggctc 60
 tgcaggccct gcccttcacc ttaggagatg ctgtgtgatt gcccaagtttc tgccttggca 120
 cagccctaac ctacgccag gttctgtgct tccacctct cctccacct catgtttctc 180
 tcacactggg attaacctct tacttctact ggtctccctt ccagtttatt acctctttgc 240
 agtccatltt cagatcttga ccattgattn ctacac 276

WO 01/46398 PCT/US00/33743

<210> 6
 <211> 252
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No.: 701918575H1

<400> 6
 gggggcccag ctggcatgct ggacaaattc tgaactora tgggcgtatt tccotaaagat 60
 tctatctac ctgagagatg agaacgcagc tggctcttcc cgaatcaca ttctgcccga 120
 gctctacatt cagcccataa tgggagctgg ttccaattat gagggtaac gtttltggtat 180
 ctctcttccc acaaalgcgc tgggtgattg tgcaccctg atcggagggat tctactgggt 240
 ctttgacaga gc 252

<210> 7
 <211> 140
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No.: 700921566H1

<400> 7
 cgcacgtgct tctctccgaa tcaaccattct gccgcagctc tacattcagc ccatgatggg 60
 agctggtttc aattatgagt gctaccgttt tgggtatccc tcttcacaa atgctctggt 120
 gatggtgccc cctgctatga 140

<210> 8
 <211> 279
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No.: 700494841H1

<220>
 <221> unsure
 <222> 44
 <223> a, t, c, g, or other

<400> 8
 ettattctgt cctgccttcc totgcctttt agtcaactac actntgctct tctcaggcc 60
 ctgagatggt gacttltccg caacccttlt atcagcacac attcaattgt ctttcagtt 120
 gtatgaccca tgtctcttta catactttgt ctgcttcttg gcactctgct gcctcagct 180
 ggggacatac ctgaggaagg atggtgatgc ggaaggactg attggtgact tcaaccatga 240
 ggtaaagtga aatgactggg aanaatgttcc aaggggtcg 279

<210> 9
 <211> 281
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No.: 701901708H1

WO 01/46398 PCT/US00/33743

<400> 9
ctggcagcaa gaaatgagca gactgacccc acacccctct caccatggca agcgetgaca 60
gcaaaagcaa ttctgttttg ggtctgata aagaccacat agaagcttc catgatgacc 120
ctcaccataa cggatgctgt ggaatgactgt gagacgcua acttgtaaca gtctgtttgg 180
gacgtggcca catcttcacc tggcagtagt attgtgggg gaaacaalca caggatgaga 240
gcatgtctta ttctgtctg cctttctctg ccttttagtc a 281

<210> 10
<211> 303
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No.: 701911752H1

<220>
<221> unsure
<222> 43, 100, 109, 189, 291
<223> a, t, c, g, or other

<400> 10
caagccttag ccttaccctt cacacatggt ggattagcca gangtaacta cctttgaa 60
agacagctct gggcattage tctgtggcag aagacatgah cagactganc caacaccc 120
ctcaccatgg caagcctgta cagcaaaccc katctgclit cggctgcat caaagacc 180
alagagantl tccatgata cggctcccat aacgtgcct gtggatgact gtgagacc 240
gaattgtaac agtcgttctt ggcctggccc acatcttca ctggccgtgg nattgctggg 300
gga 303

<210> 11
<211> 284
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No.: 701823687H1

<220>
<221> unsure
<222> 11, 153
<223> a, t, c, g, or other

<400> 11
agccgccaga ngtagcaact agagtccaga agggcaactct tctgtgttt cattaactgta 60
cltaagctct atgtcagggt tctgtgatgg ctctgcagcc cctgcccctt accttaagga 120
gtgctgtgtg attgccaagt tctgtccttg gcnagcccta acctcagccc aggtcctgtg 180
ctccaccct ctctccacc ctcalygttc tctcaccctg gattaacct cttaacttca 240
ctgggtcccc ttccagttta ttacctctt geagtcactt tcag 284

<210> 12
<211> 309
<212> DNA
<213> Rattus norvegicus

<220>
<221> misc_feature
<223> Incyte ID No.: 701596150H1

WO 01/46398

PCT/US00/33743

```

<400> 12
glatwagaac taacacttcg tacatcagct tcgacctta cagtctcga aaagtgtgt 60
gbatccttac cctgtcttgg tagatgagga gtctgggcac tgcagcagg agcccaaggy 120
caactagtaa cagcaggact cagtcaggtc ctgcgagaag ccttgcocgg agccgcaga 180
ggtagcaact agaglcocaga agggactctc tctygtttt cattactgla cttagctct 240
atgctcaggt tctgtgatgg ctctgcaggc cctgcccttc accllagga gtgtgtgtgt 300
attgcaag 309

```

```

<210> 13
<211> 158
<212> DNA
<213> Rattus norvegicus

```

```

<220>
<221> misc_feature
<223> Incyte ID No.: 700542176H1

```

```

<220>
<221> unsure
<222> 156
<223> a, t, c, g, or other

```

```

<400> 13
tcattgccca tgcattgtaa tglitlagtt cttaaatgtg ttcagaagaa acacatcaga 60
ttatttgtat caaaatctat aatccttgac aggaagttat ttaaacctaa gccagttaca 120
taaacgtgga atgaacaaat tcaaaaaaa aaaaahgy 158

```

```

<210> 14
<211> 248
<212> DNA
<213> Rattus norvegicus

```

```

<220>
<221> misc_feature
<223> Incyte ID No.: 70031389H1

```

```

<400> 14
agtcctgaa aaagtgtgtg tacccttacc ctgtcttgt agatgaggag totggcacty 60
cagcaggag cccaaaggca cttagttaaca gcaggactca gtcaggctct cgcagaagcc 120
ttgccggag ccgcccagag tagcaactag agtccagaag ggaactcttc tgygtttca 180
ttactgtact taagctctat gctcaggttc tgygatggt ctgcaggccc tgcocctca 240
cttagggag 249

```

```

<210> 15
<211> 206
<212> DNA
<213> Rattus norvegicus

```

```

<220>
<221> misc_feature
<223> Incyte ID No.: 700376870H1

```

```

<400> 15
tttagttct taaatgtgtt cagaagaaa acatcagatt atttgtatca aaatctataa 60
tccttgaag gaagttattt aaacctaaag cagtacaata aacgtggaat gaacaaatc 120
tatgtctct tcatactggt ttgtotttgc ctgttgaga accactcatt actgtctctt 180
cctaactaat aaatgggttt gaactg 206

