

SPR4D3_SEQID.2

ATGACCGATCGCTGCTGAAACCTTTATCTTTTGGGAGTCTGGTCTCTGGAGGCGTGGGATATGATGAGATCGTTA
 GCCCAACGACGACACAGAGAGATTTGTGCAGACAGTGATGTCATCGACAGCTGGGACAGATTCCTTATTAACATATTA
 CACCCCATCGGAGAGATCTATGATGGATGATGAGAGAGATCTGATGAGATGACAGGAAGATGGTGCAGACAGATTT
 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
 CATTGGTGACATCGCTGTGATTACCGCCGAGAGATGGATGTCTGCTGCTTATTTTCCCAATGGTCTGCACCAAAAT
 TCTACAAACAGCTAAAGACAGCTCAAGTATACGACGAGATCTGACCAAGCTGGATCTATGTCTTCGAGTTCT
 AACATAGAGTATATATCTACATCTGAGCAACATGATCTGCTTTGATGAGTGGTTCGCGCTTACCCCATATATATGAGCA
 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
 TCAACATCTTCTGGGACACGGCCAGCTCTGAAACACGAGC/AAAGCTTGTCGAGTCTACATAGAGAGGAGCA
 AAGATAGATATTTTGTCTCTCGCTGACATCAATCTGATCTGGGAGTATATCTTCAACATCTAACGCGTACACAAAA
 AATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
 CCTTATATAGAGTGGATGGAGTCAGCATATTTTATGACCTTAC/CAGGTCCTTCAAGAGATGGGGCGGACAGCA
 TGGGATTCCTTCTCATATAGATCTTGAAGCTGAGGACAGAGTGAAGCATATGGGTTTGTGTGCGAAGAGCTGAGAT
 CAGCCGACCATCTATGACGAGCTAGTGAGGCTGTGCTGTGCTGAGCTCGATGATGTCTGTCGAAACAACTGGGACAT

【特許請求の範囲】

【請求項 1】

単離された、濃縮された、または精製された核酸分子であって、前記核酸分子は、

(a) 配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 に記載されるアミノ酸配列およびその生物学的ドメインからなる群より選択されるアミノ酸配列を有するポリペプチドをコードする;

(b) (a) のヌクレオチド配列の相補体である; または

(c) ストリンジェントな条件下で (a) のヌクレオチド分子にハイブリダイズし、かつプロテアーゼポリペプチドをコードする

ヌクレオチド配列を含むことを特徴とする核酸分子。

【請求項 2】

ヌクレオチド配列と動作可能なように連結されたベクターまたはプロモーターをさらに含む、請求項 1 記載の核酸分子。

【請求項 3】

前記核酸分子が、哺乳動物から単離され、濃縮され、または精製されたものである、請求項 1 記載の核酸分子。

【請求項 4】

前記哺乳動物がヒトである、請求項 3 記載の核酸分子。

【請求項 5】

ストリンジェントな条件下で、配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を有するプロテアーゼポリペプチドをコードするヌクレオチド配列にハイブリダイズするヌクレオチド配列を含む核酸を含む、請求項 1 記載の核酸分子。

【請求項 6】

単離され、濃縮され、または精製されたプロテアーゼポリペプチドであって、配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列

番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および 配列番号 118 に記載される配列およびその生物学的ドメインからなる群より選択される配列と少なくとも約 90% の同一性を有するアミノ酸配列を含むポリペプチド。

【請求項 7】

前記ポリペプチドが、哺乳動物から単離され、精製され、または濃縮されたものである、10
請求項 6 記載のプロテアーゼポリペプチド。

【請求項 8】

前記哺乳動物がヒトである、請求項 7 記載のプロテアーゼポリペプチド。

【請求項 9】

プロテアーゼポリペプチドまたは前記ポリペプチドのドメインに対して特異的結合親和性を有する抗体または抗体フラグメントであって、前記ポリペプチドは、配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および 配列番号 118 に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を含むことを特徴とする抗体または抗体フラグメント。20

【請求項 10】

請求項 9 記載の抗体を産生するハイブリドーマ。30

【請求項 11】

請求項 6 記載のポリペプチドに結合する抗体および負対照抗体を含むキット。

【請求項 12】

プロテアーゼポリペプチドの活性を調節する物質を同定する方法であって、
(a) 配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および 配列番号 118 に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列と実質的に同一のプロテアーゼポリペプチドを試験物質と接触させ；40

(b) 前記ポリペプチドの活性を測定し；そして

(c) 前記物質が前記ポリペプチドの活性を調節するか否かを判定する、
の各工程を含む方法。50

【請求項 13】

細胞においてプロテアーゼポリペプチドの活性を調節する物質を同定する方法であって、
(a) プロテアーゼポリペプチドを細胞中で発現させ、ここで、前記ポリペプチドは、
配列番号 60、配列番号 61、配列番号 62、配列番号 63、配列番号 64、配列番号 65、
配列番号 66、配列番号 67、配列番号 68、配列番号 69、配列番号 70、配列番号 71、
配列番号 72、配列番号 73、配列番号 74、配列番号 75、配列番号 76、配列番号 77、
配列番号 78、配列番号 79、配列番号 80、配列番号 81、配列番号 82、配列番号 83、
配列番号 84、配列番号 85、配列番号 86、配列番号 87、配列番号 88、配列番号 89、
配列番号 90、配列番号 91、配列番号 92、配列番号 93、配列番号 94、配列番号 95、
配列番号 96、配列番号 97、配列番号 98、配列番号 99、配列番号 100、配列番号 101、
配列番号 102、配列番号 103、配列番号 104、配列番号 105、配列番号 106、配列番号 107、
配列番号 108、配列番号 109、配列番号 110、配列番号 111、配列番号 112、配列番号 113、
配列番号 114、配列番号 115、配列番号 116、配列番号 117 および配列番号 118 に記載されるア
ミノ酸配列からなる群より選択されるアミノ酸配列と実質的に同一の配列を含み；
(b) 試験物質を前記細胞に加え；そして
(c) 細胞表現型、細胞増殖、細胞分化または前記ポリペプチドと天然の結合パートナー
との間の相互作用の変化をモニターする、
の各工程を含む方法。

【請求項 14】

治療を必要とする患者に、配列番号 60、配列番号 61、配列番号 62、配列番号 63、
配列番号 64、配列番号 65、配列番号 66、配列番号 67、配列番号 68、配列番号 69、
配列番号 70、配列番号 71、配列番号 72、配列番号 73、配列番号 74、配列番号 75、
配列番号 76、配列番号 77、配列番号 78、配列番号 79、配列番号 80、配列番号 81、
配列番号 82、配列番号 83、配列番号 84、配列番号 85、配列番号 86、配列番号 87、
配列番号 88、配列番号 89、配列番号 90、配列番号 91、配列番号 92、配列番号 93、
配列番号 94、配列番号 95、配列番号 96、配列番号 97、配列番号 98、配列番号 99、
配列番号 100、配列番号 101、配列番号 102、配列番号 103、配列番号 104、配列番号 105、
配列番号 106、配列番号 107、配列番号 108、配列番号 109、配列番号 110、配列番号 111、
配列番号 112、配列番号 113、配列番号 114、配列番号 115、配列番号 116、配列番号 117 および配列
番号 118 に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列と実質的に
同一のプロテアーゼの活性を調節する物質を投与することにより疾病または疾患を治療す
る方法。

【請求項 15】

前記疾病または疾患が、癌、免疫関連疾病および疾患、心臓血管疾病、脳またはニューロン
関連疾病、代謝性疾患および炎症性疾患からなる群より選択される、請求項 14 記載の
方法。

【請求項 16】

前記疾病または疾患が、組織の癌；血液または造血細胞起源の癌；乳、結腸、肺、前立腺
、子宮頸部、脳、卵巣、膀胱または腎臓の癌からなる群より選択される、請求項 15 記載
の方法。

【請求項 17】

前記疾病または疾患が、中枢神経系または末梢神経系疾病、片頭痛；痛み；性的機能不全
；気分障害；注意障害；認識障害；低血圧症；高血圧症；精神病性疾患；神経性疾患およ
び運動異常症からなる群より選択される、請求項 15 記載の方法。

【請求項 18】

前記物質がインビトロでプロテアーゼ活性を調節する、請求項 15 記載の方法。

【請求項 19】

前記物質がプロテアーゼ阻害剤である、請求項 18 記載の方法。

10

20

30

40

50

【請求項 20】

疾病または疾患の診断道具として試料中のプロテアーゼポリペプチドを検出する方法であって、

(a) 前記試料を、ハイブリダイゼーションアッセイ条件下で配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および 配列番号 118 に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を有するプロテアーゼポリペプチドの核酸標的領域にハイブリダイズする核酸プローブと接触させ、前記プローブは、ポリペプチド、そのフラグメント、および配列およびフラグメントの相補体をコードする核酸配列を含み；そして

10

(b) 疾病の指標としてプローブ：標的領域ハイブリッドの存在または量を検出するの各工程を含む方法。

20

【請求項 21】

前記疾病または疾患が、癌、免疫関連疾病および疾患、心臓血管疾病、脳またはニューロン関連疾病、代謝性疾患および炎症性疾患からなる群より選択される、請求項 20 記載の方法。

【請求項 22】

前記疾病または疾患が、組織の癌；血液または造血細胞起源の癌；乳、結腸、肺、前立腺、子宮頸部、脳、卵巣、膀胱または腎臓の癌からなる群より選択される、請求項 21 記載の方法。

【請求項 23】

前記疾病または疾患が、中枢神経系または末梢神経系疾病、片頭痛、痛み；性的機能不全；気分障害；注意障害；認識障害；低血圧症；高血圧症；精神病性疾患；神経性疾患；および運動異常症からなる群より選択される、請求項 21 記載の方法。

30

【請求項 24】

配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および 配列番号 118 からなる群より選択される配列を有するプロテアーゼポリペプチドの生物学的ドメインをコードする核酸分子を含む、請求項 1 記載の単離された、濃縮された、または精製された核酸分子。

40

【請求項 25】

配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配

50

列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 に記載されるポリペプチドからなる群より選択されるポリペプチドと少なくとも 90% の同一性を有するアミノ酸配列を有するプロテアーゼポリペプチドをコードする核酸配列を含む, 請求項 1 記載の核酸分子。 10

【請求項 26】

分子が, 配列番号 1, 配列番号 2, 配列番号 3, 配列番号 4, 配列番号 5, 配列番号 6, 配列番号 7, 配列番号 8, 配列番号 9, 配列番号 10, 配列番号 11, 配列番号 12, 配列番号 13, 配列番号 14, 配列番号 15, 配列番号 16, 配列番号 17, 配列番号 18, 配列番号 19, 配列番号 20, 配列番号 21, 配列番号 22, 配列番号 23, 配列番号 24, 配列番号 25, 配列番号 26, 配列番号 27, 配列番号 28, 配列番号 29, 配列番号 30, 配列番号 31, 配列番号 32, 配列番号 33, 配列番号 34, 配列番号 35, 配列番号 36, 配列番号 37, 配列番号 38, 配列番号 39, 配列番号 40, 配列番号 41, 配列番号 42, 配列番号 43, 配列番号 44, 配列番号 45, 配列番号 46, 配列番号 47, 配列番号 48, 配列番号 49, 配列番号 50, 配列番号 51, 配列番号 52, 配列番号 53, 配列番号 54, 配列番号 55, 配列番号 56, 配列番号 57, 配列番号 58, および配列番号 59 からなる群より選択される配列と実質的に同一のヌクレオチド配列を含む, 請求項 1 記載の核酸分子。 20

【請求項 27】

配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 からなる群より選択されるポリペプチドをコードする核酸配列の約 10-30 個の連続するヌクレオチド塩基から本質的になる, 単離された, 濃縮された, または精製された核酸分子。 30

【請求項 28】

配列番号 1, 配列番号 2, 配列番号 3, 配列番号 4, 配列番号 5, 配列番号 6, 配列番号 7, 配列番号 8, 配列番号 9, 配列番号 10, 配列番号 11, 配列番号 12, 配列番号 13, 配列番号 14, 配列番号 15, 配列番号 16, 配列番号 17, 配列番号 18, 配列番号 19, 配列番号 20, 配列番号 21, 配列番号 22, 配列番号 23, 配列番号 24, 配列番号 25, 配列番号 26, 配列番号 27, 配列番号 28, 配列番号 29, 配列番号 30, 配列番号 31, 配列番号 32, 配列番号 33, 配列番号 34, 配列番号 35, 配列番号 36, 配列番号 37, 配列番号 38, 配列番号 39, 配列番号 40, 配列番号 41, 配列番号 42, 配列番号 43, 配列番号 44, 配列番号 45, 配列番号 46, 配列番号 47, 配列番号 48, 配列番号 49, 配列番号 50, 配列番号 51, 配列番号 52, 配列番号 53, 配列番号 54, 配列番号 55, 配列番号 56, 配列番号 57, 配列番号 58, および配列番号 59 からなる群より選択される核酸配列の約 10-30 個の連続するヌクレオチ 40 50

ド塩基から本質的になる，請求項 27 記載の単離された，濃縮された，または精製された核酸分子。

【請求項 29】

請求項 1 記載の核酸分子を含む組換え細胞。

【請求項 30】

試料中のプロテアーゼポリペプチドの存在または量を検出する方法であって，

(a) 試料をプロテアーゼ抗体免疫複合体が形成されるのに適した条件下で請求項 9 記載の抗体と接触させ；そして

(b) プロテアーゼポリペプチドとコンジュゲートした抗体の存在または量を検出する，の各工程を含む方法。

10

【発明の詳細な説明】

【0001】

本発明は，米国特許仮出願 60/214,047 (2000 年 6 月 26 日出願) に基づく優先権を主張する。この出願はその全体を本明細書の一部としてここに引用する。

【0002】

発明の分野

本発明は，プロテアーゼポリペプチド，プロテアーゼポリペプチドをコードするヌクレオチド配列，ならびに種々のプロテアーゼ関連疾病および状態の診断および治療に有用な種々の産物および方法に関する。

【0003】

20

発明の背景

プロテアーゼおよびヒト疾病

”プロテアーゼ”，”プロテイナーゼ”，および”ペプチダーゼ”は，ペプチド結合を加水分解するすべての酵素，すなわち蛋白質分解酵素に適用される類義語である。プロテアーゼは，医学研究およびバイオテクノロジーにおいて非常に重要な酵素のグループである。これらは，すべての生きている生物の生存に必要であり，すべての哺乳動物遺伝子の 1-2% によりコードされている。Rawlings および Barrett (MEROPS : ペプチダーゼデータベース. Nucleic Acids Res., 1999, 27 : 325-331) (<http://www.babraham.co.uk/Merops/Merops.htm> (図面および表を含めその全体を本明細書の一部としてここに引用する)) は，ペプチダーゼを触媒コア配列における構造的類似性に基づいて 157 のファミリーに分類した。これらのファミリーは，さらに共通の進化的関係の表示に基づいて 26 のクランに分類される。ペプチダーゼは，哺乳動物細胞において，正常な生理学および疾病関連経路の両方において鍵となる役割を果たす。例としては，アポトーシスの調節 (カスパーゼ)，血圧の調節 (レニン，アンジオテンシン変換酵素)，組織リモデリングおよび腫瘍侵襲 (コラゲナーゼ)，アルツハイマー病の発症 (β -セクレターゼ)，蛋白質ターンオーバーおよび細胞サイクル制御 (プロテオソーム)，および炎症 (TNF- α コンベルターゼ) (Barrett et al., Handbook of Proteolytic Enzymes, 1998, Academic Press, San Diego (図面および表を含めその全体を本明細書の一部としてここに引用する)) が挙げられる。

30

40

【0004】

ペプチダーゼは，エキソペプチダーゼまたはエンドペプチダーゼのいずれかに分類される。エキソペプチダーゼは，ポリペプチド鎖の末端付近においてのみ作用する。アミノペプチダーゼは遊離 N 末端で作用し，カルボキシペプチダーゼは遊離 C 末端で作用する。エンドペプチダーゼは，その作用メカニズムに基づいて以下の 6 つのサブクラスに分けられる：アスパルチルエンドペプチダーゼ (3.4.23)，システインエンドペプチダーゼ (3.4.22)，メタロエンドペプチダーゼ (3.4.24)，セリンエンドペプチダーゼ (3.4.21)，トレオニンエンドペプチダーゼ (3.4.25)，および上述のクラスのいずれにも属さない最後のグループ (3.4.99) (酵素の命名法および番号付

50

けは "Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) 1992, (<http://www.chem.qmw.ac.uk/iubmb/enzyme/EC34/intro.html>) に基づく)。

【0005】

セリンー、トレオニンーおよびシステインータイプのペプチダーゼにおいては、触媒的求核基はアミノ酸側鎖の反応基であり、ヒドロキシル基（セリンおよびトレオニンタイプペプチダーゼ）またはスルフヒドリル基（システインータイプペプチダーゼ）のいずれかである。アスパラギン酸タイプおよびメタロペプチダーゼにおいては、求核基は一般に活性化された水分子である。アスパラギン酸タイプペプチダーゼにおいては、水分子はアスパラギン酸残基の側鎖に直接結合している。メタロペプチダーゼにおいては、1または2個の金属イオンが水分子をその位置に保持し、荷電アミノ酸側鎖が金属イオンのリガンドである。金属は、亜鉛、コバルトまたはマンガンでありうる。通常は1個の金属イオンが3個のアミノ酸リガンドに結合している。ペプチダーゼのファミリーは、Rawlings & Barrett (Rawlings, N. D. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Research* 27 (1999) 325-331 (図面および表を含めその全体を本明細書の一部としてここに引用する)の番号付けシステムを用いることにより参照される。酵素の命名法および番号付けは、"Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) 1992, (<http://www.chem.qmw.ac.uk/iubmb/enzyme/EC34/intro.html>) に基づく。

10

20

【0006】

プロテアーゼファミリー

1. アスパルチルプロテアーゼ (Prosite 番号 PS00141)

アスパルチルプロテアーゼは、酸プロテアーゼとしても知られており、脊椎動物、真菌、レトロウイルスおよびある種の植物ウイルスに広く分布している蛋白質分解酵素のファミリーである。真核生物のアスパラギン酸プロテアーゼは2つのドメインから構成される単量体酵素である。各ドメインは、触媒的アスパラギン酸残基の中心に活性部位を含む。2つのドメインは、おそらく原始ドメインをコードする先祖遺伝子の複製から進化したものである。このクラスの酵素には、カテプシンE、レニン、プレセニリン (PS1)、およびAPPセクレターゼが含まれる。

30

【0007】

2. システインプロテアーゼ (Prosite PDOC00126)

真核生物システインプロテアーゼは、活性部位システインを含む蛋白質分解酵素のファミリーである。触媒作用はチオエステル中間体を通して進行し、近くのヒスチジン側鎖により促進される。アスパラギンが必須の触媒的三者系を完成する。疾病に重要な役割を果たすこのファミリーのペプチダーゼには、カスパーゼ、カルパイン、ヘッジホッグ、ユビキチンヒドロラーゼ、およびパパインが含まれる。

40

【0008】

3. メタロプロテアーゼ (Prosite PDOC00129)

メタロプロテアーゼは、マトリクスメタロプロテアーゼ (MMP)、コラゲナーゼ、ストロメリシン、ゼラチナーゼ、ネプリリシン、カルボキシペプチダーゼ、ジペプチダーゼ、および膜付随メタロプロテアーゼ、例えばADAMファミリーのプロテアーゼを含むクラスである。これらは活性に金属補因子を必要とする。頻繁に必要とされる金属イオンは亜鉛であるが、ある種のメタロプロテアーゼはコバルトおよびマンガンを利用する。

【0009】

50

細胞外マトリクスの蛋白質は細胞表面レセプターと直接相互作用し、このことにより、シグナル伝達経路を開始させ、成長因子により誘発される経路を調節する。あるものは、最適な活性のためには細胞外マトリクスに結合することを必要とする。したがって、細胞外マトリクスは、これに包まれている細胞およびこれに隣接する細胞に深い影響を有する。細胞外マトリクスのリモデリングには、いくつかのファミリーのプロテアーゼ、例えばメタロプロテアーゼ (MMP) が必要である。

【0010】

4. セリンプロテアーゼ (S1) (Prosite PS00134トリプシン-his ; PS00135トリプシン-ser)

トリプシンファミリーのセリンプロテアーゼの触媒活性は、ヒスチジン (それ自身はセリンと水素結合している) と水素結合しているアスパラギン酸残基が関与する荷電リレーションシステムにより与えられる。活性部位のセリンおよびヒスチジン残基付近の配列は、プロテアーゼのこのファミリーでよく保存されている。この大きくかつ重要なファミリーに属することが知られているプロテアーゼの部分的リストには以下のものが含まれる：血液凝固因子VIL, IX, X, XIおよびXII；トロンビン；プラスミノゲン；補体成分C1r, C1s, C2；補体因子B, DおよびI；RA-反応性因子の補体活性化成分；エラスターゼ1, 2, 3A, 3B (プロテアーゼE)；肝細胞成長因子アクチベータ；腺 (組織) カリクレイン, 例えばEGF-結合蛋白質タイプA, B, およびC；NGF- γ 鎖, γ -レニン, および前立腺特異的抗原 (PSA)；血漿カリクレイン；マスト細胞プロテアーゼ；ミエロプラスチン (プロテイナーゼ3) (ウェゲナー自己抗原)；プラスミノゲンアクチベータ (ウロキナーゼタイプおよび組織タイプ)；およびトリプシンI, II, III, およびIV。これらのペプチダーゼは、凝固、腫瘍発生、血圧の調節、成長因子の放出および他の役割を果たす。

【0011】

5. トレオニンペプチダーゼ (T1) - (Prosite PDOC00326 / PDOC00668)

トレオニンプロテアーゼは、これらの酵素の触媒部位のトレオニン残基のヒドロキシル基を利用することを特徴とする。これまでに、これらの酵素のわずか、例えば、始原細菌である *Thermoplasma acidophilum* の 20S プロテオソームしか特性決定されていない (Seemuller et al., 1995, Science, 268: 579-82, および chapter 167, Barrett et al., Handbook of Proteolytic Enzymes, 1998, Academic Press, San Diego)。

【0012】

発明の概要

本発明は、ヒトプロテアーゼの新規配列の単離および特性決定に関する。これらの配列は、新規ヒト遺伝子配列の予測アミノ酸翻訳に対するバイオインフォマティクス検索により得られる。プロテアーゼとして同定されたこれらの配列をポリペプチドに翻訳し、これをさらに特性決定する。さらに、これらのプロテアーゼの核酸配列を用いてプロテアーゼの全長cDNAクローンを得る。これらのプロテアーゼの部分配列または完全配列が、その分類、予測されたまたは推定された蛋白質構造とともに本明細書に記載される。

【0013】

これらのプロテアーゼの活性の調節は、治療上有用であることがわかるであろう。さらに、これらのプロテアーゼまたはこれらをコードするDNA配列の存在または非存在は、種々の疾病の診断または予期に有用であることがわかるであろう。この点に関して、実施例8は、本発明のプロテアーゼの染色体位置を記述し、本発明のプロテアーゼの染色体位置への疾病マッピングを記述する。

【0014】

本発明の第1の観点は、配列番号60, 配列番号61, 配列番号62, 配列番号63, 配列番号64, 配列番号65, 配列番号66, 配列番号67, 配列番号68, 配列番号69

10

20

30

40

50

、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列およびそれらの生物学的ドメインからなる群より選択されるアミノ酸配列を有する、同定された、単離された、濃縮されたまたは精製された核酸分子を特徴とする。 10

【0015】

核酸に関して”同定された”との用語は、これまでに知られていない新規プロテアーゼの一部をコードすると予測されることに基づいて、ゲノム、EST、またはcDNA配列データベースから配列が選択されたことを意味する。