```

WO 01/46398 PCT/US00/33743

<210> 16
 <211> 91
 <212> DNA
 <213> Macaca fascicularis

<220>
 <221> misc_feature
 <223> Incyte ID No.: 700709032H1

<400> 16
 ggctctctgc gsggggagca tgatcatttg aggtatogac cactcgtctg aacacaggca 60
 gtctctggta tacaccatt ccggggggga g 91

<210> 17
 <211> 252
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 5577513H1

<220>
 <221> unsure
 <222> 100, 161
 <223> a, t, c, g, or other

<400> 17
 ctggaatcaga cwaattcttc atcaacggct ccaactggga aggcatactg gggctggcct 60
 atgctagat tgccaggott tctggctctg gcttccccc caaccagctc gaagtctgg 120
 ctctctggg agggagcatg atcattggag gtatcgacca ctctctgtac acaggcagtc 180
 nctggatcac acccaaccgg cgggagtggt attataggt gatcattctg cgggtggaga 240
 tcaatggaca gg 252

<210> 18
 <211> 205
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 5094633H1

<220>
 <221> unsure
 <222> 28, 44
 <223> a, t, c, g, or other

<400> 18
 agagcattgt ggcagctggc accacatngc ccaagaaagt gttngaagct gcagtcakat 60
 ccatcaagtc agctctctcc acggagaagt tccctgatgg gttctggota ggagagcaga 120
 tgggtgctg gcaagcaggc accaccctt ggaacatctt cccagtcac tcactctacc 180
 taatgggtga ggttaccaac cagtc 205

<210> 19
 <211> 249
 <212> DNA
 <213> Homo sapiens

WO 01/46398

PCT/US00/33743

```

<220>
<221> misc_feature
<223> Incyte ID No.: 2293393H1

<220>
<221> unsure
<222> 2, 19, 36, 174, 177, 217, 221, 229
<223> a, t, c, g, or other

<400> 19
cmtgaccta ggyaaaaana gaggcaggtta cccgtntcct ggcacagaag gagagtcagt 60
cccccaaga ccaagcaata agatcagtga tttcttgggg tgycaaggtc ttctacaggc 120
taccotttct atotttctgc ttotaascaa atcacaacca aagtgttttc tagntttkta 180
aatgtgttca ggagaaaaga ctttccggga tttttambta ntgtttteng antcacaag 240
cccttgata 249

<210> 20
<211> 236
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2908594H1

<220>
<221> unsure
<222> 26, 44, 62, 91, 154, 169, 177, 204, 214
<223> a, t, c, g, or other

<400> 20
gtacaacgag tggctgatcc ctccantgat ctgggaaaaa aanagggcag gtaccctgt 60
chtggcaag aaggagagtg agtcccccaa ngacccaaga ataagatcag tgattttctg 120
ggytggcaag gtcttctaca ggtaccctt ttctcttccc tctttctana ccaatctac 180
ccaaagtgat ttctagtctc ttanattgtt tcangagaaa agactttccg ggattt 236

<210> 21
<211> 165
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2007269H1

<400> 21
gatacctcca atgatctagg gaaaaaaga ggcaggtacc cgtgtcctgg cacagaagga 60
gagtgagtec cccaaggacc aagcaataag atcagtgatt tcttggggtg gcaaggtctt 120
ctacaggcta cccctttcat ctctctgttt ctaacaaat cctac 165

<210> 22
<211> 382
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1438613F1

```

WO 01/46398

PCT/US00/33743

<220>
<221> unsure
<222> 373, 374
<223> a, t, c, g, or other

<400> 22
gaccceclila ctgctggaca ctactcaatct gttttgtctt coaagtctcg gcaagccata 60
actgctcaact gctctccagc gtytactttt aagagagacc cccctgactc aggetgggac 120
atccctgggg aagatcgtg alucggaaag actgtccact caaggaggag gctgctctga 180
tggattttag tgcagcttca aactcttctc tgggaaacg aagtcactt tcaagacctg 240
tcacattgac tcaccggca caatgatac ctcataaac cactccccc ggaagggtg 300
ataccagaga ctgctgtgtl acagcagtg gtcgatacc ccaatgatc agggaaaaa 360
aagggcagc tannctgtc ct 382

<210> 23
<211> 336
<212> DNA
<213> Homo sapiens

<220> misc_feature
<223> Incyte ID No.: 348461951

<400> 23
tgcacgggtt acctgcctct ttttttccct agatcattg aggtatcgac cactcgtgt 60
acacagccag tctctggtat acaccatcc gccgggactg gtattatgag gtatcattg 120
tgcgggtgga gctcaatgga caggatctga aatggactg caaggaatc aactatgaca 180
agagcattgt ggcacgtctc accaccacc ttctttgct caagaagtg ttgaaagctg 240
cagtcacatc catcaaggca gctctctcca cggagaagt cctgatggt ttctggctag 300
gagagcagct ggtgtgctg caagcaggca ccacc 336

<210> 24
<211> 379
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3757870H1

<400> 24
gaccattaa cccctgactg ttctaggctc aacttcaac cctttctgca ggtcctatta 60
cctctgctc atcctccaa catgataacc agagcttcc ttcacattg actgctacc 120
cccthatgtt cccaggctct cctctggttt tattaacctc ttgcagcca ttctcagatc 180
ctgtccattg atctccacc gccaatgat cactcctaa taccactcc gccggatggg 240
tgtataccag agactgcctg tgtacagca gtggtgat 279

<210> 25
<211> 305
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2900063H1

<400> 25
ctggtataca cccatccggc gggagtgcta ttatgagtg atcattgac gggctggagat 60
caatggcag gctctgaaaa tggactgcaa gggctaccg tatgcaagg gactgtgga 120
cagtgccacc accacccttc gttgcccac gaagtgttt gaagctgag tcaaatccat 180
caagggccc tctccaagg agaagttccc tgatggttcc tggctaggag agcagctgt 240

WO 01/46398 PCT/US00/33743

gtgctggcaa gcaggcacca ccccttggaa cattttccca gtcctctcac tctacctaat 300
gggta 305

<210> 26
<211> 302
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2123411H1

<400> 26
ccagagtctt cclleacatt gtaactgcta cccccttatg ttcccaggct ctccccttgg 60
tttattacct ccttgccagtc cattttcaga tccctgtccat tgatctccac ccgcacacatg 120
atgacctcat aataccaact ccgcgggatg ggtgtatacc agagactgcc tgtgtcacgc 180
gagtggtcga taccctcaat gatctaggga aaaaaagagg caggtaccgc tgtccctggca 240
cagaaggaga gtgagtccc caaggaccaa gcaataagat cagtgatttc ttggggtygc 300
aa 302