【0016】

核酸に関して”単離された”とは、互いに結合した10個（好ましくは21個、より好ましくは39個、最も好ましくは75個）またはそれ以上のヌクレオチドのポリマーを意味し、天然起源から単離された、またはセンス鎖または相補的なアンチセンス鎖として合成されるDNAおよびRNAが含まれる。本発明のある態様においては、より長い核酸が好ましく、例えば、300、600、900、1200、1500、またはそれ以上のヌクレオチドのもの、および／または配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、配列番号21、配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、配列番号28、配列番号29、配列番号30、配列番号31、配列番号32、配列番号33、配列番号34、配列番号35、配列番号36、配列番号37、配列番号38、配列番号39、配列番号40、配列番号41、配列番号42、配列番号43、配列番号44、配列番号45、配列番号46、配列番号47、配列番号48、配列番号49、配列番号50、配列番号51、配列番号52、配列番号53、配列番号54、配列番号55、配列番号56、配列番号57、配列番号58、および配列番号59に記載される配列からなる群より選択される配列と、少なくとも50%、60%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、または99%の同一性を有するものが好ましい。 20 30

【0017】

核酸とは、限定されないが、DNA、RNAまたはcDNAを意味し、核酸がRNAである場合、チミン（T）はウラシルであることが理解される。

【0018】

本発明の単離された核酸は、これが自然には純粋なまたは分離された状態で見いだされないという点において独特である。”単離された”との用語の使用は、天然に生ずる配列がその通常の細胞環境（すなわち染色体）から除かれていることを表す。すなわち、配列は、無細胞溶液中にあってもよく、または異なる細胞環境に置かれていてもよい。この用語は、この配列が存在する唯一のヌクレオチド鎖であることを意味するものではなく、天然にこれに付随する非ヌクレオチド物質を本質的に含まず（好ましくは約90%純粋、より好ましくは少なくとも95%純粋）、したがって、単離された染色体とは区別されることを意味する。 40

【0019】

核酸に関連して、”濃縮された”との用語の使用は、特定のDNAまたはRNA配列が、 50

目的とする細胞または溶液中に存在する総DNAまたはRNA中で、正常または疾病細胞、またはこの配列が由来する細胞におけるより有意に高い割合（2－5倍）を占めることを意味する。これは、存在する他のDNAまたはRNAの量の優先的減少、または特定のDNAまたはRNA配列の量の優先的増加、またはこれらの2つの組み合わせにより、人が生じさせることができる。しかし、濃縮されたとは、他のDNAまたはRNA配列が存在しないことを意味するものではなく、単に、目的とする配列の相対的な量が有意に増加されていることを意味することに注意すべきである。”有意に”との用語は、増加のレベルがそのような増加を作成した人にとって有用であることを示すために用いられ、一般に、他の核酸に比べて少なくとも約2倍、より好ましくは少なくとも5倍、より好ましくは少なくとも10倍またはそれ以上増加していることを意味する。またこの用語は、他の起源からのDNAまたはRNAが存在しないことを意味するものではない。他の起源のDNAは、例えば、酵母または細菌ゲノム、またはクローニングベクター、例えばpUC19からのDNAでありうる。この用語は、1つのmRNAのレベルが他の種のmRNAと比較して自然に増加している天然に生ずる事象、例えばウイルス感染または腫瘍タイプの成長から区別される。すなわち、この用語は、人が介在して所望の核酸の比率を上昇させる状況のみをカバーする。

【0020】

ある目的のためには、ヌクレオチド配列が精製された形であることも有利である。核酸に関して、”精製された”との用語は、絶対的純度（例えば均一な調製物）を要求するものではない。むしろ、これは配列が天然の環境におけるより比較的純粋であることを示す（天然のレベルと比較して、このレベルは、例えばmg/mLで少なくとも2－5倍高い）。cDNAライブラリから単離された個々のクローンは、電気泳動的に均一にまで精製することができる。これらのクローンから得られた本発明のDNA分子は、総DNAからまたは総RNAから直接得ることができる。cDNAクローンは天然に生じず、好ましくは部分的に精製した天然に生ずる物質（メッセンジャーRNA）の操作により得る。mRNAからのcDNAライブラリの構築は、合成物質（cDNA）の作成を含み、純粋な個々のcDNAクローンは、cDNAライブラリを有する細胞のクローン選択により合成ライブラリから単離することができる。すなわち、mRNAからcDNAライブラリを構築し、個々のcDNAクローンを単離することを含む工程により、天然のメッセンジャーのおよそ 10^6 倍の精製が得られる。すなわち、少なくとも1桁、好ましくは2または3桁、より好ましくは4または5桁の精製が明示的に企図される。

【0021】

”プロテアーゼポリペプチド”とは、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列およびその生物学的ドメインからなる群より選択されるアミノ酸配列を有するポリペプチド中の、32個（好ましくは40、より好ましくは45、最も好ましくは55個）またはそれ以上の連続するアミノ酸を意味する。ある観点においては、100、200、300、400、450、500、550、600、700、800、900個またはそれ以上のアミノ酸のポリペプチドが好ましい。プロテアーゼポリペプチドは、ポリペプチドの機能的活性が保持される限り、全長核酸配列または全長核酸配列の任意の部分によりコードされることができる。遺伝コードの縮重

のため、多数の異なる核酸配列が同じアミノ酸配列をコードしうることは当該技術分野においてよく知られている。同じく、アミノ酸の保存的変更を行って、元の機能を保持している蛋白質またはポリペプチドを得ることができることも、当該技術分野においてよく知られている。そのような置換には、アミノ酸を類似の物理化学的特性を有する残基で置き換えること、例えば、1つの脂肪族残基 (I l e, V a l, L e u または A l a) を別のもので、または塩基性残基 L y s と A r g, 酸性残基 G l u と A s p, アミド残基 G l n と A s n, ヒドロキシル残基 S e r と T y r, または芳香族残基 P h e と T y r との間で置き換えることが含まれる。蛋白質全体に対して、あったとしてもわずかの影響しか与えないアミノ酸の交換を作成することに関するさらなる情報は、B o w i e e t a l ., S c i e n c e, 1990, 247, 1306-1310に見いだすことができる (図面および表を含めその全体を本明細書の一部としてここに引用する)。すべての場合において、すべての順列が本明細書の開示によりカバーされることが意図される。

【0022】

本発明のプロテアーゼペプチドのアミノ酸配列は、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列、または対応する全長アミノ酸配列、またはそのフラグメントを有する配列に実質的に類似するであろう。

【0023】

配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載される配列からなる群より選択される配列と実質的に類似する配列は、好ましくは配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115

、配列番号116、配列番号117および配列番号118からなる群より選択される配列に少なくとも50%、60%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%または99%の同一性を有するであろう。好ましくは、プロテアーゼポリペプチドは、上述の配列の1つに少なくとも約90%、91%、92%、93%、94%、95%、96%、97%、98%または99%の同一性を有するであろう。

【0024】

”同一性”とは、その類似性または関係の尺度である配列の性質を意味する。同一性は、同一である残基の数を、既知の配列または既知の配列のドメイン中の残基の総数で割り、100を乗ずることにより測定する。”ギャップ”とは、アミノ酸の付加または欠失により生じたアラインメント中の空間である。すなわち、完全に同一の配列の2つのコピーは100%の同一性を有するが、より低い程度に保存され、欠失、付加または置換を含む配列はより低い程度の同一性を有するであろう。当業者は、標準的なパラメータを用いて配列の同一性を決定するためのいくつかのコンピュータプログラム、例えば、Gapped BLASTまたはPSI-BLAST (Altschul, et al. (1997) Nucleic Acids Res. 25:3389-3402), BLAST (Altschul, et al. (1990) J. Mol. Biol. 215:403-410), およびスミス-ウォーターマン (Smith-Waterman) (Smith, et al. (1981) J. Mol. Biol. 147:195-197) が利用可能であることを認識するであろう。好ましくは、これらのプログラムのデフォルト設定を用いるが、当業者は、これらの設定を変更することが必要であるか否かを認識しており、どのようにして変更するかを理解している。

【0025】

”類似性”は、同一の残基の数と保存的に置換された残基の数 (Bowie, et al. Science, 1999, 247, 1306-1310を参照 (図面および表を含め、その全体を本明細書の一部としてここに引用する) との合計を残基とギャップの総数で割り、100を乗ずることにより測定することができる。

【0026】

好ましい態様においては、本発明は、以下のヌクレオチド配列を含むプロテアーゼポリペプチドをコードする、単離された、濃縮されたまたは精製された核酸分子を特徴とする：

(a) 配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列およびその生物学的ドメインからなる群より選択されるアミノ酸配列を有するポリペプチドをコードする；

(b) (a) のヌクレオチド配列の相補体である；

(c) (a) のヌクレオチド分子に高度にストリンジェントな条件下でハイブリダイズし、かつ天然に生ずるプロテアーゼポリペプチドをコードする。

【0027】

好ましい態様においては、本発明は、配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、

配列番号 17, 配列番号 18, 配列番号 19, 配列番号 20, 配列番号 21, 配列番号 22, 配列番号 23, 配列番号 24, 配列番号 25, 配列番号 26, 配列番号 27, 配列番号 28, 配列番号 29, 配列番号 30, 配列番号 31, 配列番号 32, 配列番号 33, 配列番号 34, 配列番号 35, 配列番号 36, 配列番号 37, 配列番号 38, 配列番号 39, 配列番号 40, 配列番号 41, 配列番号 42, 配列番号 43, 配列番号 44, 配列番号 45, 配列番号 46, 配列番号 47, 配列番号 48, 配列番号 49, 配列番号 50, 配列番号 51, 配列番号 52, 配列番号 53, 配列番号 54, 配列番号 55, 配列番号 56, 配列番号 57, 配列番号 58, および配列番号 59 からなる群より選択される配列と実質的に同一のヌクレオチド配列を含む, 単離された, 濃縮された, または精製された核酸分子を特徴とする。好ましくは, 配列は, 上述の配列と少なくとも 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% または 100% の同一性を有する。

10

【0028】

”相補体”との用語は, 互いに多くの望ましい相互作用を形成しうる 2 つのヌクレオチドを表す。例えば, アデニンはチミンと 2 つの水素結合を形成することができるため, チミンに相補的である。同様に, グアニンとシトシンは 3 つの水素結合を形成することができるため, 相補的である。あるヌクレオチド配列は, 第 1 の配列のすべてのヌクレオチドが第 2 の配列のすべてのヌクレオチドと相補的である場合, 他のヌクレオチド配列の相補体である。

【0029】

所望の特異性および選択性に応じて, 種々の低いまたは高いストリンジェンシーのハイブリダイゼーション条件を用いることができる。これらの条件は, 当業者にはよく知られている。ストリンジェントなハイブリダイゼーション条件下では, 高度に相補的な核酸配列のみがハイブリダイズする。好ましくは, そのような条件は, 20 個の連続するヌクレオチド中に 1 または 2 個より多いミスマッチを有する核酸のハイブリダイゼーションを防止し, より好ましくは, そのような条件は, 50 個の連続するヌクレオチド中に 1 または 2 個より多いミスマッチを有する核酸のハイブリダイゼーションを防止し, 最も好ましくは, そのような条件は, 100 個の連続するヌクレオチド中に 1 または 2 個より多いミスマッチを有する核酸のハイブリダイゼーションを防止する。場合によっては, この条件は, 全長配列中に 5 個のミスマッチを有する核酸のハイブリダイゼーションを防止する。

20

30

【0030】

ストリンジェントなハイブリダイゼーションアッセイ条件とは, 少なくとも以下の程度にストリンジェントなハイブリダイゼーションアッセイ条件を意味する: 50% ホルムアミド, 5X SSC, 50mM NaH_2PO_4 , pH 6.8, 0.5% SDS, 0.1mg/mL 超音波処理サケ精子 DNA, および 5X デンハルト溶液中で 42℃ で一夜のハイブリダイゼーション; 2X SSC, 0.1% SDS で 45℃ での洗浄; および 0.2X SSC, 0.1% SDS で 45℃ での洗浄。いくつかの最もストリンジェントなハイブリダイゼーションアッセイ条件においては, 2 回目の洗浄は, 0.1X SSC で 70℃ までの温度で行うことができる (Berger et al (1987) Guide to Molecular Cloning Techniques, pg 421 (図面および表を含めその全体を本明細書の一部としてここに引用する))。しかし, 他の用途は, これらの条件の組の間に入る条件の使用を必要とするかもしれない。所望のハイブリダイゼーションを達成するのに必要な条件を決定する方法は当業者にはよく知られており, いくつかの因子, 例えば, 限定されないが, ハイブリダイズすべき配列および試験すべき試料に基づく。低いストリンジェンシーの洗浄条件は, しばしば洗浄工程の間により低い温度, 例えば, 65℃, 60℃, 55℃, 50℃, または 42℃ を用いる。

40

【0031】

”活性”との用語は, ポリペプチドがペプチド結合を加水分解することを意味する。

【0032】

本明細書において用いる場合, ”触媒活性”との用語は, プロテアーゼ触媒ドメインが基

50

質を切断する速度を規定する。触媒活性は、例えば、切断される基質の量を時間の関数として決定することにより測定することができる。触媒活性は、時間を一定にして定められた時間の後に切断された基質の濃度を決定することにより、本発明の方法により測定することができる。基質の切断は、プロテアーゼの活性部位において生ずる。活性部位は、通常は、基質がプロテアーゼに結合し、切断される空洞である。

【0033】

”生物学的ドメイン”との用語は、プロテアーゼポリペプチドの、触媒活性を有するかまたはプロテアーゼの基質に結合するドメインまたは領域を意味する。

【0034】

本明細書において用いる場合，“基質”との用語は、本発明のプロテアーゼにより切断されるポリペプチドまたは蛋白質を表す。 ”切断される”との用語は、ポリペプチドまたは蛋白質の骨格のアミノ酸残基間の共有結合の切断を表す。 10

【0035】

本明細書において用いる場合，“挿入物”との用語は、密接なホモログには存在しないプロテアーゼの一部を表す。挿入物は、エクソンの選択的スプライシングの産物であるかもしれないしそうではないかもしれない。挿入物は、蛋白質配列の非重複蛋白質データベースに対するスミスーウォーターマン配列アラインメントを用いて、またはDNAStarプログラムMegalignを用いる相同配列の多重配列アラインメントにより、同定することができる（好ましくは、このプログラムのデフォルトの設定を用いるが、当業者は、これらの設定を変更することが必要であるか否かを認識しており、どのようにして変更するかを理解している）。挿入物は、蛋白質－蛋白質相互作用のための新規なインターフェースを提示することにより、またはそのような相互作用を妨害することにより、機能的役割を果たすかもしれない。 20

【0036】

別の好ましい態様においては、本発明は、プロテアーゼポリペプチドをコードし、ヌクレオチド配列に動作可能なように連結されたベクターまたはプロモーターをさらに含む、単離された、濃縮されたまたは精製された核酸分子を特徴とする。本発明はまた、組換え核酸を特徴とし、これは好ましくは細胞または生物内にある。組換え核酸は、配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、配列番号21、配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、配列番号28、配列番号29、配列番号30、配列番号31、配列番号32、配列番号33、配列番号34、配列番号35、配列番号36、配列番号37、配列番号38、配列番号39、配列番号40、配列番号41、配列番号42、配列番号43、配列番号44、配列番号45、配列番号46、配列番号47、配列番号48、配列番号49、配列番号50、配列番号51、配列番号52、配列番号53、配列番号54、配列番号55、配列番号56、配列番号57、配列番号58、および配列番号59に記載される配列からなる群より選択される配列、またはその機能的誘導体、およびヌクレオチド配列に動作可能なように連結されたベクターまたはプロモーターを含むことができる。あるいは、組換え核酸は、細胞において機能的な転写開始領域、プロテアーゼポリペプチドをコードするRNA配列に相補的な配列および細胞において機能的な転写終止領域を含んでいてもよい。特定のベクターおよび宿主細胞の組み合わせは本明細書において議論される。 30 40

【0037】

”ベクター”との用語は、細胞にトランスフェクトすることができ、細胞ゲノム中でまたはそれとは独立に複製しうる一本鎖または二本鎖の環状核酸分子を表す。環状二本鎖核酸分子は、制限酵素で処理することにより切断し、したがって直鎖状にすることができる。核酸ベクターの分類、制限酵素、および制限酵素により切断されるヌクレオチド配列の知識は、当業者には容易に入手可能である。プロテアーゼをコードする核酸分子は、ベクタ 50

ーを制限酵素で切断し、2つの断片を一緒にライゲーションすることにより、ベクター中に挿入することができる。

【0038】

動作可能な連結とは、制御DNA配列および発現が求められているDNA配列が、遺伝子配列の発現を可能とするように接続されている連結である。遺伝子配列の発現に必要な制御領域の詳細な性質は、生物により様々であろうが、一般に、プロモーター領域（原核生物においては、プロモーター（RNA転写の開始を指示する）ならびにRNAに転写されたときに合成の開始を合図するDNA配列）の両方を含む）を含むはずである。

【0039】

”トランスフェクトする”との用語は、核酸ベクターまたは他の核酸分子を細胞性生物中に挿入する多数の方法を規定する。これらの方法には、種々の手法が含まれ、例えば、細胞を高濃度の塩、電界、界面活性剤、またはDMSOで処理することにより、細胞の外膜または壁を目的の核酸分子に対して透過性にする、または種々のウイルス伝達戦略を用いることが含まれる。

【0040】

本明細書において用いる場合、”プロモーター”との用語は、遺伝子配列の発現に必要な核酸配列を表す。プロモーター領域は生物によって様々であるが、種々の生物について当業者によく知られている。例えば、原核生物においては、プロモーター領域は、プロモーター（RNA転写の開始を指示する）ならびに、RNAに転写されたときに合成の開始を合図するDNA配列の両方を含む。そのような領域は、通常は転写および翻訳の開始に参与する5'非コーディング配列、例えばTATAボックス、キャッピング配列、CAAT配列等を含む。

【0041】

好ましい態様においては、単離された核酸は、配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、配列番号21、配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、配列番号28、配列番号29、配列番号30、配列番号31、配列番号32、配列番号33、配列番号34、配列番号35、配列番号36、配列番号37、配列番号38、配列番号39、配列番号40、配列番号41、配列番号42、配列番号43、配列番号44、配列番号45、配列番号46、配列番号47、配列番号48、配列番号49、配列番号50、配列番号51、配列番号52、配列番号53、配列番号54、配列番号55、配列番号56、配列番号57、配列番号58、および配列番号59に記載される配列からなる群より選択される核酸配列を含むか、本質的にそれからなるか、それからなり、かつ、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列、その機能的誘導体、または配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、

配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 に記載されるアミノ酸配列からなる群より選択される少なくとも 35, 40, 45, 50, 60, 75, 100, 200, または 300 個の連続するアミノ酸をコードする。核酸は, cDNA クローニングにより, またはサブトラクティブハイブリダイゼーションにより, 天然の起源から単離することができる。天然の起源は哺乳動物であることができ, 好ましくはヒト, 好ましくは血液, 精液または組織であり, 核酸はトリエステル法によりまたは自動化 DNA 合成機を用いることにより合成してもよい。

10

【0042】

”哺乳動物”とは, 好ましくはマウス, ラット, ウサギ, モルモット, ヒツジ, およびヤギ等の生物を表し, より好ましくはネコ, イヌ, 有尾サル, および無尾サルを表し, 最も好ましくはヒトを表す。

【0043】

さらに別の好ましい態様においては, 核酸は, 例えば, 追加のポリペプチドの同定およびクローニングを容易にするためのハイブリダイゼーションプローブを設計するのに, 追加のポリペプチドのクローニングを容易にするための PCR プローブを設計するのに, ポリペプチド領域に対する抗体を得るために, およびアンチセンスオリゴヌクレオチドを設計するのに有用な, 保存されたまたは独特の領域である。

20

【0044】

”保存された核酸領域”とは, プロテアーゼポリペプチドをコードする 2 つまたはそれ以上の核酸に存在する領域を意味し, 特定の核酸配列は低いストリンジェンシー条件下でこの領域にハイブリダイズすることができる。プロテアーゼポリペプチドをコードする核酸のスクリーニングに適した低ストリンジェンシー条件の例は, Wahl et al. Meth. Enzym. 152: 399-407 (1987) および Wahl et al. Meth. Enzym. 152: 415-423 (1987) (図面および表を含めその全体を本明細書の一部としてここに引用する) に提供される。好ましくは, 保存領域は, 20 ヌクレオチド中 5 個以下で異なり, より好ましくは 20 ヌクレオチド中 2 個, 最も好ましくは 20 ヌクレオチド中 1 個が異なる。

30

【0045】

”独特の核酸領域”とは, プロテアーゼポリペプチドをコードする核酸中に存在し, 天然に生ずる任意の他のポリペプチドをコードする配列中には存在しない配列を意味する。そのような領域は, 好ましくは, 配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列の, 32 個 (好ましくは 40 個, より好ましくは 45 個, 最も好ましくは 55 個) またはそれ以

40

50

上の連続するアミノ酸をコードする。特に、独特の核酸領域は、好ましくは哺乳動物起源のものである。

【0046】

好ましい態様においては、核酸プローブは、配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、配列番号21、配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、配列番号28、配列番号29、配列番号30、配列番号31、配列番号32、配列番号33、配列番号34、配列番号35、配列番号36、配列番号37、配列番号38、配列番号39、配列番号40、配列番号41、配列番号42、配列番号43、配列番号44、配列番号45、配列番号46、配列番号47、配列番号48、配列番号49、配列番号50、配列番号51、配列番号52、配列番号53、配列番号54、配列番号55、配列番号56、配列番号57、配列番号58、および配列番号59からなる群より選択される全長配列の、少なくとも12、32、75、90、105、120、150、200、250、300または350個の連続するアミノ酸をコードする核酸またはその機能的誘導体にハイブリダイズする。

【0047】

プローブを使用する方法には、ハイブリダイゼーションが生ずるような条件下で試料を核酸プローブと接触させ、プロテアーゼRNAに結合したプローブの存在または量を検出することにより、試料中のプロテアーゼRNAの存在または量を検出することが含まれる。プローブとプロテアーゼポリペプチドをコードする核酸配列との間に形成される核酸デュプレックスを、検出された核酸の配列の同定において用いることができる (Nelson et al., Nonisotopic DNA Probe Techniques, Academic Press, San Diego, Kricka, ed., p. 275, 1992 (図面および表を含めその全体を本明細書の一部としてここに引用する))。そのような方法を実施するためのキットは、その中に核酸プローブが置かれている容器手段を含むように構築することができる。

【0048】

プローブを使用する方法には、これらのプローブを用いて、例えば当該技術分野において知られる技術を用いて予測プロテアーゼのそれぞれの全長クローンを見いだすことも含まれる。これらのクローンは、コードされるプロテアーゼの触媒的活性を阻害し、癌、免疫関連疾病および疾患、心臓血管疾病、脳またはニューロン関連疾病、および代謝性疾患の治療において潜在的有用性を有する小分子化合物をスクリーニングするのに有用であろう。より詳細には、疾患には、組織、血液、または造血細胞起源の癌、特に乳、結腸、肺、前立腺、子宮頸部、脳、卵巣、膀胱、または腎臓が関与する癌；中枢神経または末梢神経系疾病および状態、例えば、片頭痛、痛み、性的機能不全、気分障害、注意障害、認識障害、低血圧症、および高血圧症；精神病性および神経学的疾患、例えば不安、精神分裂病、躁うつ病、せん妄、痴呆、重症の精神遅滞および運動異常症、例えばハンチントン病またはツレット症候群；神経変性性疾患、例えばアルツハイマー病、パーキンソン病、多発性硬化症、および筋萎縮性側索硬化症；HIV1、HIV-2または他のウイルスまたはプリオン体または真菌または細菌生物により引き起こされるウイルスまたは非ウイルス感染；代謝性疾患、例えば、特に糖尿病および肥満およびこれに関連する症候群；心臓血管疾患、例えば再灌流再狭窄、冠状動脈血栓症、凝固疾患、制御されない細胞成長疾患、アテローム性動脈硬化症；眼性疾患、例えば緑内障、網膜症、および黄斑変性；炎症性疾患、例えば慢性関節リウマチ、慢性炎症性腸疾病、慢性炎症性骨盤疾病、多発性硬化症、ぜん息、変形性関節症、乾癬、アテローム性動脈硬化症、鼻炎、自己免疫、および臓器移植拒絶が含まれる。

【0049】

別の観点においては、本発明は、配列番号60、配列番号61、配列番号62、配列番号

63, 配列番号64, 配列番号65, 配列番号66, 配列番号67, 配列番号68, 配列番号69, 配列番号70, 配列番号71, 配列番号72, 配列番号73, 配列番号74, 配列番号75, 配列番号76, 配列番号77, 配列番号78, 配列番号79, 配列番号80, 配列番号81, 配列番号82, 配列番号83, 配列番号84, 配列番号85, 配列番号86, 配列番号87, 配列番号88, 配列番号89, 配列番号90, 配列番号91, 配列番号92, 配列番号93, 配列番号94, 配列番号95, 配列番号96, 配列番号97, 配列番号98, 配列番号99, 配列番号100, 配列番号101, 配列番号102, 配列番号103, 配列番号104, 配列番号105, 配列番号106, 配列番号107, 配列番号108, 配列番号109, 配列番号110, 配列番号111, 配列番号112, 配列番号113, 配列番号114, 配列番号115, 配列番号116, 配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を有するプロテアーゼポリペプチドをコードする核酸分子を含む組換え細胞または組織を記述する。そのような細胞においては、核酸は遺伝的制御要素の制御下にあってもよく、または外来性プロモーターを含む外来性制御要素の制御下にあってもよい。”外来性”とは、通常はプロテアーゼポリペプチドのコーディング配列とインビボで転写的にカップリングしていないプロモーターを意味する。

10

【0050】

ポリペプチドは、好ましくは、配列番号60, 配列番号61, 配列番号62, 配列番号63, 配列番号64, 配列番号65, 配列番号66, 配列番号67, 配列番号68, 配列番号69, 配列番号70, 配列番号71, 配列番号72, 配列番号73, 配列番号74, 配列番号75, 配列番号76, 配列番号77, 配列番号78, 配列番号79, 配列番号80, 配列番号81, 配列番号82, 配列番号83, 配列番号84, 配列番号85, 配列番号86, 配列番号87, 配列番号88, 配列番号89, 配列番号90, 配列番号91, 配列番号92, 配列番号93, 配列番号94, 配列番号95, 配列番号96, 配列番号97, 配列番号98, 配列番号99, 配列番号100, 配列番号101, 配列番号102, 配列番号103, 配列番号104, 配列番号105, 配列番号106, 配列番号107, 配列番号108, 配列番号109, 配列番号110, 配列番号111, 配列番号112, 配列番号113, 配列番号114, 配列番号115, 配列番号116, 配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択される全長アミノ酸配列によりコードされる蛋白質のフラグメントである。”フラグメント”とは、プロテアーゼポリペプチド中に存在するアミノ酸配列を意味する。好ましくは、そのような配列は、配列番号60, 配列番号61, 配列番号62, 配列番号63, 配列番号64, 配列番号65, 配列番号66, 配列番号67, 配列番号68, 配列番号69, 配列番号70, 配列番号71, 配列番号72, 配列番号73, 配列番号74, 配列番号75, 配列番号76, 配列番号77, 配列番号78, 配列番号79, 配列番号80, 配列番号81, 配列番号82, 配列番号83, 配列番号84, 配列番号85, 配列番号86, 配列番号87, 配列番号88, 配列番号89, 配列番号90, 配列番号91, 配列番号92, 配列番号93, 配列番号94, 配列番号95, 配列番号96, 配列番号97, 配列番号98, 配列番号99, 配列番号100, 配列番号101, 配列番号102, 配列番号103, 配列番号104, 配列番号105, 配列番号106, 配列番号107, 配列番号108, 配列番号109, 配列番号110, 配列番号111, 配列番号112, 配列番号113, 配列番号114, 配列番号115, 配列番号116, 配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択される全長配列の、少なくとも32, 45, 50, 60, 100, 200, または300個の連続するアミノ酸を含む。

20

30

40

【0051】

別の観点においては、本発明は、配列番号60, 配列番号61, 配列番号62, 配列番号63, 配列番号64, 配列番号65, 配列番号66, 配列番号67, 配列番号68, 配列番号69, 配列番号70, 配列番号71, 配列番号72, 配列番号73, 配列番号74, 配列番号75, 配列番号76, 配列番号77, 配列番号78, 配列番号79, 配列番号80, 配列番号81, 配列番号82, 配列番号83, 配列番号84, 配列番号85, 配列番号

50

号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 に記載されるアミノ酸配列およびその生物学的ドメインからなる群より選択されるアミノ酸配列と実質的に同一の配列を有する, 単離された, 濃縮されたまたは精製されたプロテアーゼポリペプチドを特徴とする。好ましくは, ポリペプチド配列は上述の配列と少なくとも 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% または 100% の同一性を有する。 10

【0052】

ポリペプチドに関して”単離された”とは, 互いに結合した 6 個 (好ましくは 12 個, より好ましくは 18 個, 最も好ましくは 25, 32, 40, または 50 個) またはそれ以上のアミノ酸のポリマーを意味し, 天然起源から単離されたポリペプチドまたは合成されたポリペプチドが含まれる。ある種の観点においては, より長いポリペプチド, 例えば, 配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 に記載されるアミノ酸配列からなる群より選択される全長配列の, 100, 200, 300, 400, 450, 500, 550, 600, 700, 800, 900 個またはそれ以上の連続するアミノ酸を有するポリペプチド, および/または, 配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 からなる群より選択される配列と, 少なくとも 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% または 99% の同一性を有するポリペプチドが好ましい。 30 40

【0053】

本発明の単離されたポリペプチドは, 天然には純粋なまたは分離された形で見いだされない点において独特である。”単離された”との用語の使用は, 天然に生ずる配列がその正常な細胞性環境から除かれていることを示す。すなわち, 配列は, 無細胞溶液中にあってよく, 異なる細胞性環境に置かれていてもよい。この用語は, その配列が存在する唯一のアミノ酸鎖であることを意味するものではなく, 配列が天然にこれに付随する非アミノ 50

酸物質を本質的に含まない（少なくとも約 90% 純粋，より好ましくは少なくとも約 95% またはそれ以上純粋）ことを意味する。

【0054】

ポリペプチドに関して使用する場合，“濃縮された”との用語は，特定のアミノ酸配列が，正常または疾病細胞におけるより，または配列が由来する細胞におけるより，目的とする細胞または溶液中に存在する総アミノ酸配列の有意に高い割合（2－5 倍）を占めることを意味する。これは，存在する他のアミノ酸配列の量の優先的減少により，または目的とする特定のアミノ酸配列の量の優先的増加により，またはこれら 2 つの組み合わせにより，人が生じさせることができる。しかし，濃縮されたとは，他のアミノ酸配列が存在しないことを意味するものではなく，単に目的とする配列の相対的な量が有意に増加していることを意味することに注意すべきである。本明細書において“有意”にとの用語は，増加のレベルがそのような増加を作成した人にとって有用であることを示し，一般に，他のアミノ酸配列と比較して少なくとも約 2 倍，より好ましくは少なくとも 5－10 倍，またはそれより多い増加を意味する。この用語はまた，他の起源からのアミノ酸配列が存在しないことを意味するものではない。アミノ酸配列の他の起源は，例えば，酵母または細菌のゲノム，またはクローニングベクター，例えば pUC19 によりコードされるアミノ酸配列を含みうる。この用語は，人が介在して所望のアミノ酸配列の比率を増加させる状況のみをカバーすることを意味する。

10

【0055】

ある目的のためには，アミノ酸配列が精製された形であることも有利である。ポリペプチドに関して“精製された”との用語は絶対的純度（例えば均一調製物）を要求するものではなく，この用語は配列が天然の環境におけるより比較的純粋であることを示す。天然のレベルと比較して，このレベルは少なくとも 2－5 倍高くあるべきである（例えば mg/mL で）。少なくとも 1 桁の精製，好ましくは 2 または 3 桁，より好ましくは 4 または 5 桁の精製が明示的に意図される。物質は，好ましくは機能的に有意なレベルで夾雑物を含まず，例えば，90%，95%，または 99% 純粋である。

20

【0056】

好ましい態様においては，プロテアーゼポリペプチドは，配列番号 60，配列番号 61，配列番号 62，配列番号 63，配列番号 64，配列番号 65，配列番号 66，配列番号 67，配列番号 68，配列番号 69，配列番号 70，配列番号 71，配列番号 72，配列番号 73，配列番号 74，配列番号 75，配列番号 76，配列番号 77，配列番号 78，配列番号 79，配列番号 80，配列番号 81，配列番号 82，配列番号 83，配列番号 84，配列番号 85，配列番号 86，配列番号 87，配列番号 88，配列番号 89，配列番号 90，配列番号 91，配列番号 92，配列番号 93，配列番号 94，配列番号 95，配列番号 96，配列番号 97，配列番号 98，配列番号 99，配列番号 100，配列番号 101，配列番号 102，配列番号 103，配列番号 104，配列番号 105，配列番号 106，配列番号 107，配列番号 108，配列番号 109，配列番号 110，配列番号 111，配列番号 112，配列番号 113，配列番号 114，配列番号 115，配列番号 116，配列番号 117 および配列番号 118 に記載されるアミノ酸配列からなる群より選択される全長アミノ酸配列によりコードされる蛋白質のフラグメントである。好ましくは，プロテアーゼポリペプチドは，配列番号 60，配列番号 61，配列番号 62，配列番号 63，配列番号 64，配列番号 65，配列番号 66，配列番号 67，配列番号 68，配列番号 69，配列番号 70，配列番号 71，配列番号 72，配列番号 73，配列番号 74，配列番号 75，配列番号 76，配列番号 77，配列番号 78，配列番号 79，配列番号 80，配列番号 81，配列番号 82，配列番号 83，配列番号 84，配列番号 85，配列番号 86，配列番号 87，配列番号 88，配列番号 89，配列番号 90，配列番号 91，配列番号 92，配列番号 93，配列番号 94，配列番号 95，配列番号 96，配列番号 97，配列番号 98，配列番号 99，配列番号 100，配列番号 101，配列番号 102，配列番号 103，配列番号 104，配列番号 105，配列番号 106，配列番号 107，配列番号 108，配列番号 109，配列番号 110，配列番号 111，配列番号 112，配列

30

40

50

番号 1 1 3, 配列番号 1 1 4, 配列番号 1 1 5, 配列番号 1 1 6, 配列番号 1 1 7 および配列番号 1 1 8 に記載されるアミノ酸配列, またはその機能的誘導体からなる群より選択される全長配列の, 少なくとも 3 2, 4 5, 5 0, 6 0, 1 0 0, 2 0 0, または 3 0 0 個の連続するアミノ酸を含む。

【0057】

好ましい態様においては, プロテアーゼポリペプチドは, 配列番号 6 0, 配列番号 6 1, 配列番号 6 2, 配列番号 6 3, 配列番号 6 4, 配列番号 6 5, 配列番号 6 6, 配列番号 6 7, 配列番号 6 8, 配列番号 6 9, 配列番号 7 0, 配列番号 7 1, 配列番号 7 2, 配列番号 7 3, 配列番号 7 4, 配列番号 7 5, 配列番号 7 6, 配列番号 7 7, 配列番号 7 8, 配列番号 7 9, 配列番号 8 0, 配列番号 8 1, 配列番号 8 2, 配列番号 8 3, 配列番号 8 4 10, 配列番号 8 5, 配列番号 8 6, 配列番号 8 7, 配列番号 8 8, 配列番号 8 9, 配列番号 9 0, 配列番号 9 1, 配列番号 9 2, 配列番号 9 3, 配列番号 9 4, 配列番号 9 5, 配列番号 9 6, 配列番号 9 7, 配列番号 9 8, 配列番号 9 9, 配列番号 1 0 0, 配列番号 1 0 1, 配列番号 1 0 2, 配列番号 1 0 3, 配列番号 1 0 4, 配列番号 1 0 5, 配列番号 1 0 6, 配列番号 1 0 7, 配列番号 1 0 8, 配列番号 1 0 9, 配列番号 1 1 0, 配列番号 1 1 1, 配列番号 1 1 2, 配列番号 1 1 3, 配列番号 1 1 4, 配列番号 1 1 5, 配列番号 1 1 6, 配列番号 1 1 7 および配列番号 1 1 8 に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を有するアミノ酸配列を含む。

【0058】

ポリペプチドは, 当該技術分野においてよく知られる方法により, 天然の起源から単離することができる。天然の起源は哺乳動物であることができ, 好ましくはヒトであり, 好ましくは, 血液, 精液, または組織であり, またはポリペプチドは自動化ポリペプチド合成機を用いて合成してもよい。 20

【0059】

ある態様においては, 本発明は, (a) 配列番号 6 0, 配列番号 6 1, 配列番号 6 2, 配列番号 6 3, 配列番号 6 4, 配列番号 6 5, 配列番号 6 6, 配列番号 6 7, 配列番号 6 8, 配列番号 6 9, 配列番号 7 0, 配列番号 7 1, 配列番号 7 2, 配列番号 7 3, 配列番号 7 4, 配列番号 7 5, 配列番号 7 6, 配列番号 7 7, 配列番号 7 8, 配列番号 7 9, 配列番号 8 0, 配列番号 8 1, 配列番号 8 2, 配列番号 8 3, 配列番号 8 4, 配列番号 8 5, 配列番号 8 6, 配列番号 8 7, 配列番号 8 8, 配列番号 8 9, 配列番号 9 0, 配列番号 9 1, 配列番号 9 2, 配列番号 9 3, 配列番号 9 4, 配列番号 9 5, 配列番号 9 6, 配列番号 9 7, 配列番号 9 8, 配列番号 9 9, 配列番号 1 0 0, 配列番号 1 0 1, 配列番号 1 0 2, 配列番号 1 0 3, 配列番号 1 0 4, 配列番号 1 0 5, 配列番号 1 0 6, 配列番号 1 0 7, 配列番号 1 0 8, 配列番号 1 0 9, 配列番号 1 1 0, 配列番号 1 1 1, 配列番号 1 1 2, 配列番号 1 1 3, 配列番号 1 1 4, 配列番号 1 1 5, 配列番号 1 1 6, 配列番号 1 1 7 および配列番号 1 1 8 に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を有する組換えプロテアーゼポリペプチドを含む。”組換えプロテアーゼポリペプチド”とは, その存在位置(例えば, 天然に見いだされる物とは異なる細胞または組織に存在), 純度または構造において天然に生ずるポリペプチドと区別されるように, 組換え DNA 技術により製造されるポリペプチドを意味する。一般に, そのような組換えポリペプチドは, 天然に通常観察される量とは異なる量で細胞中に存在するであろう。 30 40

【0060】

宿主細胞中で発現させるべきポリペプチドは, 異種蛋白質からの領域を含む融合蛋白質であってもよい。そのような領域を含むことにより, 例えば, 分泌させ, 安定性を改良し, またはポリペプチドの精製を容易にすることができる。例えば, 適当なシグナルペプチドをコードする配列を発現ベクター中に組み込むことができる。シグナルペプチド(分泌リーダー)の DNA 配列を, ポリペプチドがシグナルペプチドを含む融合蛋白質として翻訳されるように, ポリヌクレオチド配列にインフレームで融合させることができる。意図する宿主細胞において機能的であるシグナルペプチドは, ポリペプチドの細胞外分泌を促進する。好ましくは, シグナル配列はポリペプチドが細胞から分泌される際にポリペプチド 50

から切断される。すなわち，プロテアーゼポリペプチドのN末端が運搬ペプチドに融合されている好ましい融合蛋白質を作成することができる。

【0061】

1つの態様においては，ポリペプチドは，ポリペプチドの精製を容易にするために用いられる異種領域を含む融合蛋白質を含む。そのような機能のために用いられる入手可能なペプチドの多くは，融合蛋白質が結合パートナーに選択的に結合することを可能とする。好ましい結合パートナーには，プロテインAのIgG結合ドメインの1またはそれ以上が含まれ，融合蛋白質は，IgG結合セファロース等のアフィニティークロマトグラフィーにより容易に均一にまで精製される。あるいは，多くのベクターは，標的蛋白質のN末端またはC末端で発現されることができるヒスチジン残基のストレッチを有するという利点を有しており，したがって，金属キレート化クロマトグラフィーにより目的とする蛋白質を回収することができる。蛋白質加水分解酵素，例えばエンテロキナーゼ，ファクターXプロコラゲナーゼまたはトロンビンの認識部位をコードするヌクレオチド配列をプロテアーゼポリペプチドの配列のすぐ上流に配置すると，融合蛋白質を切断して成熟プロテアーゼポリペプチドを得ることができる。融合蛋白質結合パートナーのさらなる例には，限定されないが，酵母I-因子，sf9昆虫細胞におけるミツバチメラチンリーダー，6-Hisタグ，チオレドキシントグ，ヘマグルチニンタグ，GSTタグ，およびOmpAシグナル配列タグが含まれる。当業者には理解されるように，ペプチドを認識しこれに結合する結合パートナーは，任意のイオン，分子または化合物であることができ，例えば金属イオン（例えば金属アフィニティークラム），抗体，またはそのフラグメント，およびペプチドに結合する任意の蛋白質またはペプチド，例えばFLAGタグが含まれる。

【0062】

抗体

別の観点においては，本発明は，プロテアーゼポリペプチドまたはプロテアーゼポリペプチドドメインまたはフラグメントに対して特異的結合親和性を有する抗体（例えば，モノクローナル抗体またはポリクローナル抗体）を特徴とし，ここで，ポリペプチドは，配列番号60，配列番号61，配列番号62，配列番号63，配列番号64，配列番号65，配列番号66，配列番号67，配列番号68，配列番号69，配列番号70，配列番号71，配列番号72，配列番号73，配列番号74，配列番号75，配列番号76，配列番号77，配列番号78，配列番号79，配列番号80，配列番号81，配列番号82，配列番号83，配列番号84，配列番号85，配列番号86，配列番号87，配列番号88，配列番号89，配列番号90，配列番号91，配列番号92，配列番号93，配列番号94，配列番号95，配列番号96，配列番号97，配列番号98，配列番号99，配列番号100，配列番号101，配列番号102，配列番号103，配列番号104，配列番号105，配列番号106，配列番号107，配列番号108，配列番号109，配列番号110，配列番号111，配列番号112，配列番号113，配列番号114，配列番号115，配列番号116，配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列と少なくとも約90%同一の配列を有する群から選択される。好ましくは，ポリペプチドは，上述の配列と少なくとも約90%，91%，92%，93%，94%，95%，96%，97%，98%，99%または100%の同一性を有する。”特異的結合親和性”とは，抗体が，特定の条件下において，他のポリペプチドに結合するより高い親和性をもって標的プロテアーゼポリペプチドに結合することを意味する。抗体または抗体フラグメントは，他のポリペプチドに結合しうる領域を含むポリペプチドである。”特異的結合親和性”との用語は，特定の条件下で，他のポリペプチドに結合するより高い親和性をもってプロテアーゼポリペプチドに結合する抗体を記述する。抗体を用いて，プロテアーゼポリペプチドの内因性起源を同定して，細胞サイクル制御をモニターすることができ，または，プロテアーゼポリペプチドの細胞中の免疫局在化に用いることができる。

【0063】

”ポリクローナル”との用語は，抗原またはその抗原性機能的誘導体で免疫した動物の血

清から誘導される抗体分子の異成分集団である抗体を表す。ポリクローナル抗体の製造のためには、種々の宿主動物に抗原を注射することにより免疫することができる。宿主の種類により、種々のアジュバントを用いて免疫学的応答を増加させることができる。

【0064】

”モノクローナル抗体”は、特定の抗原に対する抗体の実質的に均一な集団である。モノクローナル抗体は、培養連続細胞株による抗体分子の生成を与える任意の技術により得ることができる。モノクローナル抗体は、当業者に知られる方法により得ることができる（Kohler et al. Nature 256: 495-497, 1975, および米国特許4, 376, 110（これらの両方は、図面または表を含めその全体を本明細書の一部としてここに引用する））。

10

【0065】

本発明の抗体には、”ヒト化”モノクローナル抗体およびポリクローナル抗体が含まれる。ヒト化抗体は、抗体の非ヒト（典型的にはネズミ）相補性決定領域が非ヒト（例えばネズミ）免疫グロブリンの重および軽可変鎖からヒト可変ドメイン中に移され、次にネズミ対応物のフレームワーク領域中のいくつかのヒト残基が置き換えられている組換え蛋白質である。本発明にしたがうヒト化抗体は、治療方法において用いるのに適している。ネズミ免疫グロブリン可変ドメインをクローニングする一般的手法は、例えば刊行物（Orlandi et al., Proc. Nat'l Acad. Sci. USA 86: 3833 (1989)）に記載されている。ヒト化モノクローナル抗体を製造する手法は、例えば、Jonesら（Nature 321: 522 (1986)）、Riechmannら（Nature 332: 323 (1988)）、Verhoeyenら（Science 239: 1534 (1988)）、Carterら（Proc. Nat'l Acad. Sci. USA 89: 4285 (1992)）、Sandhu（Crit. Rev. Biotech. 12: 437 (1992)）、およびSingerら（J. Immun. 150: 2844 (1993)）に記載されている。

20

【0066】

”抗体フラグメント”との用語は、特定の分子に対して特異的結合親和性を表示する抗体の一部、しばしば超可変領域および周囲の重鎖および軽鎖の一部を表す。超可変領域は、ポリペプチド標的に物理的に結合する抗体の部分である。

【0067】

本発明の抗体フラグメントには、”一本鎖抗体”が含まれる。この記載において用いられる語句は、特異性をもって抗原と結合し、抗体の重鎖および軽鎖からの可変または超可変領域を含む線状ポリペプチドを示す。そのような一本鎖抗体は、一般的な方法論により製造することができる。FvフラグメントのVhおよびVl領域を共有結合させ、ジスルフィド結合を導入することにより安定化させる。Glockshuberら（Biochemistry 1362 (1990)）を参照。あるいは、ペプチドリンカーを挿入することによりVhおよびVl領域を結合させることができる。Vh、Vlおよびペプチドリンカー配列をコードする遺伝子は、組換え発現ベクターを用いて構築し発現させることができる。Colcherら（J. Nat'l Cancer Inst. 82: 1191 (1990)）を参照。VhおよびVl抗体鎖からの超可変領域を含むアミノ酸配列もまた、ジスルフィド結合またはペプチドリンカーを用いて構築することができる。

30

40

【0068】

本発明のプロテアーゼポリペプチドに対して特異的結合親和性を有する抗体または抗体フラグメントは、試料をプロテアーゼ-抗体免疫複合体の形成に適した条件下で抗体で探索し、プロテアーゼポリペプチドに結合した抗体の存在および／または量を検出することにより、試料中のプロテアーゼポリペプチドの存在および／または量を検出する方法において用いることができる。そのような方法を実施するための診断キットは、プロテアーゼに特異的な抗体または抗体フラグメント、ならびに抗体の結合パートナーまたは抗体それ自体のコンジュゲートを含むように構築することができる。

【0069】

50

本発明のプロテアーゼポリペプチドに対して特異的結合親和性を有する抗体または抗体フラグメントは、原核生物または真核生物から単離、濃縮、または精製することができる。当業者に知られる日常的な方法により、原核生物および真核生物のいずれにおいても、抗体または抗体フラグメントを製造することができる。ポリペプチド分子である抗体の精製、濃縮および単離は、上に記載される。