<210> 27
<211> 263
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1436719H1

<400> 27
gacccttca ctgctggaca ctactcaatc gttttgtctt ccaagttctg gcaagccata 60
cctgtcactc gtctcccagt gtgtactttt aagagagatc cccctgactc aggtctggac 120
atacctggcg aaggttggtg atgcggaagg actgtccact cacggaggag gctgcttga 180
tggatttgac tgcagcttca aacactttct tgggcaaacg aagtcatttt tcagatcctg 240
tccattgata tccaccgca caa 263

<210> 28
<211> 255
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 1438613H1

<400> 28
gacccttca ctgctggaca ctactcaatc gttttgtctt ccaagttctg gcaagccata 60
cctgtcactc gtctcccagt gtgtactttt aagagagatc cccctgactc aggtctggac 120
atacctggcg aaggttggtg atgcggaagg actgtccact cacggaggag gctgcttga 180
tggatttgac tgcagcttca aacactttct tgggcaaacg aagtcatttt tcagatcctg 240
tccattgata tccac 255

<210> 29
<211> 476
<212> DNA
<213> Homo sapiens

WO 01/46398

PCT/US00/33743

```

<220>
<221> misc_feature
<223> Incyte ID No.: 51571797F6

<220>
<221> unsure
<222> 11, 46, 251, 301, 312, 369
<223> a, t, c, g, or other

<400> 29
ccagagAtaa naaagagctc agccttggll tcaagcaca acggcAtaa aaaggaacc 60
cagttattta acagaatcca ttctctttta aggttcagly tgcAtctga acagaccat 120
ttccctctc cagtgcagc ttcttctctg actttgttca gatttcaatt ttataatagc 180
ttttctatt ctctctgcca gtatactttt tttttgttcc aactgataaa ttattcttgc 240
tacttagttt naaaaacatt attcaagagt attgccttat taaaaaact tgaigtctta 300
natttttttt tncctcttaa aacataaaag ctacagattc tcyttgacc tggagcagc 360
aactttgcmg tggcaggaac cccgtactcc taatagaca cgtactttga cacagagag 420
tctagacat accgctccaa ggggttttga cgytaacagt ggaagtacac acaagg 476

<210> 30
<211> 182
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 3985758H1

<220>
<221> unsure
<222> 3, 18, 39, 69, 110, 164
<223> a, t, c, g, or other

<400> 30
gcnacggltg ggcctcgncc tggagcctgc cctggcgtac cccgcggggc gcagccaagc 60
ttcttggona tggtagataa ctgcagggga ctctggcgcg ggttaactan cctggagatg 120
ctgatgggga cccccccgca gaagctacag attctcgttg acantggaag cagtaacttt 180
ga 262

<210> 31
<211> 256
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No.: 2846604H1

<400> 31
aagagtaltg ccttattaaa ataacttgat gclttctatt tttttttct ccttaaaaca 60
taaaagctac agattctcgt tgacaactgga agcagttaact ttgccctggc aggaaccccc 120
cactcctaca tagacaagta ctltgacaca gsgagctcta gncataccg ctcccagggc 180
tttagctca cagtgaagta cacacaagga agctggacgg gctctgttgg ggaagacctc 240
gtccaccatc caaag 256

<210> 32
<211> 226
<212> DNA
<213> Homo sapiens

```

WO 01/46398 PCT/US00/33743

<220>
 <221> misc_feature
 <223> Incyte ID No.: 5743028H1

<400> 32
 aacacacacac gcaaccttttg ggtagctctacc ctggtaccgc ctttcttttc aagagaccat 60
 tcttcaacag aactgtaagg ttctctctty gctgaalccg atgtgacgca tcccactctt 120
 gcgtttgagg tctagccat acgctccaa gggctttgac gtaacagtga agtacacaca 180
 aggaagctgg accgctcttg ttggggaapa cctcgtcacc atcccc 226

<210> 33
 <211> 259
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 3751907H1

<400> 33
 gggactgggt ccaagctggc gtgctggacg aattcgaaa . ctgggc ttaactccct 60
 aaaaatccca tctacctgag agatgagaa ccaagcaggt cactcctgat cacaatccct 120
 cctcaggtat gaactggat ttgctcttg ctctttttat catgcaaaag agaaagcact 180
 caatgatga caactgtata gatgtcacac ctgcattaga tctctatcac caccataatg 240
 ctgcaaaagg acagccaca 259

<210> 34
 <211> 263
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 4148749H1

<220>
 <221> unsure
 <222> 190, 239
 <223> a, t, c, g, or other

<400> 34
 attgaaccaa gttttgataa agyagacatc tggatatacc ctattaagga agagtggtag 60
 taccagatag aaattctgaa attggaatt gtaggcacaa gcettaatct ggactgcaga 120
 gagtataacg cagacaagcc catctgtggc agtggcaca cgtgtctgcy cctgcccag 180
 aaggtgtttn atgctgtgtt ggaagctgtg gccctgcgat ctctgattcc agaattcnct 240
 gatggtttct ggaactgggtc cca 263

<210> 35
 <211> 262
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 1855226H1

<400> 35
 gagacatctg gtataccctt abtaagggaag agtggtaata ccagatagaa attctgaaat 60
 tggaaattgg aggccaaaag ctlaactctg actgagaga gtatacgcg gacaaggcca 120
 tctggacag tggcaccag ctgctgcgcc tgcccacag ggtgtttgat gctgtgtgtg 180

WO 01/46398 PCT/US00/33743

aagctgtggc ccgcgcctct ctgattccag aattctctga tgyttcttgg actggytccc 240
 agctggcctg ctggacgaat tc 262

<210> 36
 <211> 268
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 1855420H1

<400> 36
 gagacatctg gtataccctt attaaggaag agtgggtacia ccagatagaa attctgcaat 60
 tggaaattgg agyccaaaagc cttaatcttgg actgcagaga gtataacgca gacaaggcca 120
 tctgtgacag tggcaccacg ctgctgcgcc tgcctcagaa ggtgtttgat ggggtggtgg 180
 aagctgtggc ccgcgcctct ctgattccag aattctctga tgyttcttgg actggytccc 240
 agctggcctg ctggacgaat toggaaac 268

<210> 37
 <211> 357
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 5114558H1

<220>
 <221> unsure
 <222> 16
 <223> a, t, c, g, or other

<400> 37
 agcactctgg tatacnccta ttaagyaaga gtggtactac cagatagaaa ttctgcaatt 60
 ggaattggag gccaagcctt taatctggac tgcagagagt ataacgaga caagycattc 120
 gggacactg gcaaccyct gctgcgctg cccagaaag tgtttgatgc ggtggtgaa 180
 gctgtgccc ggcactctct gattccagaa tctctgctg tttctgact ggtccaccg 240
 tggcgtgctg gaagatt 257

<210> 38
 <211> 430
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 839538R1

<220>
 <221> unsure
 <222> 2, 12, 13, 70, 71, 150, 208, 285, 299, 376, 377, 412, 420
 <223> a, t, c, g, or other

<400> 38
 cttctggtat amccctatta aggaagctg gtactaccag atagaattc tgaattgga 60
 aatggaggn aaagcctta atctggactg cagagaqlat aacgcagca aggccatct 120
 ggacactggc accacgtgc tgcgctgcn ccagaaggtg ttctatggg tggcgaagc 180
 tgtgcccgc gcactctctg ttccagantt ctctgatgt ttctggactg ggtccagct 240
 ggcgtgctgg acgaattgg aaacaccttg gtcttacttc cctanaatct caatctact 300

WO 01/46398

PCT/US00/33743

gagagacgag aactccagca ggtcattccg tatcacaalc ctgcctcagc ttacattca 360
 gcccatgatg ggggrrnggc ctgaattatg aatgtaaccg attcgcattt cncattcaen 420
 aatgcgctgg 430

<210> 39

<211> 317

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 5588630H1

<220>

<221> unsure

<222> 49

<223> a, t, c, g, or other

<400> 39

ctgggtatacc cctattaagg aagagtggta ctaccagata gaaattctng aaattggaaa 60
 ttggaggcca aagccttaat ctggactgca gagaglataa agacaag gccatcgtgg 120
 acagtggcac cacgtcgtcg cgcctgcccc agaaggtgtt tgatgcggtg gtggaaactg 180
 tggcccgccc atctctgatt ccagaattct ctgatgg 217

<210> 40

<211> 262

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 4999662H1

<220>

<221> unsure

<222> 2, 5, 38, 42, 45, 79, 97, 108, 118, 157, 212, 227, 235

<223> a, t, c, g, or other

<400> 40

cncgnatacc cctattaagg aagagtggta ctaccagnata cnaantctga aattggaaat 60
 tggaggccaa agccttaanc tggactgcag agagtanac gcagacang ccbctctnga 120
 cagcgcacc acgtcgtcgc gcctgcccc gaaggtnttt gatcggtgg tggaaactgt 180
 ggcccgcgca tctctgattc cagaattctc tnatggttite tggactnggt cccnctggc 240
 gtgctggacy aattcggaaa ca 262

<210> 41

<211> 273

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 3630054H1

<220>

<221> unsure

<222> 158

<223> a, t, c, g, or other

<400> 41

WO 01/46398

PCT/US00/33743

```

tgggtataccc ctattaagga agagtgggtac taccagatag aaactctgaa attggzaatt 60
gggggccaaa gcttbaatct ggaactgcaga gegtataacg cagacaaggc catcgtggac 120
agtggcacca cgtgtctgcy cctgccccag aaggtgtntg atgctgggtt ggaagctgtg 180
gccccgcgat ctctgatccc agaattctct gatggtttct ggactgggtc coagctggcg 240
tcttggagaga attcggaaac accttgggtct tac

```

```

<210> 42
<211> 272
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No.: 4108264H1

```

```

<220>
<221> unsure
<222> 19, 232, 249, 250, 261, 263, 265
<223> a, t, d, g, or other

```

```

<400> 42
tgggtactac cagatagana ttctgaatt ggaattgga ggcacaagcc ttaacttga 60
ctgaagagag tataaagcag acaaggcaat cgtggacagt ggaacacgc tgcctgcct 120
gccccagaag gttttgagc cgtgtgtgga agctgtggc cggcctctc tgaatccaga 180
attctctgal gttttctgga ctgggtccca gctgncctg tgggacgaat tncggaaaca 240
ccttggctant acttcccctan nancnccatc ta

```

```

<210> 43
<211> 1601
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No.: 1611565C31

```

```

<400> 43
agccttaabc tggactgcag agagtataac gacagacaagg ccactcgtga caactcgcag 60
ggggactctg gccggcgtca ctacctggag atgctgabcg ggaccccccc cagaaagcta 120
cagattctcg ttgacaactg agcagtaaac ttggccgtgg aaggaacccc gcaactctac 180
atagacaagc actttgacac agagaggtct agcaaatacc gctcacaagg ctttgacgtc 240
acagtgaaat acacacaagg aagctggacg ggcttcgttg ggaagacct cgtcaccatc 300
ccccaaagct tcaataactc tttttctgtc aacattgcca ctatttttga atcagagaat 360
ttotttttgc ctgggathaa atggaaatgga atacttggcc tagcttatgc cacacttggc 420
aagccatcaa gttctctgga gacctctctc gactccccgg tgacacaagc aaacatcccc 480
aacgttttct coatgcagat gtgtggagcc ggccttgcocg ttgctggatc tgggacaaac 540
ggaggttagtc ttgttctggg tggaaatgaa ccaagtittg ataaaggaga caatctggtat 600
acccttatta aggaagagtg gtaactaccg atagaaatc tgaattgga aattggaggc 660
caaaagccta atctggactg nagagagtat aacgcagaca aggcacatct ggacagtggc 720
accacgtgc tggcctgccc ccagaaggtg tttyatgcgg tggtggaagc tgtggcccgc 780
gcactctcga tccagaatc ctctgaggtt ttctggactg ggtcccagct ggcgtgctgg 840
acgactctgg aaagactctg gttctacttc cctaaatct caacttaact gagagcagc 900
aacctcaaga ggtactcccg takcaaatc ctgctcagc tttaactca gccaatgatg 960
ggggccggcc tgaattalga atgtaccga ttggcaattt ccccatccac aaatgcctg 1020
gtgatgggtg ccscgtgat ggaaggttcc acgtcaatc tgcacagac ccagaagagc 1080
gtgggtctcg cagcagccc ctgtgcagaa atlgcaagtn ctgcagtgc tgaatcttc 1140
ggccatttct caacagaaga tgtagccagc aactgtgtcc ccgctcagtc tttagagagc 1200
cccattttgt agattgtgta ctatgctctc atgaggtct gtggagccat cctctcttgc 1260
ttaactgtcc tctctgtgct gcccttccgg tctcagcctc gcccccgtga ccttgaggtc 1320
gtcaatgatg agtctctctt ggtcagacat cgtcggaaat gaatagccag gccctgacct 1380
aagcaacctt gaactcagct attaagaaaa tcaatttcc agggcagcag ccgggatgca 1440