【0070】

本発明のプロテアーゼポリペプチドに対して特異的結合親和性を有する抗体は、免疫複合体が形成するような条件下で試料を抗体と接触させ、プロテアーゼポリペプチドに結合した抗体の存在および／または量を検出することにより、試料中のプロテアーゼポリペプチドの存在および／または量を検出するための方法において用いることができる。そのような方法を実施するための診断キットは、抗体を含む第1の容器および抗体の結合パートナーおよび標識（例えば放射性同位体）を含む第2の容器を含むように構築することができる。診断キットはまた、FDAに認可された使用の通知およびその指針を含んでいてもよい。

【0071】

別の観点においては、本発明は、プロテアーゼポリペプチドまたはプロテアーゼポリペプチドドメインに対して特異的結合親和性を有する抗体を産生するハイブリドーマを特徴とし、ここで、ポリペプチドは、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択される。”ハイブリドーマ”とは、抗体、例えば本発明のプロテアーゼに対する抗体を分泌しうる不死化細胞株を意味する。好ましい態様においては、プロテアーゼに対する抗体は、本発明のプロテアーゼポリペプチドに特異的に結合することができるアミノ酸の配列を含む。

【0072】

別の観点においては、本発明はまた、上述したいずれかの核酸分子によりコードされるポリペプチドに結合する抗体、および負対照抗体を含むキットに関する。

【0073】

”負対照抗体”との用語は、特異的結合親和性を有する抗体と類似の起源に由来するが、本発明のポリペプチドに対して結合親和性を示さない抗体を表す。

【0074】

別の観点においては、本発明は、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117およ

び配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を有する群から選択されるプロテアーゼポリペプチドに結合することができるプロテアーゼポリペプチド結合剤を特徴とする。結合剤は、好ましくは、本発明のプロテアーゼポリペプチド上に存在するエピトープを認識する精製された抗体である。他の結合剤には、プロテアーゼポリペプチドに結合する分子およびプロテアーゼポリペプチドに結合する類似の分子が含まれる。そのような結合剤は、プロテアーゼ結合パートナー活性を測定するアッセイを用いて同定することができるか、またはプロテアーゼ活性、例えば基質分子に結合した蛍光発生性または放射性マーカーの放出を測定するアッセイを用いて同定することができる。

【0075】

10

プロテアーゼポリペプチドを検出するためのスクリーニング方法

本発明はまた、本発明のプロテアーゼポリペプチドまたは同等の配列を含むヒト細胞をスクリーニングする方法を特徴とする。該方法は、当該技術分野において日常的かつ標準的な技術、例えば本発明のプロテアーゼの同定のために本明細書において記載される技術（例えば、クローニング、サザンまたはノザンブロット分析、インシトゥーハイブリダイゼーション、PCR増幅等）を用いて、ヒト細胞において新規ポリペプチドを同定することを含む。

【0076】

プロテアーゼ活性を調節する物質を同定するためのスクリーニング方法

別の観点においては、本発明は、プロテアーゼ活性を調節する物質を同定する方法を特徴とする。該方法は、(a) 配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択される配列と実質的に同一のアミノ酸配列を含むプロテアーゼポリペプチドを試験物質と接触させ；(b) 前記ポリペプチドの活性を測定し；そして(c) 前記物質が前記ポリペプチドの活性を調節するか否かを判定する、の各工程を含む。より好ましくは、配列は、上述の配列に、少なくとも約90%、91%、92%、93%、94%、95%、96%、97%、98%または99%の同一性を有する。

20

30

【0077】

”調節する”との用語は、化合物が本発明のプロテアーゼの機能を変化させる能力を表す。調節剤は、好ましくは、プロテアーゼに暴露される化合物の濃度に依存して本発明のプロテアーゼの活性を活性化するかまたは阻害する。

40

【0078】

”調節する”との用語はまた、プロテアーゼと天然の結合パートナーとの間に複合体が形成される確率を増加または減少させることにより、本発明のプロテアーゼの機能を変更することを表す。調節剤は、好ましくは、プロテアーゼと天然の結合パートナーとの間にそのような複合体が形成される確率を増加させ、より好ましくは、プロテアーゼに暴露される化合物の濃度に依存してプロテアーゼと天然の結合パートナーとの間に複合体が形成される確率を増加または減少させ、最も好ましくは、プロテアーゼと天然の結合パートナーとの間に複合体が形成される確率を減少させる。

【0079】

50

”活性化する”との用語は、プロテアーゼの細胞活性を増加させることを表す。阻害するとの用語は、プロテアーゼの細胞活性を減少させることを表す。

【0080】

”複合体”との用語は、互いに結合した少なくとも2つの分子の集合を表す。シグナル伝達複合体は、しばしば、互いに結合した少なくとも2つの蛋白質分子を含む。例えば、蛋白質チロシンレセプター蛋白質キナーゼ、GRB2、SOS、RAF、およびRASは、有糸分裂促進リガンドに応答して、集合してシグナル伝達複合体を形成する。同様に、血液凝固に参与するプロテアーゼとその補因子は、細胞膜上で高分子複合体を形成することが知られている。さらに、細胞外マトリクスの修飾に参与するプロテアーゼは、その阻害剤と、および細胞外マトリクスの成分と複合体を形成することが知られている。

10

【0081】

”天然の結合パートナー”との用語は、細胞中でプロテアーゼに結合するポリペプチド、脂質、小分子、または核酸を表す。プロテアーゼと天然の結合パートナーとの間の相互作用の変化は、相互作用が形成される確率の増加または減少、またはプロテアーゼ／天然の結合パートナー複合体の濃度の増加または減少として現れることができる。

【0082】

本明細書において用いる場合、”接触させる”との用語は、試験化合物を含む溶液を、本発明の方法の細胞を浸している液体培地と混合することを表す。化合物を含む溶液はまた、該方法の細胞内への試験化合物の取り込みを容易にする他の成分、例えばジメチルスルホキシド(DMSO)を含んでいてもよい。試験化合物を含む溶液は、輸送装置、例えば、ピペットに基づく装置またはシリンジに基づく装置を用いて、細胞を浸している培地に加えることができる。

20

【0083】

別の観点においては、本発明は、細胞においてプロテアーゼ活性を調節する物質を同定する方法を特徴とする。該方法は、(a)細胞においてプロテアーゼポリペプチドを発現させ、ここで、前記ポリペプチドは、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列と実質的に同一の配列を有し；(b)試験物質を前記細胞に加え；そして(c)細胞表現型、細胞増殖、細胞分化または前記ポリペプチドと天然の結合パートナーとの間の相互作用の変化をモニターする、の各工程を含む。

30

40

【0084】

本明細書において用いる場合、”発現”との用語は、細胞中でプロテアーゼ遺伝子を含む核酸ベクターから本発明のプロテアーゼが産生されることを表す。本明細書に記載されるように、核酸ベクターを、当該技術分野においてよく知られる手法を用いて細胞中にトランスフェクトする。

【0085】

本発明の別の観点は、本発明のプロテアーゼポリペプチドに結合する化合物を同定する方法に関する。該方法は、プロテアーゼポリペプチドを化合物と接触させ、化合物がプロテアーゼポリペプチドに結合するか否かを判定することを含む。結合は、当業者によく知られる結合アッセイ、例えば、限定されないが、ゲルシフトアッセイ、ウエスタンブロット

50

、放射性標識競合アッセイ、ファージに基づく発現クローニング、クロマトグラフィーによる共分画、共沈澱、架橋、相互作用トラップ／ツーハイブリッド分析、サウスウエスタン分析、ELISA等により判定することができる。このようなアッセイは、例えば、*Current Protocols in Molecular Biology*, 1999, John Wiley & Sons, NY (その全体を本明細書の一部としてここに引用する)に記載されている。スクリーニングすべき化合物には、限定されないが、細胞外、細胞内、生物学的または化学的起源の化合物が含まれる。

【0086】

本発明の方法はまた、標識、例えば放射性標識（例えば¹²⁵I, ³⁵S, ³²P, ³H）、蛍光標識、化学発光標識、酵素標識および免疫学的標識に結合した化合物を包含する。そのような試験において用いるプロテアーゼポリペプチドは、溶液中で遊離であってもよく、固体支持体に結合していてもよく、細胞表面上に存在してもよく、細胞内に存在してもよく、または細胞の一部に付随していてもよい。当業者は、例えば、プロテアーゼポリペプチドと試験化合物との間の複合体の形成を測定することができる。あるいは、当業者は、試験化合物により引き起こされる、プロテアーゼポリペプチドとその基質との間の複合体形成の減少を調べることができる。

10

【0087】

他のアッセイを用いて、酵素活性を調べることができる。これには、限定されないが、測光法、放射線法、HPLC、電気化学的方法等が含まれ、これは、例えば、*Enzyme Assays: A Practical Approach*, eds. R. E. Senthall and M. J. Danson, 1992, Oxford University Press (その全体を本明細書の一部としてここに引用する)に記載されている。

20

【0088】

本発明の別の観点では、プロテアーゼポリペプチドを化合物と接触させ、化合物がプロテアーゼポリペプチドの活性を調節するか否かを判定することを含む、プロテアーゼポリペプチドの活性を調節する（すなわち、増加または減少させる）化合物を同定する方法に関する。これらの化合物もまた、“蛋白質プロテアーゼの調節剤”と称される。試験化合物の存在下における活性を、試験化合物の非存在下における活性に対して測定する。試験化合物を含む試料の活性が試験化合物を含まない試料の活性より高い場合、その化合物は活性を増加させたのであろう。同様に、試験化合物を含む試料の活性が試験化合物を含まない試料の活性より低い場合、その化合物は活性を阻害したのであろう。

30

【0089】

本発明は、種々の薬剤スクリーニング手法の任意のものにおいてプロテアーゼポリペプチドを用いて化合物をスクリーニングするのに特に有用である。スクリーニングすべき化合物には、限定されないが、細胞外、細胞内、生物学的または化学的起源のものが含まれる。そのような試験において用いるプロテアーゼポリペプチドは、任意の形状であることができ、好ましくは、溶液中で遊離しているか、固体支持体に結合しているか、細胞表面上に存在するか、または細胞内に存在する。当業者は、本発明のプロテアーゼが基質ポリペプチドを切断する速度の変化を測定することができる。当業者はまた、例えば、プロテアーゼポリペプチドと試験化合物の間の複合体の形成を測定することができる。あるいは、当業者は、試験化合物により引き起こされる、プロテアーゼポリペプチドとその基質との間の複合体形成の減少を調べることができる。

40

【0090】

本発明のプロテアーゼポリペプチドの活性は、例えば、化学的に合成されたペプチドリガンドに結合するかまたはそれにより活性化される能力を調べることにより決定することができる。あるいは、プロテアーゼポリペプチドの活性は、カルシウム等の金属イオン、ホルモン、ケモカイン、神経ペプチド、神経伝達物質、ヌクレオチド、脂質、臭気物質、および光子に結合する能力を調べることによりアッセイすることができる。すなわち、プロテアーゼポリペプチドの活性の調節剤は、プロテアーゼの機能、例えばプロテアーゼの結

50

合特性または蛋白質基質またはポリペプチド基質を切断する等の活性または膜局在を変化させるであろう。

【0091】

本発明の方法の種々の態様においては、アッセイは、酵母成長アッセイ、エクオリンアッセイ、ルシフェラーゼアッセイ、有糸分裂促進アッセイ、MAPプロテアーゼ活性アッセイ、ならびに当該技術分野において一般に知られている結合または機能に基づくプロテアーゼ活性の他のアッセイの形をとることができる。これらの態様のいくつかにおいては、本発明は、セリンプロテアーゼ、システインプロテアーゼ、アスパルチルプロテアーゼ、メタロプロテアーゼ、トレオニンプロテアーゼ、および他のプロテアーゼの任意のものを含む。本発明にしたがうプロテアーゼの生物学的活性には、限定されないが、天然のまたは合成のリガンドの結合、ならびに当該技術分野において知られるプロテアーゼの機能的活性のいずれかが含まれる。プロテアーゼ活性の非限定的例には、ポリペプチド鎖の切断、プロ形のポリペプチド鎖の活性産物へのプロセッシング、種々の形の貫膜シグナリング、および／または細胞外マトリクスの修飾が含まれる。

10

【0092】

本発明の調節剤は、種々の化学構造を示し、これは一般に、天然のプロテアーゼリガンドの模倣体、およびプロテアーゼのペプチドおよび非ペプチドアロステリックエフェクターに分類することができる。本発明は適切な調節剤の起源を限定するものではなく、これは、天然起源、例えば植物、動物または無機抽出物、または非天然起源、例えば小分子ライブラリ（コンビナトリアル化学方法により構築したライブラリを含む）およびペプチドラ

20

【0093】

蛋白質をコードするcDNAを薬剤発見プログラムにおいて用いることはよく知られている。ハイスループットスクリーニング（HTS）において1日に数千種類の未知の化合物を試験することができるアッセイが詳細に記載されている。文献は、薬剤発見のためにHTS結合アッセイにおいて放射性標識リガンドを使用する例を多数記載している（Williams, Medicinal Research Reviews, 1991, 11, 147-184を参照；総説については、Sweetnam, et al., J. Natural Products, 1993, 56, 441-455）。結合アッセイHTSのためには、組換え蛋白質が好ましい。これは特異性がより高く（より高い相対純度）、大量のレセプター材料を生成する能力を提供し、広範な種類のフォーマットで用いることができるためである（Hodgson, Biol Technology, 1992, 10, 973-980を参照；その全体を本明細書の一部としてここに引用する）。組換え蛋白質の機能的発現には種々の異種システムが利用可能であり、当業者にはよく知られている。そのようなシステムには、細菌（Strosberg, et al., Trends in Pharmacological Sciences, 1992, 13, 95-98）、酵母（Pausch, Trends in Biotechnology, 1997, 15, 487-494）、何種類かの昆虫細胞（Vanden Broeck, Int. Rev. Cytology, 1996, 164, 189-268）、両生類細胞（Jayawickreme et al., Current Opinion in Biotechnology, 1997, 8, 629-634）およびいくつかの哺乳動物細胞株（CHO, HEK293, COS等；Gerhardt, et al., Eur. J. Pharmacology, 1997, 334, 1-23を参照）が含まれる。これらの例は他の可能な細胞発現システム、例えば線虫から得られる細胞株を使用することを除外するものではない（PCT出願WO98/37177）。

30

40

【0094】

発現されたプロテアーゼを、その規定されたリガンド（この場合にはこれを活性化する対応するペプチド）とともにHTS結合アッセイにおいて用いることができる。同定されたペプチドは、適当な放射性同位体、例えば、限定されないが、 ^{125}I 、 ^3H 、 ^{35}S または ^{32}P で、当業者によく知られる方法により標識する。あるいは、ペプチドは、適当

50

な蛍光誘導体を用いてよく知られる方法により標識してもよい (Baindur, et al., Drug Dev. Res., 1994, 33, 373-398; Rogers, Drug Discovery Today, 1997, 2, 156-160)。組換え蛋白質を発現する細胞株から作成した膜調節物において、レセプターに特異的に結合する放射活性リガンドは、いくつかの標準的な方法の1つ、例えば、レセプター-リガンド複合体を濾過して結合したリガンドを未結合リガンドから分離する方法において、HTSアッセイにより検出することができる (Williams, Med. Res. Rev., 1991, 11, 147-184.; Sweetnam, et al., J. Natural Products, 1993, 56, 441-455)。別の方法としては、シンチレーション近接アッセイ (SPA) またはそのような分離を必要としないフラッシュプレート (Flash Plate) フォーマットが挙げられる (Nakayama, Cur. Opinion Drug Disc. Dev., 1998, 1, 85-91 Bosse, et al., J. Biomolecular Screening, 1998, 3, 285-292)。蛍光リガンドの結合は、例えば、蛍光エネルギー転移 (FRET)、結合したリガンドの直接分光蛍光分析、または蛍光分極等の種々の方法により検出することができる (Rogers, Drug Discovery Today, 1997, 2, 156-160; Hill, Cur. Opinion Drug Disc. Dev., 1998, 1, 92-97)。

10

【0095】

プロテアーゼおよび異種プロテアーゼポリペプチドの機能的発現に必要な天然の結合パートナーは、宿主細胞の天然の構成物であってもよく、またはよく知られる組換え技術により導入してもよい。プロテアーゼポリペプチドは無傷であってもよく、キメラであってもよい。プロテアーゼ活性化により他の天然蛋白質が刺激または阻害され、これらの事象を測定可能な応答に結びつけることができる。

20

【0096】

そのような生物学的応答の例には、限定されないが、以下のものが含まれる：特別に遺伝子工学処理した酵母細胞において制限栄養の非存在下で生存する能力 (Pausch, Trends in Biotechnology, 1997, 15, 487-494)；蛍光色素により測定される細胞内Ca²⁺濃度の変化 (Murphy, et al., Cur. Opinion Drug Disc. Dev., 1998, 1, 192-199)。蛍光の変化を用いてリガンド誘導性の膜電位または細胞内pHの変化をモニターすることもできる。このような目的のためにHTSに適した自動化システムが記載されている (Schroeder, et al., J. Biomolecular Screening, 1996, 1, 75-80)。共通のセカンドを測定するためのアッセイもまた利用可能であるが、これらは一般的にはHTSには好ましくない。

30

【0097】

本発明は、プロテアーゼポリペプチドに結合するリガンドの、またはプロテアーゼポリペプチドによる基質切断の阻害剤をスクリーニングし同定する多数のアッセイを企図する。1つの例においては、プロテアーゼポリペプチドを固定化し、候補調節剤、例えば阻害剤化合物の存在下および非存在下において、結合パートナーまたは基質との相互作用を評価することができる。別の例においては、プロテアーゼポリペプチドとその結合パートナーまたは基質との間の相互作用は、溶液アッセイにおいて、候補阻害剤化合物の存在下および非存在下の両方で評価することができる。いずれのアッセイにおいても、阻害剤はプロテアーゼポリペプチドとその天然の結合パートナーとの間の結合またはプロテアーゼポリペプチドが基質分子を切断する活性を減少させる化合物として同定される。別の企図されるアッセイには、PCT国際公開WO95/20652 (1995年8月3日公開；図面および表を含め本明細書の一部としてここに引用する) に記載されるような、トランスフォームしたまたはトランスフェクトした宿主細胞におけるポジティブシグナルの検出により蛋白質/蛋白質相互作用の阻害剤を同定する、ジハイブリッドアッセイの変種が含まれる。

40

50

【0098】

本発明により企図される調節剤の候補には、潜在的活性化剤または潜在的阻害剤のいずれかのライブラリから選択される化合物が含まれる。小分子調節剤の同定に用いられる多くの多様なライブラリが存在し、これには、例えば、(1)化学ライブラリ、(2)天然産物ライブラリ、および(3)ランダムペプチド、オリゴヌクレオチドまたは有機分子から構成されるコンビナトリアルライブラリが含まれる。化学ライブラリはランダムな化学構造から構成され、その一部は既知の化合物の類似体または他の薬剤発見スクリーニングにおいて"ヒット"または"リード"として同定された化合物の類似体であるが、別のものは天然産物から誘導され、さらに別のものは、無指向性有機合成化学から作成する。天然産物ライブラリは、(1)土壌、植物または海洋生物からの培養液の発酵および抽出、または(2)植物または海洋生物の抽出、によりスクリーニング用の混合物を作成するために用いられる、微生物、動物、植物、または海洋生物の収集物である。天然産物ライブラリには、ポリケタイド、非リボソームペプチド、およびそれらの変種(天然に生じない)が含まれる。総説については、Science 282:63-68(1998)を参照。コンビナトリアルライブラリは、混合物としての多数のペプチド、オリゴヌクレオチド、または有機化合物から構成される。これらのライブラリは、伝統的自動化合成法、PCR、クローニング、または私有の合成法により比較的容易に製造することができる。特に興味深いものは非ペプチドコンビナトリアルライブラリである。さらに興味深い他のライブラリには、ペプチド、蛋白質、ペプチド模倣物、多重平行合成収集物、リコンビナトリアル、およびポリペプチドライブラリが含まれる。コンビナトリアルケミストリおよびこれから作成されるライブラリの総説については、Myers, Curr. Opin. Biotechnol. 8:701-707(1997)を参照。本明細書に記載される、種々のライブラリを用いる調節剤の同定により、候補物質である"ヒット"(または"リード")を改変して、"ヒット"が活性を調節する能力を最適化することができる。

10

20

30

【0099】

本発明により企図される他の候補阻害剤を設計することができ、これには溶解型の結合パートナー、ならびにキメラ、または融合蛋白質等の結合パートナーが含まれる。本明細書において用いる場合、"結合パートナー"は、上述したような天然の結合パートナーならびにキメラポリペプチド、天然のリガンド以外のペプチド調節剤、抗体、抗体フラグメント、および同定されたプロテアーゼ遺伝子の発現産物に免疫特異的な抗体ドメインを含む改変型化合物を広く包含する。

【0100】

プロテアーゼポリペプチドの特異的ペプチドリガンドを同定するために他のアッセイを用いることができ、これには、試験リガンドの標的蛋白質への直接結合の測定により標的蛋白質のリガンドを同定するアッセイ、ならびにイオンスプレー質量分析計/HP LC方法または他の物理学および分析的方法を用いてアフィニティー限外濾過により標的蛋白質のリガンドを同定するアッセイが含まれる。あるいは、そのような結合相互作用は、Fields et al., Nature, 340:245-246(1989)、およびFields et al., Trends in Genetics, 10:286-292(1994)(いずれも本明細書の一部としてここに引用する)に記載される酵母ツーハイブリッド系を用いて間接的に評価することができる。ツーハイブリッド系は2つの蛋白質またはポリペプチドの間の相互作用を検出するための遺伝子アッセイである。これを用いて、目的とする既知の蛋白質に結合する蛋白質を同定し、相互作用に重要なドメインまたは残基を線引きすることができる。この方法論の変形が、DNA結合蛋白質をコードする遺伝子のクローニング、蛋白質に結合するペプチドの同定、および薬剤のスクリーニング用に関連されている。ツーハイブリッド系は、1対の相互作用する蛋白質が転写活性化ドメインをレポーター遺伝子上流活性化配列(UAS)に結合するDNA結合ドメインの近くに配置させる能力を利用しており、一般に酵母において行われる。アッセイにおいては、(1)第1の蛋白質に融合されたDNA結合ドメイン、および(2)第2の蛋白質に融合された活性化ドメイン、をコードする2つのハイブリッド遺伝子を構築する

40

50

ことが必要である。DNA結合ドメインは、第1のハイブリッド蛋白質をレポーター遺伝子のUASにターゲティングする。しかし、ほとんどの蛋白質は活性化ドメインを有していないため、このDNA結合ハイブリッド蛋白質はレポーター遺伝子の転写を活性化しない。活性化ドメインを含む第2のハイブリッド蛋白質は、UASに結合しないため、それ自身ではレポーター遺伝子の発現を活性化しない。しかし、両方のハイブリッド蛋白質が存在する場合には、第1の蛋白質と第2の蛋白質との非共有結合的相互作用が活性化ドメインをUASにつなぎ、レポーター遺伝子の転写を活性化させる。例えば、第1の蛋白質が、他の蛋白質または核酸と相互作用することが知られているプロテアーゼ遺伝子産物、またはそのフラグメントである場合、このアッセイを用いて結合相互作用を妨害する薬剤を検出することができる。異なる試験薬剤を系に添加してレポーター遺伝子の発現をモニターする。阻害剤が存在すると、レポーターシグナルが発生しない。

10

【0101】

プロテアーゼポリペプチド遺伝子産物の機能が未知であり、遺伝子産物に結合するリガンドが知られていない場合においても、酵母ツーハイブリッドアッセイを用いて遺伝子産物に結合する蛋白質を同定することができる。プロテアーゼポリペプチドまたはそのフラグメントに結合する蛋白質を同定するアッセイにおいては、プロテアーゼポリペプチド（またはフラグメント）およびUAS結合ドメイン（すなわち、第1の蛋白質）の両方をコードする融合ポリヌクレオチドを用いることができる。さらに、それぞれが活性化ドメインに融合された異なる第2の蛋白質をコードする多数のハイブリッド遺伝子を製造し、このアッセイにおいてスクリーニングすることができる。典型的には、第2の蛋白質は総cDNAまたはゲノムDNA融合ライブラリの1またはそれ以上のメンバーによりコードされ、ここで、それぞれの第2の蛋白質のコーディング領域は活性化ドメインに融合されている。この系は広範な種類の蛋白質に適用することができ、第2の結合蛋白質の種類または機能を知ることすら必要ではない。この系は非常に感度が高く、他の方法では明らかにされない相互作用を検出することができる。過渡的相互作用であっても、転写を誘発して安定なmRNAを生成し、これは繰り返し翻訳されてレポーター蛋白質を生成することができる。

20

【0102】

他のアッセイを用いて、標的蛋白質に結合する薬剤を検索してもよい。試験リガンドの標的蛋白質への直接結合を同定する1つのそのようなスクリーニング方法が、米国特許5,585,277（本明細書の一部としてここに引用する）に議論されている。この方法は、蛋白質は一般に折り畳まれた状態と折り畳まれていない状態の混合物として存在し、2つの状態の間で継続的に交替しているという原理に基づく。試験リガンドが折り畳まれた形の標的蛋白質に結合する場合（すなわち、試験リガンドが標的蛋白質のリガンドである場合）、リガンドに結合した標的蛋白質分子は折り畳まれた状態のままである。すなわち、折り畳まれた標的蛋白質は、標的蛋白質に結合する試験リガンドの存在下で、リガンドの非存在下におけるよりも多く存在する。リガンドの標的蛋白質への結合は、折り畳まれた状態と折り畳まれていない状態の標的蛋白質とを区別する任意の方法により検出することができる。このアッセイを行うためには、標的蛋白質の機能がわかっている必要はない。この方法により事実上すべての薬剤、例えば、限定されないが、金属、ポリペプチド、蛋白質、脂質、多糖類、ポリヌクレオチドおよび小有機分子を試験リガンドとして評価することができる。

30

40

【0103】

標的蛋白質のリガンドを同定する別の方法は、Wieboldt et al., Anal. Chem., 69:1683-1691 (1997)（本明細書の一部としてここに引用する）に記載されている。この手法は、1回に溶液相中の20-30種類の薬剤のコンビナトリアルライブラリを標的蛋白質に対する結合についてスクリーニングする。簡単な膜洗浄により標的蛋白質に結合する薬剤を他のライブラリ成分から分離する。次にフィルター上に保持された特定の選択された分子を標的蛋白質から遊離させ、HPLCおよび気体補助エレクトロスプレー（イオンスプレー）イオン化質量分析により分析する。この

50

方法は、標的蛋白質に対して最も高い親和性を有するライブラリ成分を選択し、特に小分子ライブラリについて有用である。

【0104】

本発明の好ましい態様においては、プロテアーゼ活性を調節する化合物をスクリーニングする方法は、試験化合物をプロテアーゼポリペプチドと接触させ、化合物とプロテアーゼポリペプチドとの間の複合体の存在についてアッセイすることを含む。そのようなアッセイにおいては、典型的にはリガンドを標識する。適当にインキュベートした後、遊離のリガンドを結合した形のリガンドから分離する。遊離のまたは複合体を形成していない標識の量は、特定の化合物がプロテアーゼポリペプチドに結合する能力の尺度である。

【0105】

本発明の別の態様においては、プロテアーゼポリペプチドに対して適切な結合親和性を有する化合物のハイスループットスクリーニングを用いる。簡単には、多数の様々な小ペプチド試験化合物を固体支持体上で合成する。ペプチド試験化合物をプロテアーゼポリペプチドと接触させ、洗浄する。次に、結合したプロテアーゼポリペプチドを当該技術分野においてよく知られる方法により検出する。本発明の精製ポリペプチドはまた、上述した薬剤スクリーニング手法において用いるために、直接プレート上にコーティングすることができる。さらに、非中和抗体を用いて蛋白質を捕獲し、これを固体支持体上に固定化することができる。

【0106】

本発明の別の態様は、本発明のポリペプチドに結合しうる中和抗体がポリペプチドに対する結合について試験化合物と特異的に競合する、競合スクリーニングアッセイを用いることを含む。このようにして、抗体を用いて、プロテアーゼポリペプチドと1またはそれ以上の抗原性決定基を共有するペプチドの存在を検出することができる。放射性標識した競合結合実験は、A. H. Lin et al. *Antimicrobial Agents and Chemotherapy*, 1997, vol. 41, no. 10, pp. 2127-2131に記載されており、この開示はその全体を本明細書の一部としてここに引用する。

【0107】

治療方法

本発明は、治療を必要とする患者に、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117、および配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列と実質的に同一のプロテアーゼポリペプチド、および本発明の任意の他のプロテアーゼポリペプチドを投与することにより、疾病または疾患を治療する方法を含む。”遺伝子治療”の節において説明されるように、本発明のプロテアーゼポリペプチドはまた、プロテアーゼポリペプチドのインビボ発現のための適当なポリヌクレオチド手段を投与することにより、間接的に投与することができる。好ましくは、プロテアーゼポリペプチドは、上述の配列の1つに、少なくとも約90%、91%、92%、93%、94%、95%、96%、97%、98%、99%または100%の同一性を有する。

【0108】

別の観点においては、本発明は、治療を必要とする患者に、配列番号60、配列番号61

10

20

30

40

50

、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択される配列と実質的に同一のプロテアーゼの活性を調節する物質を投与することにより、疾病または疾患を治療または予防する方法を提供する。

【0109】

好ましくは、疾病は、癌、免疫関連疾病および疾患、心臓血管疾病、脳またはニューロン関連疾病、および代謝性疾患からなる群より選択される。より詳細には、これらの疾病には、組織、血液、または造血細胞起源の癌、特に、乳、結腸、肺、前立腺、子宮頸部、脳、卵巣、膀胱、または腎臓が関与する癌；中枢神経または末梢神経系疾病および状態、例えば、片頭痛、痛み、性的機能不全、気分障害、注意障害、認識障害、低血圧症、および高血圧症；精神病性および神経学的疾患、例えば、不安、精神分裂病；躁うつ病、せん妄、痴呆、重症の精神遅滞および運動異常症、例えば、ハンチントン病、またはツレット症候群；神経変性性疾患、例えばアルツハイマー病、パーキンソン病、多発性硬化症、および筋萎縮性側方硬化症；HIV-1、HIV-2または他のウイルスまたはプリオン体または真菌または細菌生物により引き起こされるウイルスまたは非ウイルス感染；代謝性疾患、例えば、特に糖尿病および肥満、およびこれらに関連する症候群；心臓血管疾患、例えば、再灌流再狭窄、冠状動脈血栓症、凝固疾患、制御されない細胞成長疾患、アテローム性動脈硬化症；眼性疾患、例えば、緑内障、網膜症、および黄斑変性；炎症性疾患、例えば、慢性関節リウマチ、慢性炎症性腸疾病、慢性炎症性骨盤疾病、多発性硬化症、ぜん息、変形性関節症、乾癬、アテローム性動脈硬化症、鼻炎、自己免疫、および臓器移植拒絶が含まれる。

【0110】

好ましい態様においては、本発明は、治療を必要とする患者に、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を有するプロテアーゼポリペプチドの活性を調節する物質を投与することにより、疾病または疾患を治療または予防する方法を提供する。好ましくは、疾病は、癌、免疫関連疾病および疾患、心臓血管疾病、脳またはニューロン関連疾病、および代謝性疾患からなる群より選択される。より詳細には、これらの疾病には、組織、血液、または造血細胞起源の癌、特に乳、結腸、肺、前立腺、子宮頸部、脳、卵巣、膀胱、または腎臓が関与する癌；中枢神経または末梢神経系疾病および状態、例えば、片頭痛、痛み、性的機能不全、気分障害、注意障害、認識障害、低血圧症、および高血圧症；精神

病性および神経学的疾患，例えば不安，精神分裂病，躁うつ病，せん妄，痴呆，重症の精神遅滞および運動異常症，例えばハンチントン病またはツレット症候群；神経変性性疾患，例えばアルツハイマー病，パーキンソン病，多発性硬化症，および筋萎縮性側方硬化症；H I V-1，H I V-2または他のウイルスまたはプリオン体または真菌または細菌生物により引き起こされるウイルスまたは非ウイルス感染；代謝性疾患，例えば，特に糖尿病および肥満およびこれらに関連する症候群；心臓血管疾患，例えば再灌流再狭窄，冠状動脈血栓症，凝固疾患，制御されない細胞成長疾患，アテローム性動脈硬化症；眼性疾患，例えば緑内障，網膜症，および黄斑変性；炎症性疾患，例えば慢性関節リウマチ，慢性炎症性腸疾患，慢性炎症性骨盤疾患，多発性硬化症，ぜん息，変形性関節症，乾癬，アテローム性動脈硬化症，鼻炎，自己免疫，および臓器移植拒絶が含まれる。

10

【0111】

好ましくは，疾病は，免疫関連疾患および疾患，心臓血管疾患，および癌からなる群より選択される。最も好ましくは，免疫関連疾患および疾患は，慢性関節リウマチ，慢性炎症性腸疾患，慢性炎症性骨盤疾患，多発性硬化症，ぜん息，変形性関節症，乾癬，アテローム性動脈硬化症，鼻炎，自己免疫，および臓器移植からなる群より選択される。

【0112】

プロテアーゼ関連疾患または疾病の治療に有用な物質は，好ましくは，問題とする疾病または疾患の治療に対応する活性についての1またはそれ以上のインビトロアッセイにおいて陽性の結果を示す（そのようなアッセイの例は，本明細書，例えば実施例7において提供される）。望ましい活性についてスクリーニングすることができる物質の例は，この節（プロテアーゼ活性を調節する物質を同定するためのスクリーニング方法）を含む，本明細書全体で提供され参照される。プロテアーゼの活性を調節する物質は，好ましくは，限定されないが，この節および以下の実施例7で参照される方法およびスクリーニングにより，または他の任意の適当な方法により決定される，アンチセンスオリゴヌクレオチド，リボザイムおよび他のプロテアーゼ阻害剤を含む。アンチセンスオリゴヌクレオチドおよびリボザイムの使用は，以下の“遺伝子治療”の節において詳細に説明される。

20

【0113】

“予防する”との用語は，生物が異常な状態に罹患するかこれを発達させる可能性を減少させることを表す。

【0114】

“治療する”との用語は，生物において治療効果を有し，異常な状態を少なくとも部分的に緩和するかまたは除外することを表す。

30

【0115】

“治療効果”との用語は，異常な状態を引き起こすかまたはこれに寄与する阻害または活性化率を表す。治療効果は，異常な状態の1またはそれ以上の症状をある程度緩和する。異常な状態の治療に関して，治療効果は，以下の1またはそれ以上を表すことができる：（a）細胞の増殖，成長，および／または分化の増加または減少；（b）細胞死の活性化または阻害（すなわち，遅延または停止）；（c）変性の阻害；（d）異常な状態に伴う1またはそれ以上の症状のある程度の緩和；および（e）影響を受けた細胞の集団の機能の増強。異常な状態に対して有効性を示す化合物は，本明細書に記載されるようにして同定することができる。

40

【0116】

“異常な状態”との用語は，生物の細胞または組織における，その生物における正常な機能からはずれた機能を表す。異常な状態は，細胞増殖，細胞分化，または細胞生存に関連する。

【0117】

異常な細胞増殖状態には，癌，例えば繊維性およびメサンギウム疾患，異常な新脈管形成および脈管形成，創傷治癒，乾癬，真性糖尿病，および炎症が含まれる。

【0118】

異常な分化状態には，限定されないが，神経変性性疾患，遅い創傷治癒速度，および遅い

50

組織移植治癒速度が含まれる。

【0119】

異常な細胞生存状態は、プログラムされた細胞死（アポトーシス）経路が活性化されているかまたは排除されている状態に関連する。多くのプロテアーゼがアポトーシス経路に関連している。プロテアーゼのいずれかの機能の異常は、細胞の不死または未成熟細胞死につながりうる。

【0120】

シグナル伝達プロセスにおけるプロテアーゼの機能に関連して，“異常な”との用語は、生物において過剰発現または過小発現されているか，その触媒活性が野生型プロテアーゼ活性より低いかまたは高いように変異しているか，天然の結合パートナーともはや相互作用できないように変異しているか，別の蛋白質によりもはや修飾されないか，または天然の結合パートナーともはや相互作用しない，プロテアーゼを表す。

10

【0121】

”投与する”との用語は，化合物を生物の細胞または組織内に取り込ませる方法に関連する。異常な状態は，生物の細胞または組織が生物中または生物外に存在する場合，予防または治療することができる。生物の外に存在する細胞は，細胞培養皿中で維持または成長させることができる。生物中に含まれる細胞については，当該技術分野には化合物を投与する多くの手法が存在し，例えば，限定されないが，経口，非経口，経皮，注入およびエアロゾル外用が含まれる。生物外の細胞については，当該技術分野には化合物を投与する多数の手法が存在し，例えば，限定されないが，細胞マイクロインジェクション手法，トランスフォーメーション手法，および担体手法が含まれる。

20

【0122】

異常な状態はまた，シグナル伝達経路に異常を有する一群の細胞を有する生物に化合物を投与することにより，予防または治療することができる。次に，化合物の投与が生物機能に及ぼす影響をモニターすることができる。生物は，好ましくはマウス，ラット，ウサギ，モルモット，またはヤギであり，より好ましくは有尾サルまたは無尾サルであり，最も好ましくはヒトである。

【0123】

別の観点においては，本発明は，疾病または疾患の診断道具として，試料中でプロテアーゼポリペプチドを検出する方法を特徴とする。該方法は，（a）試料を，ハイブリダイゼーションアッセイ条件下で，配列番号60，配列番号61，配列番号62，配列番号63，配列番号64，配列番号65，配列番号66，配列番号67，配列番号68，配列番号69，配列番号70，配列番号71，配列番号72，配列番号73，配列番号74，配列番号75，配列番号76，配列番号77，配列番号78，配列番号79，配列番号80，配列番号81，配列番号82，配列番号83，配列番号84，配列番号85，配列番号86，配列番号87，配列番号88，配列番号89，配列番号90，配列番号91，配列番号92，配列番号93，配列番号94，配列番号95，配列番号96，配列番号97，配列番号98，配列番号99，配列番号100，配列番号101，配列番号102，配列番号103，配列番号104，配列番号105，配列番号106，配列番号107，配列番号108，配列番号109，配列番号110，配列番号111，配列番号112，配列番号113，配列番号114，配列番号115，配列番号116，配列番号117および配列番号118に記載されるアミノ酸配列からなる群より選択されるアミノ酸配列を有するプロテアーゼポリペプチドの核酸標的領域にハイブリダイズする核酸プローブと接触させ，ここで，前記プローブは，ポリペプチド，そのフラグメントをコードする核酸配列，および配列およびフラグメントの相補体を含み；そして（b）プローブ：標的領域ハイブリッドの存在または量を疾病の指標として検出する，の各工程を含む。

30

40

【0124】

本発明の好ましい態様においては，疾病または疾患は，慢性関節リウマチ，動脈硬化症，自己免疫疾患，臓器移植，心筋梗塞，心筋症，発作，腎不全，酸化的ストレス関連神経変

50

性疾患，および癌からなる群より選択される。好ましくは，疾病は，癌，免疫関連疾病および疾患，心臓血管疾病，脳またはニューロン関連疾病，および代謝性疾患からなる群より選択される。より詳細には，これらの疾病には，組織，血液，または造血細胞起源の癌，特に乳，結腸，肺，前立腺，子宮頸部，脳，卵巣，膀胱，または腎臓が関与する癌；中枢神経または末梢神経系疾病および状態，例えば，片頭痛，痛み，性的機能不全，気分障害，注意障害，認識障害，低血圧症，および高血圧症；精神病性および神経学的疾患，例えば，不安，精神分裂病，躁うつ病，せん妄，痴呆，重症の精神遅滞および運動異常症，例えば，ハンチントン病またはツレット症候群；神経変性性疾患，例えば，アルツハイマー病，パーキンソン病，多発性硬化症，および筋萎縮性側方硬化症；H I V - 1，H I V - 2 または他のウイルスまたはプリオン体または真菌または細菌生物により引き起こされるウイルスまたは非ウイルス感染；代謝性疾患，例えば，特に糖尿病および肥満，およびこれらに関連する症候群；心臓血管疾患，例えば，再灌流再狭窄，冠状動脈血栓症，凝固疾患，制御されない細胞成長疾患，アテローム性動脈硬化症；眼性疾患，例えば，緑内障，網膜症，および黄斑変性；炎症性疾患，例えば，慢性関節リウマチ，慢性炎症性腸疾病，慢性炎症性骨盤疾病，多発性硬化症，ぜん息，変形性関節症，乾癬，アテローム性動脈硬化症，鼻炎，自己免疫，および臓器移植拒絶が含まれる。

10

【0125】

プロテアーゼ”標的領域”とは，配列番号1，配列番号2，配列番号3，配列番号4，配列番号5，配列番号6，配列番号7，配列番号8，配列番号9，配列番号10，配列番号11，配列番号12，配列番号13，配列番号14，配列番号15，配列番号16，配列番号17，配列番号18，配列番号19，配列番号20，配列番号21，配列番号22，配列番号23，配列番号24，配列番号25，配列番号26，配列番号27，配列番号28，配列番号29，配列番号30，配列番号31，配列番号32，配列番号33，配列番号34，配列番号35，配列番号36，配列番号37，配列番号38，配列番号39，配列番号40，配列番号41，配列番号42，配列番号43，配列番号44，配列番号45，配列番号46，配列番号47，配列番号48，配列番号49，配列番号50，配列番号51，配列番号52，配列番号53，配列番号54，配列番号55，配列番号56，配列番号57，配列番号58，および配列番号59に記載される配列からなる群より選択されるヌクレオチド塩基配列，または対応する全長配列，核酸プローブが特異的にハイブリダイズするであろうその機能的誘導体，またはそのフラグメントである。特異的なハイブリダイゼーションとは，他の核酸の存在下において，プローブが本発明のプロテアーゼの標的領域にのみ検出可能なようにハイブリダイズすることを示す。推定標的領域は，当該技術分野においてよく知られる，データベース中の最も近い関連配列のアラインメントおよび比較からなる方法により同定することができる。

20

30

【0126】

好ましい態様においては，核酸プローブは，配列番号60，配列番号61，配列番号62，配列番号63，配列番号64，配列番号65，配列番号66，配列番号67，配列番号68，配列番号69，配列番号70，配列番号71，配列番号72，配列番号73，配列番号74，配列番号75，配列番号76，配列番号77，配列番号78，配列番号79，配列番号80，配列番号81，配列番号82，配列番号83，配列番号84，配列番号85，配列番号86，配列番号87，配列番号88，配列番号89，配列番号90，配列番号91，配列番号92，配列番号93，配列番号94，配列番号95，配列番号96，配列番号97，配列番号98，配列番号99，配列番号100，配列番号101，配列番号102，配列番号103，配列番号104，配列番号105，配列番号106，配列番号107，配列番号108，配列番号109，配列番号110，配列番号111，配列番号112，配列番号113，配列番号114，配列番号115，配列番号116，配列番号117および配列番号118に記載されるアミノ酸配列，または対応する全長アミノ酸配列，またはそれらの機能的誘導体からなる群より選択される配列の，少なくとも6，12，75，90，105，120，150，200，250，300または350個の連続するアミノ酸をコードするプロテアーゼ標的領域にハイブリダイズする。ハイブリダイゼ

40

50

ーション条件は、他の核酸分子の存在下でプロテアーゼ遺伝子とのみハイブリダイゼーションが生ずるような条件であるべきである。ストリンジェントなハイブリダイゼーション条件下では、高度に相補的な核酸配列のみがハイブリダイズする。好ましくは、そのような条件は、20個の連続するヌクレオチド中に1または2個より多いミスマッチを有する核酸のハイブリダイゼーションを防止する。そのような条件は、Bergerら(1987, Guide to Molecular Cloning Techniques, pg 421, (図面および表を含めその全体を本明細書の一部としてここに引用する))において定義される。

【0127】

試料中のプロテアーゼ遺伝子の検出により診断することができる疾病には、プロテアーゼ核酸(DNAおよび/またはRNA)が正常細胞と比較して増幅されている疾病が含まれる。"増幅"とは、正常細胞と比較して、細胞中でプロテアーゼDNAまたはRNAの数が増加していることを意味する。正常細胞においては、プロテアーゼは典型的には1コピーの遺伝子として見いだされる。選択された疾病においては、プロテアーゼ遺伝子の染色体位置が増幅されて、遺伝子の多数のコピーまたは増幅が生ずる。遺伝子増幅は、そのようなプロテアーゼRNAを増幅させることができ、またはプロテアーゼRNAはプロテアーゼDNAの増幅なしに増幅することができる。

【0128】

RNAに関する場合、"増幅"は、細胞においてプロテアーゼRNAが検出可能なように存在することでありうる。ある正常細胞においては、プロテアーゼRNAの基底発現がないためである。他の正常細胞においては、プロテアーゼの発現の基底レベルが存在し、したがってこれらの場合には増幅は基底レベルと比較して少なくとも1-2倍、好ましくはそれより多い、検出可能な存在の増加を示すプロテアーゼRNAの増加である。

【0129】

試料中のプロテアーゼ核酸の検出により診断することができる疾病には、好ましくは癌が含まれる。本発明の核酸探索方法に適した試験試料には、例えば、細胞または細胞の核酸抽出物、または体液が含まれる。上述の方法において用いられる試料は、アッセイフォーマット、検出方法、およびアッセイする組織、細胞または抽出物の性質により様々でありうる。細胞の核酸抽出物を調製する方法は当該技術分野においてよく知られており、用いる方法に適合した試料を得るために容易に適合させることができる。

【0130】

最後の観点においては、本発明は、疾病または疾患の診断道具として、試料中のプロテアーゼポリペプチドを検出する方法を特徴とする。該方法は、(a)試料中のプロテアーゼポリペプチドをコードする核酸標的領域を、プロテアーゼポリペプチド、またはその1またはそれ以上フラグメントをコードする対照核酸標的領域と比較し、ここで、プロテアーゼポリペプチドは、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載される配列、またはその1またはそれ以上フラグメントからなる群より選択されるアミノ酸配列を有しており；そして(b)標的領域と対照標的領域との間の配列または量の相違を疾病または疾患の指標として検出することを含む。好ましくは、疾病は、癌、免疫関連疾病および疾患、心臓血管疾病、脳またはニューロン関連疾病、および代謝性

疾患からなる群より選択される。

【0131】

より詳細には、これらの疾病には、組織、血液、または造血細胞起源の癌、特に乳、結腸、肺、前立腺、子宮頸部、脳、卵巣、膀胱、または腎臓が関与する癌；中枢神経または末梢神経系疾病および状態、例えば、片頭痛、痛み、性的機能不全、気分障害、注意障害、認識障害、低血圧症、および高血圧症；精神病性および神経学的疾患、例えば、不安、精神分裂病、躁うつ病、せん妄、痴呆、重症の精神遅滞および運動異常症、例えば、ハンチントン病またはツレット症候群；神経変性性疾患、例えば、アルツハイマー病、パーキンソン病、多発性硬化症、および筋萎縮性側方硬化症；H I V-1、H I V-2または他のウイルスまたはプリオン体または真菌または細菌生物により引き起こされるウイルスまたは非ウイルス感染；代謝性疾患、例えば、特に糖尿病および肥満、およびこれらに関連する症候群；心臓血管疾患、例えば、再灌流再狭窄、冠状動脈血栓症、凝固疾患、制御されない細胞成長疾患、アテローム性動脈硬化症；眼性疾患、例えば、緑内障、網膜症、および黄斑変性；炎症性疾患、例えば、慢性関節リウマチ、慢性炎症性腸疾患、慢性炎症性骨盤疾患、多発性硬化症、ぜん息、変形性関節症、乾癬、アテローム性動脈硬化症、鼻炎、自己免疫、および臓器移植拒絶が含まれる。