```

WO 01/46398 PCT/US00/33743

tgggtggcgt tctctctgtg ccccccctg tcaactctct gttctgtctc cagatgcctt 1500
 ctatgattcac tgtcttttga tctctgattt tcaagcttcc aatctctccc tacttccaaq 1560
 aaaaaataat aaaaaaaaaa cttcattcta aacccaaaaa a 1601

<210> 44
 <211> 1911
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No.: 1869869

<400> 44
 atacgcctca catagggaat ttggccctcg aggcagaat tgggcagagc tcaagtccct 60
 aagatggtct tccccaggga gcctccccca agccttcaaa ttggcctcc agtlttagaca 120
 ggaagctggc ctctgccttt agggcctgat gaggactcac tccctgtctc tttctgcaaa 180
 agcctgaaga ctctcctggg cctttctttg actctctgtt aaagcagacc caagtccca 240
 acctctctc ctgcagctt tctggtgctg gcttccccct caaccagctc gaagtgcctg 300
 cctctgtcgg agggagcag atcattggag gtatcgacca ctctctgtac acaggcagtc 360
 tctggtatcc acccatccgg cgggwtgggt attatgaggt c--catt-gtg cgggtgaga 420
 tcaatggaca ggaactgaaa atggactgca aggagtacaa ctatgacaag agcattgtgg 480
 acagctggcac caaccaactt cgtttgccc aaaaagtgtt tgaagctgca gtcaatcca 540
 tcaaggcagc ctctctcagc gagaagtcc ctgatggtt ctggttagga gageagctgg 600
 tctgtcggca agcaggcacc accccttgga acalltctcc agtcatctca ctctacotaa 660
 tgggtggcgl taaccaaccg tctctccgca tcaccatcct tccgcagcaa tacctggcgc 720
 cagtggaaga tctggccacg tccccagcag actgttaca gtttgcctc tcaacatcat 780
 caaggggcac tgttatggga gctgttatea tggaggcctt ctacgttctc ttgatcggg 840
 cccgnaaacg aattggtctt gctgtcagcg ctgtccatgt gcacgttag ttcaaggcgg 900
 cagcgtgaga aggcctcttt gtcacctgg acatggaaga ctgtgctac aacatccac 960
 agcagatga gtcnaacctc atgacatag cctatgtcat ggtgctcctc tgcgccctct 1020
 tcatgtctcc actctgcctc atggtgtgtc agtggcctg cctcctctgc ctggccagc 1080
 acagatgca ctttctgat gacatctccc tctgaaagt aggagccca tgggcagaag 1140
 acagagcttc cctgggaca caectcctg gttcactttg gtcacaagta ggagacacag 1200
 atggcaactg tggccagagc acctkagagc cctccccacc caccnaatgc ctctgccttg 1260
 atggagaagg aaangctgg caaggtgggt tccagggact gtacctgtag gaaacagaaa 1320
 agagagcaaa gaagcactc cctggcagga atctcttgg tcaactcaa tttagctgg 1380
 gaattctgc tggctgaaac ttgagcctg aactttgtc cccactcct ttaaalctc 1440
 caaccocaaq tattectctt ttcttagtt cagaagtact ggcctcacac gaaggttacc 1500
 tggcgtatg tccctgtggt acctgcccag aagaagacc aagcttgttt cctctgctgg 1560
 caaagtcagt agggagggat gcaagtttg ctatttctt tagagacagc gactglataa 1620
 acaagctcaa cttgtgtgca agatgtcct cttgaattaa aaaaaaaa tagattgaact 1680
 atttatacaa atggggcggg ctggaagag gagaaggaga gggagtcaa agacagggaa 1740
 tagtgggatc aaagctagga aagcagaaa caaacacct caccagctct agtlttagac 1800
 ctcatctcca agatagcctc ccatctcaga agatgggggt tgttttcaa tggttctctt 1860
 tccgtggtgg cagccggacc aaaaatgaga tggggaagg cctatctaga c 1911

<210> 45
 <211> 1366
 <212> DNA
 <213> Homo sapiens

<300>
 <309> g1658285

<400> 45
 cccagaactc agagctgtct tctctctgtg gccagttggg gaaccagctc atgaagtgga 60
 tgggtggtgt cttggtctgc ctccagctct tggaggcagc agtggteaaa gtgcccctga 120
 agaaatttaa gtctatccgt gagaccatga aggagaaggg cttgctgggg gagtctctga 180
 ggaccocaaa gtatgatcct gcttggagtt accgctttgg tgacctcagc gtgacctagc 240

WO 01/46398

PCT/US00/33743

```

agcccatggc ctacatggat gctgcclact ttggtagat cagcctcggg actccacccc 300
agaacttctt ggtctctttt gacaccggct cctccaactt gtggglgccc tctgtctact 360
gccagagaca ggctcgcacc agtcaactcc gctcaacc ccagcagctg tccacctact 420
cczccaatgg gcagaccttc tccctgcagt atggcagtgg cagcctcacc ggcttctttg 480
gctatgacac cctgactgtc cagagcator aggtcccca ccaggagttc ggcttgagtg 540
agaatgagcc tggtaaccaac ttctctatg cgcagtttga tggcctcctg ggcttgacct 600
acctctctct gtcctggat gaggccaca cagctatgca ggccatggtg caggcgggct 660
ccctcaccag ccctgtcttc aggtctaac tcagcaaca ccagggtctc agcaggggag 720
cgtttgtctt tgggggtgta atagcagcc tglacaaggg ccagatctar tggcggctg 780
tcaccacaga actctactgg cagattggca ttgaaagatt cctcatcgcc gcccaggct 840
ccggctgggt ttctgagggt ttgcaggcca tctgtgmac aggcacctct ctgtaactg 900
tgccccacga gtaaatgagt gctctctgc agccacagc gccccaggag gatgagtatg 960
gacagtttct cgtgaactgt aacagcattc agaatctgcc cagcttgacc ttcactatca 1020
atgggtggga gttccctctg ccacctctct cctatatctt cagtaacaac ggctactgca 1080
ccgtgggagt ccagccaccc taccgtctct ccccaacagg ccagcccttg tggatctctg 1140
gggatgtctt cctcaggtcc tactattccg tctacagctt gggcaacac agagtggct 1200
ttgcaactgc ccctagact tctgtctctg aacgtgggc tccccctt ctcttgacc 1260
tgcaccctcc tagggcattg tctctgtctt tccactctgg attcagctt cttttctgg 1320
actctggact ttcttaata ataantagt ttcttaaaaa aaaaaa cttttctgg 1386