10

【0132】

本明細書において用いる場合，“比較する”との用語は、試料から単離された核酸標的領域と対照核酸標的領域との間の不一致を同定することを表す。不一致は、ヌクレオチド配列における、例えば挿入、欠失、または点突然変異であってもよく、または所定のヌクレオチド配列の量におけるものであってもよい。配列におけるこれらの不一致を判定する方法は当業者にはよく知られている。“対照”核酸標的領域とは、正常細胞、例えば、先に記載したような疾病を有しない細胞において見いだされる配列または配列の量を表す。

20

【0133】

“ドメイン”との用語は、特定の機能を含むポリペプチドの領域を表す。例えば、シグナル伝達蛋白質のN末端またはC末端ドメインは、例えば、限定されないが、シグナル伝達分子を細胞の異なる領域に局在させる分子に結合し、特定の細胞シグナルを伝播するのに直接関与する他のシグナリング分子に結合する、等の機能を提供することができる。あるドメインは蛋白質の残部と別々に発現させてそれ自身で機能することができるが、他のドメインはその機能を保持するためには無傷の蛋白質の一部のままでなければならない。後者は蛋白質の機能的領域とも称され、これもまたドメインと関連する。

30

【0134】

プロテアーゼの発現は、シグナル伝達経路、例えばR a s / M A Pキナーゼシグナリング経路により調節されうる。さらに、プロテアーゼの活性は、M A Pキナーゼシグナル伝達経路の活性を調節することができる。さらに、プロテアーゼは、異なるシグナル伝達経路の間のコミュニケーションの手段であることが示されるかもしれない。

【0135】

“シグナル伝達経路”との用語は、細胞外シグナルを細胞膜を通して伝播し、細胞内シグナルとなる分子を表す。このシグナルは、次に細胞性応答を刺激することができる。シグナル伝達プロセスに関与するポリペプチド分子は、典型的にはレセプターおよび非レセプター蛋白質チロシンキナーゼ、レセプターおよび非レセプター蛋白質ホスファターゼ、S R Cホモロジー2および3ドメイン、ホスホチロシン結合蛋白質（S R Cホモロジー2（S H 2）およびホスホチロシン結合（P T BおよびP H）ドメイン含有蛋白質）、プロリンリッチ結合蛋白質（S H 3ドメイン含有蛋白質）、G T P a s e、ホスホジエステラーゼ、ホスホリパーゼ、プロリルイソメラーゼ、プロテアーゼ、C a ²⁺結合蛋白質、c A M P結合蛋白質、グアニルシクラーゼ、アデニルシクラーゼ、N O生成蛋白質、ヌクレオチド交換因子および転写因子である。

40

【0136】

上述した本発明の概要は限定的なものではなく、本発明の他の特徴および利点は、以下の本発明の詳細な説明および特許請求の範囲から明らかであろう。

50

【0137】

図面の簡単な説明

図1 A-WWは、ヒト蛋白質プロテアーゼのヌクレオチド配列（配列番号1，配列番号2，配列番号3，配列番号4，配列番号5，配列番号6，配列番号7，配列番号8，配列番号9，配列番号10，配列番号11，配列番号12，配列番号13，配列番号14，配列番号15，配列番号16，配列番号17，配列番号18，配列番号19，配列番号20，配列番号21，配列番号22，配列番号23，配列番号24，配列番号25，配列番号26，配列番号27，配列番号28，配列番号29，配列番号30，配列番号31，配列番号32，配列番号33，配列番号34，配列番号35，配列番号36，配列番号37，配列番号38，配列番号39，配列番号40，配列番号41，配列番号42，配列番号43，配列番号44，配列番号45，配列番号46，配列番号47，配列番号48，配列番号49，配列番号50，配列番号51，配列番号52，配列番号53，配列番号54，配列番号55，配列番号56，配列番号57，配列番号58，および配列番号59）を5'から3'の方向で示す。配列中、Nは任意のヌクレオチドを意味する。

【0138】

図2 A-Sは、配列番号1-59によりコードされるヒトプロテアーゼのアミノ酸配列（配列番号60，配列番号61，配列番号62，配列番号63，配列番号64，配列番号65，配列番号66，配列番号67，配列番号68，配列番号69，配列番号70，配列番号71，配列番号72，配列番号73，配列番号74，配列番号75，配列番号76，配列番号77，配列番号78，配列番号79，配列番号80，配列番号81，配列番号82，配列番号83，配列番号84，配列番号85，配列番号86，配列番号87，配列番号88，配列番号89，配列番号90，配列番号91，配列番号92，配列番号93，配列番号94，配列番号95，配列番号96，配列番号97，配列番号98，配列番号99，配列番号100，配列番号101，配列番号102，配列番号103，配列番号104，配列番号105，配列番号106，配列番号107，配列番号108，配列番号109，配列番号110，配列番号111，配列番号112，配列番号113，配列番号114，配列番号115，配列番号116，配列番号117および配列番号118）を翻訳の方向で示す。配列中、Xは任意のアミノ酸を意味する。

【0139】

発明の詳細な説明

以下の本発明の背景の記載は、本発明の理解を助けるために提供され、本発明に対する先行技術であるかまたはそれを記載すると認めるものではない。

【0140】

プロテアーゼは、他の蛋白質のアミノ酸骨格を切断しうる酵素であり、生体内の多数の多様なプロセスに関与している。その正常な機能には、アポトーシスの調節（カスパーゼ）（Salvesen and Dixon, Cell, 1997, 91: 443-46），血圧の制御（レニン，アンジオテンシン変換酵素）（vanHoof et al., 1991, N Engl J Med. 324 (19): 1305-11，およびBarrett et al., Handbook of Proteolytic Enzymes, 1998, 第254節および第359節, Academic Press, San Diego），組織リモデリングおよび腫瘍侵襲（コラゲナーゼ）（Vuet al., 1998, Cell 93: 411-22, Werb, 1997, Cell, 91: 439-442），アルツハイマー病の発達（ β -セクレターゼ）（De Strooper et al., 1999, Nature 398: 518-22），蛋白質ターンオーバーおよび細胞サイクル制御（プロテオソーム）（Bastians et al., 1999, Mol. Biol. Cell. 10: 3927-41, Gottesman, et al., 1997, Cell, 91: 435-38, Larsen et al., 1997, Cell, 91: 431-34），炎症（TNF- α コンベルターゼ）（Black et al., Nature, 1997, 385: 729-33），および蛋白質ターンオーバー（Bochtler et al., 1999, Annu. Re

v. *Biophys Biomol Struct.* 28:295-317) が含まれる。プロテアーゼは、いくつかの主要なグループに分類することができる：セリンプロテアーゼ、システインプロテアーゼ、アスパルチルプロテアーゼ、メタロプロテアーゼ、トレオニンプロテアーゼ、および他のプロテアーゼ。

【0141】

1. アスパルチルプロテアーゼ (Al:Prosite 番号 PS00141) : 酸プロテアーゼとしても知られるアスパルチルプロテアーゼは、脊椎動物、真菌、植物、レトロウイルスおよびある種の植物ウイルスにおいて広く分布している蛋白質分解酵素のファミリーである。真核生物のアスパラギン酸プロテアーゼは2つのドメインから構成される単量体酵素である。各ドメインは触媒的アスパラギン酸残基の中心に活性部位を含む。2つのドメインは、おそらくは原始的ドメインをコードする先祖遺伝子の重複から進化したものと考えられる。このクラスの酵素には、カテプシンE、レニン、プレセニリン (PS1)、およびAPPセクレターゼが含まれる。

10

【0142】

カテプシンE

カテプシンEは胃腸管に見いだされる免疫学的に独特のアスパラギン酸プロテアーゼである (Azuma et al., 1992, *J. Biol. Chem.*, 267:1609-1614)。カテプシンEは、食物蛋白質の消化には関与していないようである細胞内プロテイナーゼである。これは胃の上皮粘膜産生細胞の表面に最も高い濃度で見いだされる。これは胎児胃において発現される最初のアスパラギン酸プロテイナーゼであり、胃癌の半分以上に見いだされる。したがって、これは '胎児発癌性 (oncofetal)' 抗原であるようである。この胃癌との関連性は、これがこの疾病の発達において役割を果たすかもしれないことを示唆する。

20

【0143】

レニン

レニンは腎臓の傍糸球体細胞により放出され、アンジオテンシノーゲの活性化経路 (アルドステロン放出、血管収縮、および血圧の上昇をもたらすカスケード) の第1段階を触媒する。レニンはアンジオテンシノーゲンを切断して、アンジオテンシンIを形成し、これは血圧および電解質平衡の重要なレギュレータであるアンジオテンシンI変換酵素によりアンジオテンシンIIに変換される。レニンは腎臓以外の臓器、例えば脳でも生じ、ここで多くの活性の制御に関与していることが示唆されている。

30

【0144】

プレセニリン蛋白質

遺伝性型の疾病を有するアルツハイマー病 (AD) 患者は、プレセニリン蛋白質 (PSEN1; PSEN2) またはアミロイド前駆体蛋白質 (APP) における変異を有する。これらの疾病連鎖変異により、より長い形のアミロイドペプタ (AD脳に見いだされるアミロイド沈着の主要成分) の産生が増加する (Safftig et al., *Eur. Arch. Psychiatr. Clin. Neurosci.*, 1999, 249:271-79)。プレセニリンは、ガンマーセクレターゼ (APPを切断する酵素) に影響を及ぼすことによりAPPプロセッシングを制御すると仮定されている (Cruts et al., 1998, *Hum. Mutat.*, 11:183-190, Haass et al., *Science*, 1999, 286:916-19)。また、プレセニリンはこれらがガンマーセクレターゼ活性を直接制御するかまたはそれ自身がプロテアーゼ酵素であるように、ノッチレセプターの切断に関与すると考えられている (Destrooper et al., *Nature*, 1999, 398:518-22)。PSEN2の2つの選択的転写産物が同定されている (Sato et al., 1999, *J. Neurochem.* 72(6):2498-505)。PS1遺伝子中の点突然変異により、アミロイド前駆体蛋白質 (APP) の蛋白質分解プロセッシングによるアミロイド発生性ペプチドアミロイドペプタ (1-42) の産生が選択的に増加する (Lemere et al., 1996, *Nat. Med.* 2(10):1146-50)。正常なAP

40

50

PプロセッシングにおけるPS1の役割の可能性は、Destrooperら(Nature 391:387390, 1998)によりPS1-欠損マウス胚に由来する神経培養物において研究された。彼らは、APPの細胞外ドメインのアルファ-およびベータ-セクレターゼによる切断はPS1の非存在によって影響されないが、APPの貫膜ドメインのガンマ-セクレターゼによる切断は妨害され、このことによりAPPのC-末端フラグメントが蓄積し、アミロイドペプチドの産生が5倍低下したことを見いだした。パルスチェイス実験は、PS1欠損はAPPの膜付随フラグメントのターンオーバーを特異的に減少させることを示した。すなわち、PS1は、APPの完全な膜ドメインを切断する蛋白質分解活性を促進するようである。著者によれば、この結果は、臨床的に明白なPS1における変異は機能の獲得を引き起こし、PS1活性の阻害はアルツハイマー病における抗アミロイド発生性治療の潜在的標的であることを示す。

10

【0145】

ベータ-セクレターゼ

ベータ-セクレターゼは、脳において特異的に発現されており、アルツハイマー病に伴うアミロイド前駆体蛋白質(APP)の蛋白質分解プロセッシングを担う(Potter et al., 2000, Nat. Biotechnol. 18(2):125-26)。これはAPPのベータ-ペプチド配列のアミノ末端の残基671と672の間で切断し、ベータ-切断型可溶性APPを生成および細胞外放出させ、後にガンマ-セクレターゼによりカルボキシ末端フラグメントを放出させる(Kimberly et al., 2000, J. Biol. Chem. 275(5):3173-78)。Yanら(Nature, 1999402:533-37)は、ベータ-セクレターゼ活性を有する新規な膜結合アスパルチルプロテアーゼ(Asp2)を同定した。Asp2遺伝子は、脳および他の組織において広く発現されている。細胞におけるAsp2の発現の減少は、アミロイドベータ-ペプチド産生を低下させ、ベータ-セクレターゼ切断により生成するカルボキシ末端APPフラグメントの蓄積を妨害する。Asp2は、アミロイドベータ-ペプチドペプチドの産生およびその結果としてのアルツハイマー病におけるアミロイドプラークの形成を妨害するよう設計される薬剤の新規な蛋白質標的である。

20

【0146】

ヒト胎盤形成に参与する2つのアスパルチルプロテアーゼが最近単離された：脱落膜アスパルチルプロテアーゼ(DAP-1)、およびDAP-2(Moses et al., Mol. Hum. Repro., 1999, 5:983-89)。

30

【0147】

アスパルチルペプチダーゼファミリーの別のメンバーは、ヒト免疫不全ウイルスタイプ1のHIV-1レトロペプシンである。この酵素はウイルスポリ蛋白質のプロセッシングおよび成熟ビリオンの成熟に重要である。

【0148】

2. システインプロテアーゼ

体内で広範な機能を行う別のクラスのプロテアーゼはシステインプロテアーゼである。これらは、特に、前駆体蛋白質のプロセッシング、および処分とマークされた蛋白質のユビキチン経路による細胞内分解に作用する。真核生物システインプロテアーゼは、活性部位システインを含む蛋白質分解酵素のファミリーである。触媒作用はチオエステル中間体を通じて進行し、近くのヒスチジン側鎖により促進される。アスパラギンにより必須の触媒的三者系が完成する。疾病において重要な役割を果たすこのファミリーのペプチダーゼには、カスパーゼ、カルパイン、ヘッジホッグ、およびユビキチンヒドロラーゼが含まれる。

40

【0149】

システインプロテアーゼは、免疫系(マクロファージ、単球等)の細胞などの多数の細胞により産生される。これらの免疫細胞は炎症の部位に移動し分子を分泌することにより、部分的には身体における保護的役割を果たす。分泌分子の例はシステインプロテアーゼである。

【0150】

50

ある条件下では、免疫系のシステインプロテアーゼの不適切な制御は、自己免疫疾病、例えば慢性関節リウマチにつながりうる。例えば、システインプロテアーゼであるカテプシンCの過剰分泌は、エラスチン、コラーゲン、ラミニンおよび骨において見いだされる他の構造蛋白質の分解を引き起こす。この不適切な消化が生じた骨は、転移をより受けやすい。

【0151】

カスパーゼ (C14) - アポトーシス

プロテアーゼ反応のカスケードは、プログラムされた細胞死をうける哺乳動物細胞において観察されるアポトーシス性の変化の原因であると考えられている。このカスケードには、カスパーゼファミリーのアスパラギン酸特異的システインプロテアーゼの多くのメンバー、例えばカスパーゼ2, 3, 6, 7, 8, および10が関与する (Salvesen and Dixit, Cell, 1997, 91: 443-446)。細胞毒性化学療法または正常な細胞生存シグナルの欠失により生ずるアポトーシス性シグナルを回避した (転移細胞におけるように) 癌細胞は、触診しうる腫瘍まで発達することができる。

10

【0152】

他のカスパーゼはまた、プロ炎症性サイトカインの活性化にも関与する。カスパーゼ1は、IL-1 β およびIL-18 (インターフェロン- γ -誘導因子) の前駆体の特異的にプロセシングする (Salvesen and Dixit, Cell, 1997)。

【0153】

カルパイン (C2) - 軸索死, ジストロフィー

カルシウム依存性システインプロテアーゼは、集合的にカルパインと称され、哺乳動物細胞に広く分布している (Wang, 2000, Trends Neurosci. 23 (1): 20-26)。カルパインは非リソソーム性細胞内システインプロテアーゼである。哺乳動物カルパインには、2種類の偏在性蛋白質, CAPN1およびCAPN2, ならびに2種類の胃特異的蛋白質, および筋肉特異的なCAPN3が含まれる (Herasse et al., 1999, Mol. Cell. Biol. 19 (6): 4047-55)。偏在性酵素は、共通の小サブユニットを伴う別個の大サブユニットを有するヘテロダイマーから構成され、これらはすべて異なる遺伝子によりコードされている。カルパインの大サブユニットは4個のドメインに分けることができる。ドメインIおよびIIIは、その機能が不明であり、既知の蛋白質と相同性を示さない。しかし、前者は、蛋白質分解活性の制御に重要であろう。ドメインIIは他のシステインプロテアーゼとの類似性を示し、共通してその活性部位にヒスチジン, システイン, およびアスパラギン残基を含む。ドメインIVはカルモジュリン様である。CAPN5およびCAPN6は、カルモジュリン様ドメインIVを欠失している点において、先に同定された脊椎動物カルパインと異なる (Ohno et al., 1990, Cytogenet. Cell Genet. 53 (4): 225-29)。

20

30

【0154】

CAPN3遺伝子における変異は肢帯筋ジストロフィー, タイプ2A (LGMD2A) と関連づけられてきた (Allamand et al., 1995, Hum. Molec. Genet. 4: 459-463)。この疾病に伴うゆっくり進行する筋肉柔弱は、通常は下肢帯に最初に生じ、次に上肢に広がるが、顔面筋肉ではゆるやかである。カルパインはまたアルツハイマー病に伴う神経細胞死につながる過剰活性Cdk5の発達における関与が示唆されている (Patrick et al., 1999, Nature 402: 615-622)。

40

【0155】

ヘッジホッグ (C46) - 癌

発生している胚の構成および形態は、一連の誘導性相互作用により達成される。脊椎動物遺伝子の1つのファミリーが胚形成の間の誘導性シグナルをコードするショウジョウバエ遺伝子 'ヘッジホッグ' (hh) と関連すると記述されている (Johnson and Tabin, 1997, Cell 90: 979-990)。

50

ウジョウバエの発生の際のいくつかの時点において細胞の運命の確立に關与する分泌蛋白質をコードする (Marigo et al., 1995, Genomics 28: 44-51)。3つの既知のhhの哺乳動物ホモログが存在する: ソニックヘッジホッグ (Shh), インディアンヘッジホッグ (Ihh), およびデザートヘッジホッグ (Dhh) (Johnson and Tabin, 1997, Cell 90: 979-990)。Shhは, そのショウジョウバエの同族体と同様に, 初期胚のパターニングを助けるシグナルをコードする。これは, ヘンゼン結節, 神経管の床板, 初期腸内胚葉, 肢芽の後方, および脊索全体において発現されている (Chiang et al., 1996, Nature 383: 407-413)。これは腹側神経管, 前方-後方肢軸, および腹側体節のパターニングにおける鍵となる誘導性シグナルであると示唆されている。Oroら ("Basal cell carcinomas in mice overexpressing sonic hedgehog," Science 276: 817-821, 1997) は, 皮膚においてSHHを過剰発現するトランスジェニックマウスが基底細胞母斑症候群の多くの特徴を発達させることを示し, このことは, SHHがマウスにおいて基底細胞癌 (BCC) を誘導するのに十分であることを示す。データは, SHHがヒト腫瘍発生において役割を有するかもしれないことを示唆する。SHHまたは他の 'ヘッジホッグ' 遺伝子の活性化変異はヒトにおけるBCC形成の別の経路であるかもしれない。ヒト変異his133tyr (マウスにおいてはhis134tyr) が候補である。これは, 全前脳症を有する個体について報告されている機能変異喪失とは異なるものである (Oro et al., 1997, Science 276: 817-821)。His133は, 触媒部位において触媒に必要であると考えられているの1つである保存残基であるhis134と隣接して存在する。SHHは, 多数のヒト腫瘍における優性オンコジンであって, 反対の 'パッチド' (PTCH) 遺伝子の腫瘍サプレッサー活性の鏡であるかもしれない (Aszterbaum et al., 1998, J Invest. Derm. 110: 885-888)。マウスにおけるShh誘導性腫瘍の急速かつ頻繁な出現は, SHH-PTC経路の破壊がBCCの形成に十分であることを示唆した。

【0156】

脊椎動物ヘッジホッグファミリーのメンバー (ソニック, インディアンおよびデザート) は, 種々の臓器系, 例えば, 神経, 体節, 肢, 骨格, および雄においては生殖腺形態発生の発達に必須であることが示されている。デザートヘッジホッグは発生中の網膜において発現されるが, インディアンヘッジホッグ (Ihh) は, 胚第13日から開始して発生中のおよび成熟した網膜色素沈着上皮において発現される (Levine et al., J Neurosci., 1997, 17 (16): 6277-88)。Dhhはまた, 精子形成の制御において役割を有することが示唆されている。セルトリー細胞前駆体はSry (性決定遺伝子) を発現し, これは哺乳動物における精巢の発達につながる。セルトリー細胞前駆体におけるDhh発現は, Sryの活性化の後にまもなく開始され, 成人になるまで精巢中で持続する。Bitgoodら (Curr. Biol., 1996, 6 (3): 298-304) は, Dhh-null変異のホモ接合を有する雌マウスは明らかな表現型を示さないが, 雄は生存可能であるが成熟精子を完全に欠失しており生殖不能であることを開示し, Dhhシグナリングが哺乳動物精子形成の制御に必須の役割を果たすことを示している。Dhhはまた, 保護的神経鞘神経内膜, 神経周膜および神経上膜の維持に役割を有することが見いだされている。Dhhノックアウトマウスにおいては, 成人神経の結合組織鞘は, 電子顕微鏡によれば非常に異常であるように見える。Mirskyら (Ann. N. Y. Acad. Sci., 1999, 883: 196-202) は, シュワン細胞から間葉へのDhhシグナリングが形態学的かつ機能的に正常な神経周膜の形成に關与することを示す。

【0157】

胚形成および器官形成の際の発達生物学および分子生物学の最近の進歩は, 骨形成のメカニズムに新たな洞察を提供した。Iwasakiら (J. Bone Joint Sur 50

g Br., 1999, 81 (6): 1076-82) は、インディアンヘッジホッグ (Ihh) は軟骨細胞前駆体で、後に成熟した肥大性軟骨細胞で発現されることを示す。Ihh は、脊椎動物骨格の形態形成において重要な役割を果たす。Beckerら (Dev. Biol., 1997, 187 (2): 298-310) は、Ihh がマウスの初期胚形成の間の胚外の内胚葉の分化の媒介にも関与していることを示唆するデータを提供する。軟骨細胞増殖の減少および大きい肥大を有する短肢小人症は、Ihh の標的化された欠失の結果である (Karp et al., 2000, Development 127 (3): 543-48)。Ihh mRNA および蛋白質の発現は、F9 細胞がレチノイン酸に応答して外側内臓外胚葉層を含有する壁側内胚葉または胚様体のいずれかに分化するにつれて劇的に非制御となる。胚盤胞生長培養物の RT-PCR 分析は、Ihh メッセージは胚盤胞にはほとんどまたは全く存在しないが、培養すると後日壁側内胚葉細胞の出現と一致して有意なレベルで現れることを示す。 10

【0158】

ユビキチンヒドロラーゼ (C12) - アポトーシス, チェックポイントの完全性 本発明においては、14 の遺伝子、すなわち、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、配列番号17、配列番号18 はユビキチンヒドロラーゼファミリーに属する。これらの遺伝子によりコードされるポリペプチドは、以下の活性の1またはそれ以上を有するかもしれない。 20

【0159】

ユビキチンカルボキシ末端ヒドロラーゼ (3. 1. 2. 15) (脱ユビキチン化酵素) は、ユビキチンのC末端のグリシンのペプチド結合を認識しこれを加水分解するチオールプロテアーゼである。これらの酵素は、ポリユビキチン前駆体のプロセッシングならびにユビキチン化蛋白質のプロセッシングに関与する。真核生物細胞においては、ユビキチンの蛋白質への共有結合は、種々の細胞プロセスにおいて役割を果たす。多くの場合、ユビキチン化により26Sプロテオソームによる蛋白質分解が生ずる。蛋白質のユビキチン化は可逆的であり、ユビキチンの除去は、脱ユビキチン化酵素、すなわちDUBにより触媒される。ユビキチン化蛋白質からのユビキチンの除去を触媒するこれらの酵素の欠損は、神経変性性疾患、例えばアルツハイマー病、パーキンソン病、進行性核上麻痺、およびピック病およびカフ病に特徴的でありうる。 30