```

```

<210> 46
<211> 423
<212> PRF
<213> Homo sapiens

```

```

<220>
<221> Misc_feature
<223> Incyte ID No.: 161156501

```

```

<400> 46
Met Leu Ile Gly Thr Pro Pro Gln Lys Leu Gln Ile Leu Val Asp
1 5 10 15
Thr Gly Ser Ser Asn Phe Ala Val Ala Gly Thr Pro His Ser Tyr
20 25 30 35
Ile Asp Thr Tyr Phe Asp Thr Glu Arg Ser Ser Thr Tyr Arg Ser
35 40 45
Lys Gly Phe Asp Val Thr Val Lys Tyr Thr Gln Gly Ser Trp Thr
50 55 60
Gly Phe Val Gly Glu Asp Leu Val Thr Ile Pro Lys Gly Phe Asn
65 70 75
Thr Ser Phe Leu Val Asn Ile Ala Thr Ile Phe Glu Ser Glu Asn
80 85 90
Phe Phe Leu Pro Gly Ile Lys Trp Asn Gly Ile Leu Gly Leu Ala
95 100 105
Tyr Ala Thr Leu Ala Lys Pro Ser Ser Ser Leu Glu Thr Phe Phe
110 115 120
Asp Ser Leu Val Thr Gln Ala Asn Ile Pro Asn Val Phe Ser Met
125 130 135
Gln Met Cys Gly Ala Gly Leu Pro Val Ala Gly Ser Gly Thr Asn
140 145 150
Gly Gly Ser Leu Val Leu Gly Gly Ile Glu Pro Ser Leu Tyr Lys
155 160 165
Gly Asp Ile Trp Tyr Thr Pro Ile Lys Glu Glu Trp Tyr Tyr
170 175 180
Ile Glu Ile Leu Lys Leu Glu Ile Gly Gly Gln Ser Leu Asn Leu
185 190 195
Asp Cys Arg Glu Tyr Asn Ala Asp Lys Ala Ile Val Asp Ser Gly
200 205 210
Thr Thr Leu Leu Arg Leu Pro Gln Lys Val Phe Asp Ala Val Val
215 220 225
Glu Ala Val Ala Arg Ala Ser Leu Ile Pro Glu Phe Ser Asp Gly
230 235 240

```

WO 01/46398

PCT/US00/33743

```

Phe Trp Thr Gly Ser Gln Leu Ala Cys Trp Thr Asn Ser Glu Thr
245 250 255
Pro Trp Ser Tyr Phe Pro Iys Ile Ser Ile Tyr Leu Arg Asp Glu
260 265 270
Asn Ser Ser Arg Ser Phe Arg Ile Thr Ile Leu Pro Gln Leu Tyr
275 280 285
Ile Gln Pro Met Met Gly Ala Gly Leu Asn Tyr Glu Cys Tyr Arg
290 295 300
Phe Gly Ile Ser Pro Ser Thr Asn Ala Leu Val Ile Gly Ala Thr
305 310 315
Val Met Glu Gly Phe Tyr Val Ile Phe Asp Arg Ala Gln Lys Arg
320 325 330
Val Gly Phe Ala Ala Ser Pro Cys Ala Glu Ile Ala Gly Ala Ala
335 340 345
Val Ser Glu Ile Ser Gly Pro Phe Ser Thr Glu Asp Val Ala Ser
350 355 360
Asn Cys Val Pro Ala Gln Ser Leu Ser Glu Pro Ile Leu Trp Ile
365 370 375
Val Ser Tyr Ala Leu Met Ser Val Cys Gly Ala Ile Leu Leu Val
380 385 390
Leu Ile Val Leu Leu Leu Leu Pro Phe Arg Cys Gln Arg Arg Pro
395 400 405
Arg Asp Pro Glu Val Val Asn Asp Glu Ser Ser Leu Val Arg His
410 415 420
Arg Trp Lys

```

<210> 47

<211> 323

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No.: 1869868

<400> 47

```

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

```

WO 01/46398

PCT/US00/33743

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

<210> 48
 <211> 388
 <212> PRT
 <213> Homo sapiens

<300>
 <308> gl658286

<400> 48
 Met Lys Trp Met Val Val Val Leu Val Cys Leu Gln Leu Leu Glu
 1 5 10 15
 Ala Ala Val Val Lys Val Pro Leu Lys Lys Phe Lys Ser Ile Arg
 20 25 30
 Glu Thr Met Lys Glu Lys Gly Leu Leu Gly Glu Phe Leu Arg Thr
 35 40 45
 His Lys Tyr Asp Pro Ala Trp Lys Tyr Arg Phe Gly Asp Leu Ser
 50 55 60
 Val Thr Tyr Glu Pro Met Ala Tyr Met Asp Ala Ala Tyr Phe Gly
 65 70 75
 Glu Ile Ser Ile Gly Thr Pro Pro Gln Asn Phe Leu Val Leu Phe
 80 85 90
 Asp Thr Gly Ser Ser Asn Leu Trp Val Pro Ser Val Tyr Cys Gln
 95 100 105
 Ser Gln Ala Cys Thr Ser His Ser Arg Phe Asn Pro Ser Glu Ser
 110 115 120
 Ser Thr Tyr Ser Thr Asn Gly Gln Thr Phe Ser Leu Gln Tyr Gly
 125 130 135
 Ser Gly Ser Leu Thr Gly Phe Phe Gly Tyr Asp Thr Leu Thr Val
 140 145 150
 Gln Ser Ile Gln Val Pro Asn Gln Glu Phe Gly Leu Ser Glu Asn
 155 160 165
 Glu Pro Gly Thr Asn Phe Val Tyr Ala Gln Phe Asp Gly Ile Met
 170 175 180
 Gly Leu Ala Tyr Pro Ala Leu Ser Val Asp Glu Ala Thr Thr Ala
 185 190 195
 Met Gln Gly Met Val Gln Glu Gly Ala Leu Thr Ser Pro Val Phe
 200 205 210
 Ser Val Tyr Leu Ser Asn Gln Gln Gly Ser Ser Gly Gly Ala Val
 215 220 225
 Val Phe Gly Gly Val Asp Ser Ser Leu Tyr Thr Gly Gln Ile Tyr
 230 235 240
 Trp Ala Pro Val Thr Gln Glu Leu Tyr Trp Gln Ile Gly Ile Glu

WO 01/46398 PCT/US00/33743

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

<210> 49
 <211> 396
 <212> FAT
 <213> Neurospora crassa

<300>
 <308> g1032445

<400> 49
 Met Lys Gly Ala Leu Leu Thr Ala Ala Met Leu Leu Gly Ser Ala
 1 5 10 15
 Gln Ala Gly Val His Thr Met Lys Leu Lys Lys Val Pro Leu Ala
 20 25 30
 Asp Glu Leu Glu Ser Val Pro Ile Asp Val Gln Val Gln His Leu
 35 40 45
 Gly Gln Lys Tyr Thr Gly Leu Arg Thr Glu Ser His Thr Gln Ala
 50 55 60
 Met Phe Lys Ala Thr Asp Ala Gln Val Ser Gly Asn His Pro Val
 65 70 75
 Pro Ile Thr Asn Phe Met Asn Ala Gln Tyr Phe Ser Glu Ile Thr
 80 85 90
 Ile Gly Thr Pro Pro Gln Thr Phe Lys Val Val Leu Asp Thr Gly
 95 100 105
 Ser Ser Asn Leu Trp Val Pro Ser Ser Gln Cys Gly Ser Ile Ala
 110 115 120
 Cys Tyr Leu His Asn Lys Tyr Glu Ser Ser Glu Ser Ser Thr Tyr
 125 130 135
 Lys Lys Asn Gly Thr Ser Phe Lys Ile Glu Tyr Gly Ser Gly Ser
 140 145 150
 Leu Ser Gly Phe Val Ser Gln Asp Arg Met Thr Ile Gly Asp Ile
 155 160 165
 Thr Ile Asn Asp Gln Leu Phe Ala Glu Ala Thr Ser Glu Pro Gly
 170 175 180
 Leu Ala Phe Ala Phe Gly Arg Phe Asp Gly Ile Leu Gly Leu Gly
 185 190 195
 Tyr Asp Arg Leu Ala Val Pro Gly Ile Thr Pro Pro Phe Tyr Lys
 200 205 210
 Met Val Glu Gln Lys Leu Val Asp Glu Pro Val Phe Ser Phe Tyr
 215 220 225
 Leu Ala Asp Gln Asp Gly Glu Ser Glu Val Val Phe Gly Gly Val
 230 235 240
 Asn Lys Asp Arg Tyr Thr Gly Lys Ile Thr Thr Ile Pro Leu Arg

WO 01/46398

PCT/US00/33743

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

【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
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number
WO 01/46398 A3

- (51) International Patent Classification: C12N 9/64
- (72) Inventors; and
- (75) Inventors/Applicants (or US only): KASER, Matthew, R. [GB/US]; 4793 Ewing Road, Castro Valley, CA 94546 (US); COCKS, Benjamin, G. [AU/US]; 1215 Van Dyke Drive, Sunnyvale, CA 94087 (US); LORING, Jennie [US/US]; 701 Santa Cruz Lane, Foster City, CA 94404 (US); TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US); YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).
- (21) International Application Number: PC17/US0033743
- (74) Agents: TURNER, Christopher et al., Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (22) International Filing Date: 12 December 2000 (12.12.2000)
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CI, 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, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NG, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (39) Priority Data: 09/470,954 22 December 1999 (22.12.1999) US
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, ND, SI, SZ, TZ, UG, ZW); Eurasian
- (62) Related by continuation (CON) or continuation-in-part (CIP) to earlier application: US 09/470,954 (CIP) Filed on 22 December 1999 (22.12.1999)
- (71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

[Continued on next page]

(54) Title: ASPARTYL PROTEASES

```

11      20      29      38      47      56
5' AGC CCT CAC ATA GGG GAT CTG GGC CTC GAG GCA AGA ATT CCG CCG AGC TCA GTG
---
85      94      103      112      121
CCT GAG GAT GGT CTT TTC GAA GGA GGA TCC GCG AAG GCT GCA AAT TCG CCT TCC
---
139     148     157     166     175     184
AGT TTA GAG AGG AAG CTG GCT TCT GCG TTT AAG GCG GTA TGA GGA CTC ACT CTT
---
173     182     191     200     209     218
TGT CTC TTT CTG CTA AAG GTT GAC GAC TCC CTC GAG CCT TTC TTT GAG TAT GTG
---
227     236     245     254     263     272
GTA AAG CAG AGC GAC GTT CCC AAC CTC TTC TCC TGG CAG CTT TGT GGT GCT GGC
---
281     290     299     308     317     326
TTC CCC TTC AAC GAG TCC GAA GTG CTG GCT TCT GGT GGG GGG AAG ATC ATT
---
P F L N Q S E V L A S V G C S N I I

```

A

(57) Abstract: The invention provides mammalian nucleic acid molecules and fragments thereof. It also provides for the use of the mammalian nucleic acid molecules for the characterization, diagnosis, evaluation, treatment, or prevention of conditions, diseases and disorders associated with Alzheimer's disease and Down syndrome. The invention additionally provides expression vectors and host cells for the production of the protein encoded by the mammalian nucleic acid molecules.



WO 01/46398 A3

WO 01/46398 A3



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, CI,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report:
10 January 2002

Published:
with international search report

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

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 00/33743
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N/64 According to International Patent Classification (IPC) as to date mentioned in priority and IP.		
B. FIELDS SEARCHED *Inventive search(s) searched (classification system followed by classification symbols) IPC 7 C12N Document(s) searched other than minimum document(s) to the extent that such document(s) are included in the fields searched *Database(s) used (including internet) (the international search (or other) data base(s), where practical, search fields used) STRAND, EP0-Internal, WPI Data, PAJ		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Character of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 848 062 A (SMITHKLINE BEECHAM PLC ; SMITHKLINE BEECHAM CORP (US)) 17 June 1998 (1998-06-17) See especially residues 96-518 of table 2 which are 100% identical to SEQ ID NO:46 of the present application. page 1-5; claims 1-19; tables 1,2 ---	I-20
A	YAN ET AL.: "Membrane-anchored aspartyl protease with Alzheimer's disease beta-secretase activity" NATURE, vol. 402, 2 December 1999 (1999-12-02), pages 533-537, XP002167053 the whole document --- -/--	I-20
<input checked="" type="checkbox"/> Father documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex		
* Special categories of cited documents: "A" documents defining the general state of the art which is not considered to be of particular relevance "E" earlier document published on or after the international filing date "L" document which may have priority, essential to which is cited to establish the publication date of another claim or other specific provision of the application "O" document relating to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but after the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principles or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or inventive as disclosed, or involve an inventive step when the document is taken into account "Y" document of the closest reference; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
10 May 2001		01.08.01
Name and mailing address of the ISA Examiner Patent Office, P.O. Box 1618, Patent van 2 NL - 2202 HV Rijswijk Tel. (+31-70) 360 2040, Fax. (+31-651) 460 01, Fax. (+31-70) 342 3816		Author of the report Sprinks, M