【0160】

パpain (C1) - カテプシンK, SおよびB, - 骨再吸収, Agプロセッシング (Prosite PS00139)

本発明においては、1つの遺伝子、配列番号4がパpainファミリーに属する。この遺伝子によりコードされるポリペプチド遺伝子は、以下の1またはそれ以上の活性を有するかもしれない。

【0161】

カテプシンKは、ペプチダーゼのパpainファミリーのメンバーであり、破骨再吸収に関与する。これは、細胞外分解において重要な役割を果たし、骨リモデリングの疾患、例えばピノコジストーシス (骨硬化症および低い身長を特徴とする常染色体劣性骨軟骨異形成症) において役割を有するかもしれない。主要組織適合性複合体 (MHC) クラスII分子による抗原提示には、異なるプロテアーゼがエンドサイトーシス経路に関与して、エンドサイトーシスされた抗原ならびにMHCクラスII付随不変鎖を分解することが必要である。カテプシンS (パpainファミリーのメンバー) のみが、不変鎖の完全な破壊に必須であるようである。カテプシンBは、肺、前立腺、結腸、乳、および胃の腫瘍において過剰発現されている。Hughesら (Proc. Nat. Acad. Sci. 95: 12410-12415, 1998) は、8p23-p22に食道腺癌においてカテプシンBを過剰発現させるアンプリコンを見いだした。免疫組織科学的分析により、40の食道腺癌標本のうち29 (72.5%) でCTS B蛋白質の大量の細胞外発現が見いだされた。この知見は、食道腺癌および潜在的には他の腫瘍におけるCTS Bの重要な役割を裏付 50

けると考えられる。

【0162】

カテプシンBはリソソームプロテアーゼであり、種々の癌（乳、肺腺癌）における予測マーカーとして研究されている。

【0163】

システインプロテアーゼAEP

システインプロテアーゼAEPは、免疫機能において別の役割を果たす。これは、B細胞における抗原プロセッシングに必要なプロテアーゼ段階における関与が示唆されている（Manoury et al. Nature 396:695-699 (1998)）。

【0164】

A型肝炎ウイルスプロテアーゼ（C3E）

A型肝炎ウイルスゲノムは、インビボで酵素的に切断して成熟蛋白質を生ずるのに必要なシステインプロテアーゼをコードする（Wang, 1999, Prog. Drug Res. 52:197219）。この酵素および他のウイルスにおけるそのホモログ（例えばE型肝炎ウイルス）は、化学療法的介入の潜在的標的である。

【0165】

3. メタロプロテアーゼ

コラゲナーゼ（M10）－侵襲

本発明においては、配列番号19および配列番号20の2つの遺伝子がM10ファミリーのメンバーである。これらの遺伝子によりコードされるポリペプチドは、以下の活性の1またはそれ以上を有するかもしれない。

【0166】

マトリクス分解は癌が広がるために必須の段階である。72kDおよび92kDのタイプI Vコラゲナーゼは、分泌される垂鉛メタロプロテアーゼのグループのメンバーであり、哺乳動物において細胞外マトリクスのコラーゲンを分解する。このグループの他のメンバーには、腸コラゲナーゼおよびストロメリシンが含まれる（Nagase et al., 1992, Matrix Suppl. 1:421-424）。Vura（Cell, 1998, 934:11-22）は、胚性幹細胞における標的化した破壊により、MMP9／ゼラチナーゼB遺伝子にnull変異を有するホモ接合マウスを作成した。これらのマウスは骨格成長板血管新生および骨形成の異常なパターンを示した。培養MMP9-nullマウスからの成長板は新脈管形成アクチベータの放出が遅いことが示され、このことは、新脈管形成の制御におけるこのプロテイナーゼの役割を確立する。MMP2（ゼラチナーゼA）はヒト癌の攻撃性と関連づけられている（Chenard et al., 1999, Int. J. Cancer, 82:208-12）。基底細胞癌（BCC）とより攻撃的な扁平上皮癌（SCC）とを比較した研究においては、MMP2およびMMP9の両方ともSCCにおいてより高いレベルで発現されていた（Dumas et al., 1999, Anticancer Res., 19(4B):2929-38）。さらに、Tリンパ球におけるMMP2およびMMP9の発現がRas／MAPキナーゼシグナリング経路により調節されることが最近示された（Esparza et al., 1999, Blood, 94:2754-66）（Li et al., 1998, Biochim. Biophys. Acta, 1405:110-20も参照）。

【0167】

ADAM（M12）－TNF，炎症，成長因子プロセッシング

ADAMペプチダーゼはジスインテグリンおよびメタロプロテイナーゼ（ADAM）ドメインを含む蛋白質のファミリーである（Werb and Yan, Science, 1998, 282:1279-1280）。このファミリーのメンバーは、独特の構造を有する細胞表面蛋白質であり、潜在的接着およびプロテアーゼドメインの両方を有する（Primakoff and Myles, Trends in Genet., 2000, 16:83-87）。これらのプロテアーゼの活性は、TNF，炎症，および／または成長因子プロセッシングと関連づけることができる。

10

20

30

40

50

【0168】

A D A Mプロテアーゼは、プローおよびメタロプロテイナーゼドメイン、ジスインテグリンドメイン、システインリッチ領域およびE G F反復を有することを特徴とする（B l o b e l, 1997, C e l l, 90:589-592, 図面および表を含めその全体を本明細書の一部としてここに引用する）。これは、多くの蛋白質、例えば腫瘍壊死因子- α （T N F- α ）、k i t-リガンド、T G F α 、F a s-リガンド、サイトカインレセプター、例えばI L-6レセプターおよびN G Fレセプター、ならびに接着蛋白質、例えばL-セレクチン、および β アミロイド前駆体蛋白質の形質膜からの放出に関連づけられている（B l o b e l, 1997, C e l l, 90:589-592）。

【0169】

腫瘍壊死因子- α は、233アミノ酸の前駆体からプロ炎症性サイトカインとして合成される。膜結合前駆体の分泌された成熟蛋白質への変換は、T N F- α コンベルターゼと称されるプロテアーゼにより媒介される。T N F- α は種々の疾病に関与している。A D A M 17は、ジスインテグリンおよびメタロプロテイナーゼドメインを含み、'腫瘍壊死因子- α 変換酵素'（T A C E）とも称される（B l a c k e t a l., N a t u r e, 1997, 385:729-33）。遺伝子はA D A Mファミリーの特徴、すなわち分泌シグナル配列、ジスインテグリンドメイン、およびメタロプロテイナーゼドメインを含む824アミノ酸のポリペプチドをコードする。発現実験は、コードされる蛋白質が前駆体腫瘍壊死因子- α を切断して成熟型にすることを示した。この酵素はまた、トランスフォーミング成長因子- α （T G F- α ）のプロセッシングにおいても役割を果たすかもしれない。この遺伝子を欠失したマウスはT G F- α を欠失したものと類似する表現型を示すためである（P e s c h o n e t a l., S c i e n c e, 282:1281-1284）。

【0170】

ネプリリシン（M13）-エンドセリン-変換酵素

本発明においては、1つの遺伝子、配列番号21がこのファミリーのメンバーである。この遺伝子によりコードされるポリペプチドは、以下の活性の1またはそれ以上を有するかもしれない。

【0171】

ネプリリシンは、エンケファリンおよび他の生物活性ペプチドを分解する作用をするメタロペプチダーゼであり、高血圧および腎臓疾病の薬剤の標的である（O e f n e r, e t a l., J M o l. B i o l., 2000, 296:341-49）。

【0172】

カルボキシペプチダーゼ（M14）-神経伝達物質プロセッシング

本発明においては、3つの遺伝子、配列番号1、配列番号2、および配列番号3がZ nカルボキシペプチダーゼである。これらの遺伝子によりコードされるポリペプチドは、以下の活性の1またはそれ以上を有するかもしれない。

【0173】

カルボキシペプチダーゼはC O O H-末端塩基性アミノ酸（アルギニンまたはリシン）を特異的に除去する。これらは、多くの生物学的プロセス、例えば、ペプチドホルモン活性の活性化、不活性化もしくは調節、神経伝達物質プロセッシング、および蛋白質および酵素の物理学的特性の変更において重要な機能を有する。

【0174】

ジペプチダーゼ（M2）-A C E

本発明においては、1つのプロテアーゼ、配列番号22がM2ファミリーのメンバーである。この遺伝子によりコードされるポリペプチドは、以下の活性の1またはそれ以上を有するかもしれない。

【0175】

アンジオテンシンI変換酵素（E C 3. 4. 15. 1）、すなわちキニナーゼI Iは、アンジオテンシンIを加水分解してアンジオテンシンI I（強力な血管収縮剤でありアルド

10

20

30

40

50

ステロン刺激ペプチドである)とすることにより, 血圧調節および電解質平衡において重要な役割を果たすジペプチジルカルボキシペプチダーゼである。この酵素はまた, ブラジキニン(強力な血管拡張剤)を不活性化させることができる。アンジオテンシン変換酵素は主としてその血圧調節における役割に関して研究されてきたが, この広く分布している酵素は多くの他の生理学的機能を有する。ACEには2つの型, すなわち精巣特異的アイソザイムおよび2つの活性中心を有する体細胞アイソザイムがある。

【0176】

マトリクスメタロプロテアーゼ(M10B)ー組織リモデリングおよび炎症

マトリクスメタロプロテアーゼ(MMP)は, 発生および炎症の間の組織リモデリングおよび修復に重要な関連するマトリクス分解酵素のファミリーである。異常な発現は, 種々の疾病, 例えば腫瘍侵襲性, 関節炎, およびアテローム性動脈硬化症と関連している。MMP活性はまた, タバコ誘導性肺気腫とも関連している。 10

【0177】

マトリクスメタロプロテアーゼ(MMP)は, 発生および炎症の間の組織リモデリングおよび修復に重要な関連するマトリクス分解酵素のファミリーである(Belotti et al., 1999, Int. J. Biol. Markers 14(4):232-38)。異常な発現は, 種々の疾病, 例えば腫瘍侵襲性(Johansson and Kahari, 2000, Histol. Histopathol. 15(1):225-37), 関節炎(Malemud et al., 1999, Front. Biosci. 4:D762-71), およびアテローム性動脈硬化症(Nagase, 1997, Biol. Chem. 378(3-4):151-60)と関連している。MMP活性はまた, タバコ誘導性肺気腫とも関連している(Dhami et al., Am. J. Respir. Cell Mol. Biol., 2000, 22:244-52)。 20

【0178】

SREBPプロテアーゼ(M50)

ステロール制御要素結合蛋白質プロテアーゼは, 膜内蛋白質分解およびステロール制御性結合蛋白質(SREBP)の放出において機能する(Duncan et al., 1997, J. Biol. Chem. 272:12778-85)。SREBPはコレステロールおよび脂肪酸代謝の遺伝子を活性化し, このためSREBPプロテアーゼは治療的調節の魅力的な標的である(Brown et al., 1997, Cell 89:331-340)。 30

【0179】

成長因子のメタロプロテアーゼプロセッシング

上述のTGF- α のプロセッシングに加えて, メタロプロテアーゼは, 他の成長因子, 例えば, ヘパリン結合EGF(proHB-EGF)(Izumi et al., EMBO J, 1998, 17:7260-72), およびアンフィレギュリン(Brown et al., 1998, J. Biol. Chem., 27:17258-68)の前駆体のプロセッシングにおいて作用することが直接示されている。

【0180】

さらに, 最近, メタロプロテアーゼが, GPCR経路の刺激によりMAPキナーゼ経路が刺激されるという伝達を助けることが示された(Prenzel et al., 1999, Nature, 402:884-888)。経路における成長因子中間体であるHB-EGFは, 特性決定されていないメタロプロテアーゼが関与するGPCR経路により制御される蛋白質分解工程において細胞により放出される。放出後, HB-EGFは細胞外マトリクスにより結合され, 次に表面上のEGFレセプターに対して提示され, その結果MAPキナーゼ経路が活性化される(Prenzel et al., 1999, Nature, 402:884-888)。 40

【0181】

Gallea-Robacheら(1997)による最近の研究では, 異なる基質特異性を示すメタロプロテアーゼファミリーが, 他の成長因子, 例えばマクロファージコロニー 50

刺激因子 (M-CSF) および幹細胞因子 (SCF) の発散 (shedding) に関与することが示唆されている (Gallean-Robache et al., 1997, Cytokine 9:340-46)。M-CSF (CSF-1としても知られる) の発散は、ホルボルエステルによる蛋白質キナーゼCの活性化と関連づけられている (Stein et al., 1991, Oncogene, 6:601-05)。

【0182】

4. セリンプロテアーゼ

セリンプロテアーゼは、トリプシン、カリクレイン、キモトリプシン、エラスターゼ、トロンビン、組織プラスミノゲンアクチベータ (tPA)、ウロキナーゼプラスミノゲンアクチベータ (uPA)、プラスミン (Werb, Cell, 1997, 91:439-442), カリクレイン (Clements, Biol. Res., 1998, 31:51-59), およびカテプシンG (Shamamian et al., Surgery, 2000, 127:142-47) を含むクラスである。これらのプロテアーゼは、一般に、その活性部位において、ヒスチジン-57, アスパラギン酸-102, およびセリン-195 (キモトリプシンの番号付けシステムを用いる) からなる、よく保存されたアミノ酸残基の触媒三者系を有する。セリンプロテアーゼ活性は凝固と関連づけられており、腫瘍マーカーとしての用途を有するかもしれない。

【0183】

セリンプロテアーゼは基質の特異性によりさらに細分類することができる。エラスターゼは、小さい脂肪族残基に、例えばバリンに隣接する基質を好んで切断し、カイメースは大きい芳香族疎水性残基の近くを好んで切断し、トリプターゼは正に荷電した残基を好む。プロリンの隣りを好んで切断するセリンプロテアーゼの別のクラスが最近記載されている。このプロリルエンドペプチダーゼは、アルツハイマー患者における記憶喪失の進行への関与が示唆されている (Toide et al., 1998, Rev. Neurosci. 9(1):17-29)。

【0184】

この大きくかつ重要なファミリーに属することが知られているプロテアーゼの部分的リストには以下のものが含まれる: 血液凝固因子VII, IX, X, XIおよびXII; トロンビン; プラスミノゲン; 補体成分C1r, C1s, C2; 補体因子B, DおよびI; RA反応性因子の補体活性化成分; エラスターゼ1, 2, 3A, 3B (プロテアーゼE); 肝臓成長因子アクチベータ; 腺 (組織) カリクレイン, 例えばEGF結合蛋白質タイプA, B, およびC; NGF- γ 鎖, γ -レニン, および前立腺特異的抗原 (PSA); 血漿カリクレイン; マスト細胞プロテアーゼ; ミエロプラスチン (プロテイナーゼ3) (ウェゲナー自己抗原); プラスミノゲンアクチベータ (ウロキナーゼタイプ, および組織タイプ); およびトリプシンI, II, III, およびIV。これらのペプチダーゼは、凝固, 腫瘍発生, 血圧の調節, 成長因子の放出, および他の役割において鍵となる役割を果たす (<http://www.babraham.co.uk/Merops/Merops.htm>)。

【0185】

本発明のトリプシンファミリーのプロテアーゼには, SGPr434, 配列番号24; SGPr446_1, 配列番号25; SGPr447, 配列番号26; SGPr432_1, 配列番号27; SGPr529, 配列番号28; SGPr428_1, 配列番号29; SGPr425, 配列番号30; SGPr548, 配列番号31; SGPr396, 配列番号32; SGPr426, 配列番号33; SGPr552, 配列番号34; SGPr405, 配列番号35; SGPr485_1, 配列番号36; SGPr534, 配列番号37; SGPr390, 配列番号38; SGPr521, 配列番号39; SGPr530_1, 配列番号40; SGPr520, 配列番号41; SGPr455, 配列番号42; SGPr507_2, 配列番号43; SGPr559, 配列番号44; SGPr567_1, 配列番号45; SGPr479_1, 配列番号46; SGPr489_1, 配列番号47; SGPr465_1, 配列番号48; SGPr524_1, 配列番号49; SGPr

422, 配列番号50; SGPr538, 配列番号51; SGPr527_1, 配列番号52; SGPr542, 配列番号53; SGPr551, 配列番号54; SGPr451, 配列番号55; SGPr452_1, 配列番号56; SGPr504, 配列番号57; SGPr469, 配列番号58; SGPr400, 配列番号59が含まれる。配列番号23はズブチラーゼサブファミリーのセリンプロテアーゼである。ほとんどの大きな蛋白質前駆体の限定された蛋白質分解は, インビボでズブチリシン様プロ蛋白質コンベルターゼにおいて行われる。多くの重要な生物学的プロセス, 例えば, ペプチドホルモン合成, ウイルス蛋白質プロセッシングおよびレセプター成熟にはこれらの酵素による蛋白質分解性プロセッシングが関与しており, このため, これらは新規治療剤を開発するための潜在的な標的である (Bergeron F, J Mol Endocrinol 2000 Feb 10; 24(1): 1-22)。

【0186】

5. トレオニンペプチダーゼ (T1) - (Prosite PDOC00326/PDOC00668)

プロテオソームサブユニット (T1A)

プロテオソームは, 細胞性基質のATP/ユビキチン依存性非リソソーム蛋白質分解に関与する多触媒性トレオニンプロテイナーゼ複合体である。これは異常な構造を有する蛋白質, ならびに代謝制御および細胞サイクル進行に関連する天然に生ずる寿命の短い蛋白質の選択的排除を担う (Momand et al., 2000, Gene 242(1-2): 15-29, Bochtler et al., 1999, *Adv. Biochem. Biophys. Biomol. Struct.* 28: 295-317)。プロテオソーム阻害剤であるラクタシスチンはヒト内皮細胞の増殖を可逆的に阻害し, このことは, プロテオソームが新脈管形成において役割を果たすことを示唆する (Kumeda, et al., *Anticancer Res.* 1999 Sep-Oct; 19(5B): 3961-8)。高等脊椎動物におけるプロテオソームの別の重要な機能は, MHCクラス1分子上で循環リンパ球に提示されるペプチドを生成することである (Castelli et al., 1997, *Int. J. Clin. Lab. Res.* 27(2): 103-10)。プロテオソームは沈降係数26Sを有し, 20S触媒コアおよび22S制御複合体から構成される。真核生物20Sプロテオソームは, 700-800kDの分子量を有し, 15種類以上の21-32kDのポリペプチドの組から構成される。すべての真核生物20Sプロテオソームサブユニットは, アークエスバクテリア *Thermoplasma acidophilum* の α または β サブユニットのいずれかに対する高い類似性に基づいて, おおまかに α および β の2つのサブファミリーに分類することができる (Mayr et al., 1999, *Biol. Chef.* 380(10): 1183-92)。成分のいくつかはトレオニンペプチダーゼであると同定されており, このことは, このクラスのペプチダーゼが種々の機能の中でも特に代謝経路および細胞サイクル進行の制御に鍵となる役割を果たすことを示唆する (Yorgin et al., 2000, *J. Immunol.* 164(6): 2915-23)。

【0187】

6. 触媒作用メカニズムが不明のペプチダーゼ

Rasプロトオンコジンおよび他のプレニル化蛋白質の翻訳後プロセッシングを担うプレニル蛋白質特異的プロテアーゼはこのクラスに分類される。このクラスにはまた, 哺乳動物感染において役割を果たすかもしれないいくつかのウイルスペプチダーゼ, 例えば, カルジオウイルスエンドペプチダーゼ2A (脳心筋炎ウイルス) (Molla et al., 1993, *J. Virol.* 67(8): 4688-95), NS2-3プロテアーゼ (C型肝炎ウイルス) (Bright et al., 1998, *Antivir. Ther.* 3(Suppl 3): 71-81), エンドペプチダーゼ (感染性脾臓壊死ウイルス) (Lejal et al., *J. Gen. Virol.*, 2000, 81: 983-992), およびNproエンドペプチダーゼ (豚コレラウイルス) (Tratsch et al., 1998, *J. Virol.* 72(9): 7681-84) が含ま

れる。

【0188】

核酸プローブ、プロテアーゼを検出するための方法およびキット

本発明の核酸プローブを用いて、通常のハイブリダイゼーション方法により適当な染色体またはcDNAライブラリを探索して、本発明の他の核酸分子を得ることができる。染色体DNAまたはcDNAライブラリは、当該技術分野において認識されている方法にしたがって、適当な細胞から調製することができる（"Molecular Cloning: A Laboratory Manual", 第2版. Cold Spring Harbor Laboratory, Sambrook, Fritsch, & Maniatis, eds., 1989を参照）。 10

【0189】

別の方法においては、目的とするポリペプチドのアミノ酸配列のN末端、およびC末端部分に対応するヌクレオチド配列を有する核酸プローブを得るために化学合成を行うことができる。合成した核酸プローブを、ポリメラーゼ連鎖反応（PCR）においてプライマーとして使用して、認識されているPCR手法にしたがって、本質的にPCR Protocols, "A Guide to Methods and Applications", Academic Press, Michael, et al., eds., 1990にしたがって、適当な染色体またはcDNAライブラリを用いて、本発明のフラグメントを得ることができる。 20

【0190】

当業者は、本明細書に開示される配列に基づいて、当該技術分野において知られるコンピュータアラインメントおよび配列分析の方法を用いて、そのようなプローブを容易に設計することができる（"Molecular Cloning: A Laboratory Manual", 1989, 上掲）。本発明のハイブリダイゼーションプローブは、標準的な標識技術、例えば放射性標識、酵素標識、蛍光標識、ビオチン-アビジン標識、化学発光等を用いる技術により、標識することができる。ハイブリダイゼーション後、プローブは既知の方法を用いて可視化することができる。 30

【0191】

本発明の核酸プローブはRNAならびにDNAプローブを含み、そのようなプローブは、例えば当該技術分野において知られる技術を用いて生成される。核酸プローブは、固体支持体上に固定化してもよい。そのような固体支持体の例としては、限定されないが、プラスチック、例えばポリカーボネート、複合炭水化物、例えばアガロースおよびセファロース、およびアクリル樹脂、例えばポリアクリルアミドおよびラテックスビーズが含まれる。核酸プローブをそのような固体支持体に結合させる技術は当該技術分野においてよく知られている。 40

【0192】

本発明の核酸探索方法に適した試験試料には、例えば、細胞または細胞の核酸抽出物、または体液が含まれる。上述の方法において用いられる試料は、アッセイフォーマット、検出方法、およびアッセイすべき組織、細胞、または抽出物の性質により様々であろう。細胞の核酸抽出物を調製する方法は当該技術分野においてよく知られており、用いる方法に適合する試料を得るために容易に適合させることができる。 50

【0193】

試料中で本発明の核酸の存在を検出する1つの方法は、（a）前記試料をハイブリダイゼーションが生ずるような条件下で上述の核酸プローブと接触させ、そして（b）前記核酸分子に結合した前記プローブの存在を検出する、ことを含む。当業者は、当該技術分野において知られる技術にしたがって、上述のように核酸プローブを選択するであろう。試験すべき試料には、限定されないが、ヒト組織のRNA試料が含まれる。

【0194】

試料中で本発明の核酸の存在を検出するためのキットは、その中に上述の核酸プローブが置かれている少なくとも1つの容器手段を含む。キットはさらに、以下の1またはそれ以 50

上を含む他の容器を含んでいてもよい：洗浄試薬および結合した核酸プローブの存在を検出することができる試薬。検出試薬の例としては、限定されないが、放射性標識プローブ、酵素標識プローブ（西洋ワサビペルオキシダーゼ、アルカリホスファターゼ）、および親和性標識プローブ（ビオチン、アビジン、またはストレプトアビジン）が挙げられる。好ましくは、キットはさらに使用の指針を含む。