Form PCT/ISA/210 (second sheet), July 2002

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 00/33743

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Location of document, with references, where appropriate, of the relevant passages	Publication No.
A	DATABASE EMBL [Online] EMBL:AF190727, 26 October 1999 (1999-10-26) VASSAR ET AL.: "Rattus norvegicus beta-site APP cleaving enzyme (Bace) mRNA" XP002167054 abstract -----	1-20

1

Form PCT/ISA/210 (Continuation of International Search Report) July 2001

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/33743

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

- 1. Claims Nos. :
because they relate to subject matter not required to be searched by this Authority, namely:
- 2. Claims Nos. :
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
- 3. Claims Nos. :
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(p).

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. As a fee required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
- 2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not make payment of any additional fee.
- 3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claim nos.:
- 4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the inventions first mentioned in the claims; it is covered by claims Nos.:

1,14-20 totally; 2-13 partially

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
- No protest accompanied the payment of additional search fees.

International Application No. PCT/US 00/33743

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. Claims: 1,14-20 totally; 2-13 partially

Invention 1: An aspartyl protease nucleic acid sequence (SEQ ID NO:43) and subject-matter relating thereto.

2. Claims: 2,3,4-13 partially

Inventions 2-33: Further aspartyl protease nucleic acid sequences corresponding to SEQ ID NOs:1-5,7-29,31-33,37 respectively and subject-matter relating thereto.

3. Claims: 2,4-13 partially

Inventions 34-44: Further aspartyl protease nucleic acid sequences corresponding to SEQ ID NOs:6,30,34-36,38-42 and 44 respectively and subject-matter relating thereto.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 00/33743

Patent document used in search report	Publication date	Patent family member(s)	Publication date
EP 0848062 A	17-06-1998	JP 11069981 A	16-03-1999
		US 6162630 A	19-12-2000
		US 6025180 A	15-02-2000

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
C 1 2 N 1/21	C 1 2 N 9/64	Z 4 C 0 8 4
C 1 2 N 5/10	C 1 2 Q 1/68	A
C 1 2 N 9/64	G 0 1 N 33/15	Z
C 1 2 Q 1/68	G 0 1 N 33/50	Z
G 0 1 N 33/15	G 0 1 N 33/53	M
G 0 1 N 33/50	G 0 1 N 33/566	
G 0 1 N 33/53	C 1 2 N 5/00	A
G 0 1 N 33/566	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, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW

(72) 発明者 コックス、ベンジャミン・ジー
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・ファンダイクドライブ 1 2 1 5

(72) 発明者 ローリング、ジャンヌ
アメリカ合衆国カリフォルニア州 9 4 4 0 4 ・ フォスターシティ・サンタクルスレーン 7 0 1

(72) 発明者 タング、ワイ・トム
アメリカ合衆国カリフォルニア州 9 5 1 1 8 ・ サンノゼ・ランウィックコート 4 2 3 0

(72) 発明者 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

F ターム(参考) 2G045 AA25 AA29 AA40 BA13 BB20 CB17 CB21 DA12 DA13 DA14
DA36 DA77 FB02 FB03 FB04 FB07
4B024 AA01 AA11 BA14 CA02 CA04 CA09 FA02 HA01 HA14
4B050 CC03 DD11 LL01 LL03
4B063 QA18 QQ44 QR55 QS34
4B065 AB01 BA01 CA33 CA44 CA46
4C084 AA02 AA07 BA01 BA23 CA53 CA56 DC01 NA14 ZA02 ZA16

专利名称(译)	天冬氨酰蛋白酶		
公开(公告)号	JP2004500811A	公开(公告)日	2004-01-15
申请号	JP2001546896	申请日	2000-12-12
[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	カイザーマシューアール コックスベンジャミンジー ローリングジャンヌ タングワイトム ユエヘンリー		
发明人	カイザー、マシュー・アール コックス、ベンジャミン・ジー ローリング、ジャンヌ タング、ワイトム ユエ、ヘンリー		
IPC分类号	G01N33/50 A61K38/00 A61K38/43 A61P25/28 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/64 C12N15/09 C12Q1/68 G01N33/15 G01N33/53 G01N33/566		
CPC分类号	A61K38/00 A61P25/28 C12N9/6478		
FI分类号	C12N15/00.ZNA.A A61P25/28 C12N1/15 C12N1/19 C12N1/21 C12N9/64.Z C12Q1/68.A G01N33/15.Z G01N33/50.Z G01N33/53.M G01N33/566 C12N5/00.A A61K37/48		
F-TERM分类号	2G045/AA25 2G045/AA29 2G045/AA40 2G045/BA13 2G045/BB20 2G045/CB17 2G045/CB21 2G045 /DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/DA77 2G045/FB02 2G045/FB03 2G045/FB04 2G045/FB07 4B024/AA01 4B024/AA11 4B024/BA14 4B024/CA02 4B024/CA04 4B024/CA09 4B024 /FA02 4B024/HA01 4B024/HA14 4B050/CC03 4B050/DD11 4B050/LL01 4B050/LL03 4B063/QA18 4B063/QQ44 4B063/QR55 4B063/QS34 4B065/AB01 4B065/BA01 4B065/CA33 4B065/CA44 4B065 /CA46 4C084/AA02 4C084/AA07 4C084/BA01 4C084/BA23 4C084/CA53 4C084/CA56 4C084/DC01 4C084/NA14 4C084/ZA02 4C084/ZA16		
优先权	09/470954 1999-12-22 US		
外部链接	Espacenet		

摘要(译)

本发明提供了哺乳动物核酸分子及其片段。它还提供了哺乳动物核酸分子用于表征，诊断，评估，治疗或预防与阿尔茨海默氏病和唐氏综合症有关的状况，疾病和病症的用途。本发明另外提供表达载体和宿主细胞，用于产生由哺乳动物核酸分子编码的蛋白质。