【0195】

詳細には、コンパートメント化されたキットには、試薬が別々の容器に入れられている任意のキットが含まれる。そのような容器には、小ガラス容器、プラスチック容器またはプラスチックまたは紙のストリップが含まれる。そのような容器は、試料および試薬が互いに夾雑しないように、かつ各容器の薬剤または溶液を定量的様式で1つのコンパートメントから別のコンパートメントに加えることができるように、1つのコンパートメントから別のコンパートメントに試薬を効率よく移すことができるものである。そのような容器には、試験試料を入れる容器、アッセイにおいて用いるプローブまたはプライマーを含有する容器、洗浄試薬（例えばリン酸緩衝化食塩水、T r i s緩衝液等）を含有する容器、およびハイブリダイズしたプローブ、結合した抗体、増幅産物等を検出するために用いられる試薬を含有する容器が含まれる。当業者は、本発明において記載される核酸プローブを、当該技術分野においてよく知られる確立されたキットフォーマットの1つに容易に組み込むことができることを容易に認識するであろう。

10

【0196】

プロテアーゼ核酸分子を含むDNA構築物およびこれらの構築物を含む細胞：

20

本発明はまた、宿主細胞において転写を開始するのに有効なプロモーターおよび上述の核酸分子を5'から3'方向に含む組換えDNA分子に関する。さらに、本発明は、ベクターおよび上述の核酸分子を含む組換えDNA分子に関する。本発明はまた、細胞において機能的な転写領域、上述のポリペプチドに対応するアミノ酸配列をコードするRNA配列に相補的な配列、および前記細胞において機能的な転写終止領域を含む核酸分子に関する。上述の分子は、単離されたおよび／または精製されたDNA分子でありうる。

【0197】

本発明はまた、上述の核酸分子を含み、したがってポリペプチドを発現しうる細胞または生物に関する。ポリペプチドは、そのポリペプチドを発現するよう変更されている細胞から精製することができる。細胞が、細胞が通常は産生しないかまたは細胞が通常はより低いレベルで産生する蛋白質を産生するように遺伝子操作により作成されている場合、細胞は”所望のポリペプチドを発現するよう変更されている”と言われる。当業者は、ゲノム、cDNA、または合成配列のいずれかを真核生物または原核生物細胞のいずれかに導入して発現させるための方法を容易に適用することができる。

30

【0198】

核酸分子、例えばDNAは、これが転写および翻訳制御情報を含むヌクレオチド配列を含み、かつそのような配列がポリペプチドをコードするヌクレオチド配列に”動作可能なように連結されている”場合、ポリペプチドを”発現しうる”と言われる。動作可能な連結とは、制御DNA配列および発現が求められているDNA配列が、遺伝子配列の発現を可能とするような様式で接続されているような連結である。遺伝子配列の発現に必要な制御領域の詳細な性質は生物によって異なるであろうが、これは一般にプロモーター領域を含む。原核生物においては、プロモーター領域は、プロモーター（RNA転写の開始を指示する）ならびにRNAに転写されたときに合成の開始を合図するであろうDNA配列の両方を含む。そのような領域は、通常は転写および翻訳の開始に関与する5'-非コーディング配列、例えばTATAボックス、キャッピング配列、CAAT配列等を含むであろう。

40

【0199】

所望の場合には、本発明のプロテアーゼをコードする配列の3'側の非コーディング領域を上述の方法により得ることができる。この領域は、その転写終止制御配列、例えば終止およびポリアダニル化のために保持することができる。すなわち、本発明のプロテアーゼ

50

をコードするDNA配列に天然に隣接している3'領域を保持することにより、転写終止シグナルを提供することができる。発現宿主細胞において転写終止シグナルが十分に機能性でない場合には、宿主細胞において機能的な3'領域で置き換えることができる。

【0200】

2つのDNA配列（例えば、プロモーター領域配列および本発明のプロテアーゼをコードする配列）は、2つのDNA配列の連結の性質が、プロテアーゼ配列の転写を可能とする場合、すなわち、連結が（1）フレームシフト変異を導入しない、（2）プロモーター領域配列が本発明のプロテアーゼをコードする遺伝子配列の転写を指示する能力を妨害しない、または（3）本発明のプロテアーゼの遺伝子配列がプロモーター領域配列により転写されることを妨害しない場合、動作可能なように連結されていると言われる。すなわち、プロモーター領域は、プロモーターがそのDNA配列の転写を行うことができる場合、DNA配列に動作可能なように連結されているであろう。すなわち、本発明のプロテアーゼをコードする遺伝子を発現させるためには、適当な宿主により認識される転写および翻訳シグナルが必要である。

10

【0201】

本発明は、本発明のプロテアーゼ（またはその機能的誘導体）をコードする遺伝子を原核生物または真核生物細胞のいずれかにおいて発現させることを包含する。原核生物宿主は、一般に、組換え蛋白質の製造において非常に有効でありかつ便利であり、したがって、本発明のプロテアーゼのための1つの好ましい発現システムである。原核生物は、しばしば*E. coli*の種々の株により代表される。しかし、他の細菌株等の他の微生物株もまた用いることができる。

20

【0202】

原核生物系においては、宿主と適合性のある種に由来する複製部位および制御配列を含むプラスミドベクターを用いることができる。適当なプラスミドベクターの例には、pBR322, pUC118, pUC119等が含まれ、適当なファージまたはバクテリオファージベクターの例には、 λ gt10, λ gt11等が含まれ、適当なウイルスベクターの例には、pMAM-neo, pKRC等が含まれる。好ましくは、本発明の選択されたベクターは、選択された宿主細胞において複製する能力を有する。

【0203】

認められている原核生物宿主には、細菌、例えば*E. coli*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*等が含まれる。しかし、そのような条件下においては、ポリペプチドはグリコシル化されない。原核生物宿主は発現プラスミド中のレプリコンおよび制御配列と適合性でなければならない。

30

【0204】

本発明のプロテアーゼ（またはその機能的誘導体）を原核生物細胞において発現させるためには、本発明のプロテアーゼをコードする配列が、機能的原核生物プロモーターと動作可能なように連結されていることが必要である。そのようなプロモーターは、構成的であってもよく、より好ましくは制御可能（すなわち、誘導可能または抑制解除可能）である。構成的プロモーターの例には、バクテリオファージ λ のintプロモーター、pBR322の β -ラクタマーゼ遺伝子配列のblaプロモーター、およびpPR325のクロラムフェニコールアセチルトランスフェラーゼ遺伝子配列のcatプロモーター等が含まれる。誘導可能な原核生物プロモーターの例には、バクテリオファージ λ の主要右および左プロモーター（ P_L および P_R ）、*E. coli*のtrp, recA, lacZ, lacI, およびgalプロモーター、 α -アミラーゼ（Ulmanen et al., J. Bacteriol. 162:176-182, 1985）および*B. subtilis*の σ -28-特異的プロモーター（Gilman et al., Gene Sequence 32:11-20, 1984）、*Bacillus*のバクテリオファージのプロモーター（Gryczan, The Molecular Biology of the Bacilli, Academic Press, Inc., NY, 1982）、お

40

50

よび *Streptomyces* プロモーター (Ward et al., Mol. Gen. Genet. 203:468-478, 1986) が含まれる。原核生物プロモーターは, Glick (Ind. Microbiol. 1:277-282, 1987), Cenatiempo (Biochimie 68:505-516, 1986), および Gottesman (Ann. Rev. Genet. 18:415-442, 1984) により概説されている。

【0205】

原核生物細胞における適切な発現には、遺伝子コーディング配列の上流にリボソーム結合部位の存在が必要である。そのようなリボソーム結合部位は、例えば、Gold et al. (Ann. Rev. Microbiol. 35:365-404, 1981) に記載されている。制御配列、発現ベクター、トランスフォーメーション方法等の選択は、遺伝子を発現させるために用いられる宿主細胞のタイプに依存する。本明細書において用いる場合、“細胞”、“細胞株”および“細胞培養”は互換的に用いることができ、そのような名称はすべて子孫を含む。すなわち、“トランスフォーマント”または“トランスフォームした細胞”との語句は、継代の数にかかわらず、初代対象細胞およびここから誘導された培養物を含む。また、意図的なまたは偶然の変異のために、すべての子孫がDNA含有物において精密に同一ではないかもしれないことが理解される。しかし、定義されるように、変異体子孫は元々のトランスフォームした細胞と同じ機能性を有する。

【0206】

本発明の発現システムにおいて用いることができる宿主細胞は、目的とするプロテアーゼポリペプチドの発現において用いるのに適当である限り、特に限定されない。適当な宿主はしばしば真核生物細胞を含むことができる。好ましい真核生物宿主には、例えば、酵母、真菌、昆虫細胞、哺乳動物細胞（インビボまたは組織培養のいずれか）が含まれる。宿主として有用でありうる哺乳動物細胞には、HeLa細胞、線維芽細胞由来の細胞、例えばVEROまたはCHO-K1、またはリンパ球由来の細胞、およびこれらの誘導体が含まれる。好ましい哺乳動物宿主細胞には、SP2/0およびJ558L、ならびに神経芽細胞腫細胞株、例えばIMR332が含まれ、これらは正しい翻訳後プロセッシングのよりすぐれた能力を提供することができる。

【0207】

さらに、植物細胞もまた宿主として利用可能であり、植物細胞と適合しうる制御配列、例えばカリフラワーモザイクウイルス35Sおよび19S、およびノパリンシンターゼプロモーターおよびポリアデニル化シグナル配列が利用可能である。他の好ましい宿主は昆虫細胞、例えば、*Drosophila* larvaeである。昆虫細胞を宿主として用いる場合、*Drosophila* アルコールデヒドロゲナーゼプロモーターを用いることができる (Rubin, Science 240:1453-1459, 1988)。あるいは、バキュロウイルスベクターを、昆虫細胞において本発明のプロテアーゼを大量に発現するよう遺伝子工学処理することができる (Jasny, Science 238:1653, 1987; Miller et al., Genetic Engineering, Vol. 8, Plenum, Setlow et al., eds., pp. 277-297, 1986)。

【0208】

酵母がグルコースの豊富な培地中で成長するときに大量に産生される解糖系酵素をコードする活発に発現されている配列からのプロモーターおよび終止要素を組み込んだ一連の酵母発現システムの任意のものを用いることができる。既知の解糖系遺伝子配列はまた、非常に効率的な転写制御シグナルを提供することができる。酵母は、翻訳後修飾を行うこともできる点において、実質的な利点を与える。強いプロモーター配列および高コピー数プラスミドを用いる多くの組換えDNA戦略が存在し、これを酵母において所望の蛋白質を製造するために用いることができる。酵母はクローン化された哺乳動物遺伝子のリーダー配列を認識して、リーダー配列を有するペプチド（すなわちプレペプチド）を分泌する。哺乳動物宿主における本発明のプロテアーゼの発現にはいくつかの可能なベクター系が利

用可能である。

【0209】

宿主の性質に応じて、広範な種類の転写および翻訳制御配列を用いることができる。転写および翻訳制御シグナルは、制御シグナルが高レベルの発現を有する特定の遺伝子配列に関連しているウイルス起源、例えばアデノウイルス、ウシパピローマウイルス、サイトメガロウイルス、サルウイルス等から誘導することができる。あるいは、哺乳動物発現産物、例えばアクチン、コラーゲン、ミオシンなどからのプロモーターを用いることができる。転写開始制御シグナルは、遺伝子配列の発現を調節することができるように、抑制または活性化を可能とするものを選択することができる。興味深いものは、温度を変化させることにより発現を抑制または開始することができるように温度感受性であるか、化学物質（例えば代謝産物）制御を行うことができる制御シグナルである。 10

【0210】

本発明のプロテアーゼの真核生物宿主における発現には、真核生物制御領域を使用することが必要である。そのような領域は、一般に、RNA合成の開始を指示するのに十分なプロモーター領域を含む。好ましい真核生物プロモーターには、例えば、マウスメタロチオネインI遺伝子配列のプロモーター（Hamer et al., J. Mol. Appl. Gen. 1: 273-288, 1982）；ヘルペスウイルスのTKプロモーター（McKnight, Cell 31: 355-365, 1982）；SV40初期プロモーター（Benoit et al., Nature (London) 290: 304-31, 1981）；および酵母gal4遺伝子配列プロモーター（Johnston et al., Proc. Natl. Acad. Sci. (USA) 79: 6971-6975, 1982；Silver et al., Proc. Natl. Acad. Sci. (USA) 81: 5951-5955, 1984）が含まれる。 20

【0211】

真核生物mRNAの翻訳は、最初のメチオニンをコードするコドンから開始される。この理由のため、真核生物プロモーターと、本発明のプロテアーゼ（またはその機能的誘導体）をコードするDNA配列との間の連結には、メチオニンをコードしうる介在コドン（すなわちAUG）が含まれないことを確実にすることが好ましい。そのようなコドンの存在は、融合蛋白質（AUGコドンが本発明のプロテアーゼのコーディング配列と同じリーディングフレームにある場合）またはフレームシフト変異（AUGコドンが本発明のプロテ 30
アーゼのコーディング配列と同じリーディングフレームにない場合）の形成のいずれかをもたらす。

【0212】

本発明のプロテアーゼをコードする核酸分子および動作可能なように連結されたプロモーターは、レシピエントである原核生物または真核生物細胞中に、非複製DNAまたはRNA分子のいずれかとして導入することができ、これは、直線状分子またはより好ましくは閉環状分子のいずれでもよい。そのような分子は自己複製することができないため、遺伝子の発現は導入された配列の過渡的発現により生ずる。あるいは、導入されたDNA配列が宿主染色体中にインテグレートされることにより永久発現が得られる。

【0213】

所望の遺伝子配列を宿主細胞染色体中にインテグレートしうるベクターを用いることができる。導入されたDNAをその染色体中に安定にインテグレートしている細胞は、発現ベクターを含有する宿主細胞の選択を可能とする1またはそれ以上のマーカーを導入することにより選択することができる。マーカーは、栄養要求性宿主に対する栄養、殺生物剤耐性（例えば抗生物質、または重金属、例えば銅）等を提供することができる。選択マーカー遺伝子配列は、発現させるべきDNA遺伝子配列に直接連結してもよく、または同じ細胞にコトランスフェクションにより導入してもよい。mRNAの最適な合成には追加の要素も必要であろう。これらの要素には、スプライシングシグナル、ならびに転写プロモーター、エンハンサー、および終止シグナルが含まれる。そのような要素を組み込んだcDNA発現ベクターには、Okayama (Mol. Cell. Biol. 3: 280-2 40 50

89, 1983) により記載されるものが含まれる。

【0214】

導入された核酸分子は、レシピエント宿主中で自己複製可能なプラスミドまたはウイルスベクター中に取り込ませることができる。この目的のために広範な種類のベクターの任意のものをを用いることができる。特定のプラスミドまたはウイルスベクターの選択において重要な因子には、ベクターを含むレシピエント細胞を認識しベクターを含まないレシピエント細胞から選択することの容易性；特定の宿主において所望されるベクターのコピー数；およびベクターを異なる種の宿主細胞間で”シャトル”しうるということが望ましいか否かが含まれる。

【0215】

好ましい原核生物ベクターには、E. coli 中で複製しうるプラスミド（例えば、pBR322, ColE1, pSC101, pACYC184, Σ VX; "Molecular Cloning: A Laboratory Manual", 1989, 上掲）などのプラスミドが含まれる。Bacillus プラスミドには、pC194, pC221, pT127 等が含まれる (Gryczan, The Molecular Biology of the Bacilli, Academic Press, NY, pp. 307-329, 1982)。適当な Streptomyces プラスミドには、p1J101 (Kendall et al., J. Bacteriol. 169: 4177-4183, 1987), および Streptomyces バクテリオファージ, 例えば Φ C31 (Chater et al., Sixth International Symposium on Actinomycetiales Biology, Akademiai Kiado, Budapest, Hungary, pp. 45-54, 1986) が含まれる。Pseudomonas プラスミドは John et al. (Rev. Infect. Dis. 8: 693-704, 1986), および Izaki (Jpn. J. Bacteriol. 33: 729-742, 1978) により概説されている。

【0216】

好ましい真核生物プラスミドには、例えば、BPV, ワクチニア, SV40, 2-ミクロンサークル等、またはそれらの誘導体が含まれる。そのようなプラスミドは当該技術分野においてよく知られている (Botstein et al., Miami Wnter. Symp. 19: 265-274, 1982; Broach, "The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance", Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470, 1981; Broach, Cell 28: 203-204, 1982; Bollon et al., J. Clin. Hematol. Oncol. 10: 39-48, 1980; Maniatis, Cell Biology: A Comprehensive Treatise, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608, 1980)。

【0217】

構築物を含むベクターまたは核酸分子を発現用に用意した後、DNA 構築物は、種々の適当な手段、すなわち、トランスフォーメーション、トランスフェクション、コンジュゲーション、プロトプラスト融合、エレクトロポレーション、粒子銃技術、リン酸カルシウム沈澱、直接マイクロインジェクション等により、適当な宿主細胞中に導入することができる。ベクターを導入した後、レシピエント細胞を、ベクター含有細胞の成長を選択する選択培地中で成長させる。クローニングされた遺伝子の発現により、本発明のプロテアーゼまたはそのフラグメントが産生される。これは、トランスフォームした細胞中でそのまま起こるか、またはこれらの細胞を分化させるよう誘導した後に起こる（例えば、プロモデオキシウラシルを神経芽細胞腫等に投与することにより）。本発明のペプチドを形成するために、種々のインキュベーション条件を用いることができる。最も好ましい条件は

，生理学的条件を模倣した条件である。

【0218】

抗体，ハイブリドーマ，プロテアーゼの検出のための使用方法およびキット

本発明は，本発明のプロテアーゼに結合親和性を有する抗体に関する。プロテアーゼポリペプチドは，配列番号60，配列番号61，配列番号62，配列番号63，配列番号64，配列番号65，配列番号66，配列番号67，配列番号68，配列番号69，配列番号70，配列番号71，配列番号72，配列番号73，配列番号74，配列番号75，配列番号76，配列番号77，配列番号78，配列番号79，配列番号80，配列番号81，配列番号82，配列番号83，配列番号84，配列番号85，配列番号86，配列番号87，配列番号88，配列番号89，配列番号90，配列番号91，配列番号92，配列番号93，配列番号94，配列番号95，配列番号96，配列番号97，配列番号98，配列番号99，配列番号100，配列番号101，配列番号102，配列番号103，配列番号104，配列番号105，配列番号106，配列番号107，配列番号108，配列番号109，配列番号110，配列番号111，配列番号112，配列番号113，配列番号114，配列番号115，配列番号116，配列番号117および配列番号118に記載されるアミノ酸配列，またはその機能的誘導体，またはその少なくとも9個の連続するアミノ酸（好ましくは，その少なくとも20，30，35，または40個の連続するアミノ酸）からなる群より選択されるアミノ酸配列を有することができる。

【0219】

本発明はまた，本発明のプロテアーゼに対して特異的結合親和性を有する抗体に関する。そのような抗体は，本発明のプロテアーゼに対するその結合親和性を他のポリペプチドに対する結合親和性と比較することにより単離することができる。本発明のプロテアーゼに選択的に結合する抗体を，本発明のプロテアーゼと他のポリペプチドとを区別することを必要とする方法において用いるために選択することができる。そのような方法には，限定されないが，他のポリペプチドを含む組織におけるプロテアーゼ発現の変化の分析が含まれる。

【0220】

本発明のプロテアーゼは，種々の手順および方法，例えば抗体の生成，医薬組成物の同定，およびDNA／蛋白質相互作用の研究において用いることができる。

【0221】

本発明のプロテアーゼは，抗体またはハイブリドーマを生成するために用いることができる。当業者は，抗体が望まれる場合，そのようなペプチドを本明細書に記載されるように生成し，免疫原として用いることができることを認識するであろう。本発明の抗体には，モノクローナル抗体およびポリクローナル抗体，ならびにこれらの抗体のフラグメント，およびヒト化型が含まれる。本発明の抗体のヒト化型は，キメラ化またはCDRグラフトリング等の当該技術分野において知られる手法の1つを用いて生成することができる。

【0222】

本発明はまた，上述のモノクローナル抗体またはその結合フラグメントを産生するハイブリドーマに関する。ハイブリドーマは，特定のモノクローナル抗体を分泌することができる不死化細胞株である。

【0223】

一般に，モノクローナル抗体およびハイブリドーマを製造する手法は当該技術分野においてよく知られている（Campbell，"Monoclonal Antibody Technology：Laboratory Techniques in Biochemistry and Molecular Biology"，Elsevier Science Publishers，Amsterdam，The Netherlands，1984；St. Groth et al.，J. Immunol. Methods 35：1-21，1980）。抗体を生成することが知られている任意の動物（マウス，ウサギ等）を，選択されたポリペプチドで免疫することができる。免疫の方法は当該技術分野においてよく知られている。そのような方法には，ポリペプチドの皮下また

は腹膜内注射が含まれる。当業者は、免疫に用いるポリペプチドの量は、免疫する動物、ポリペプチドの抗原性、および注入部位により様々であることを認識するであろう。

【0224】

ポリペプチドは、ペプチドの抗原性を増加させるために、修飾するかまたはアジュバント中で投与することができる。ポリペプチドの抗原性を増加させる方法は当該技術分野においてよく知られている。そのような方法には、抗原を異種蛋白質（例えばグロブリンまたはβ-ガラクトシダーゼ）とカップリングさせるか、または免疫の間にアジュバントを含めることが含まれる。

【0225】

モノクローナル抗体については、免疫した動物から脾臓細胞を切除し、ミエローマ細胞、例えばSP2/O-Ag14ミエローマ細胞と融合させ、モノクローナル抗体産生ハイブリドーマ細胞とさせる。当該技術分野においてよく知られる多数の方法の任意のものをを用いて、所望の特性を有する抗体を産生するハイブリドーマ細胞を同定することができる。これらには、ハイブリドーマをELISAアッセイ、ウエスタンブロット分析、またはラジオイムノアッセイを用いてスクリーニングすることが含まれる（Lutz et al., Exp. Cell Res. 175:109-124, 1988）。所望の抗体を分泌するハイブリドーマをクローニングし、当該技術分野において知られる方法を用いてクラスおよびサブクラスを決定する（Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", 上掲, 1984）。 10 20

【0226】

ポリクローナル抗体については、免疫した動物から抗体を含有する抗血清を単離し、上述の方法の1つを用いて所望の特異性を有する抗体の存在についてスクリーニングする。上述の抗体は、検出可能なように標識することができる。抗体は、放射性同位体、アフィニティー標識（例えばビオチン、アビジン等）、酵素標識（例えば西洋ワサビペルオキシダーゼ、アルカリホスファターゼ等）、蛍光標識（例えばFITCまたはローダミン等）、常磁性原子等を用いて、検出可能なように標識することができる。そのような標識を行う方法は当該技術分野においてよく知られている。例えば、Stemberger et al., J. Histochem. Cytochem. 18:315, 1970; Bayer et al. Meth. Enzym. 62:308, 1979; Engvall et al., Immunol. 109:129, 1972; Goding, J. Immunol. Meth. 13:215, 1976を参照。本発明の抗体は、二次標識抗体、例えば標識された抗ウサギ抗体を用いることにより間接的に標識してもよい。本発明の標識された抗体は、インビトロ、インビボ、およびインシトゥーアッセイに用いて、特定のペプチドを発現する細胞または組織を同定することができる。 30

【0227】

上述の抗体は、固体支持体上に固定化してもよい。そのような固体支持体の例には、プラスチック、例えばポリカーボネート、複合炭水化物、例えばアガロースおよびセファロース、アクリル樹脂、例えばポリアクリルアミドおよびラテックスビーズが含まれる。抗体をそのような固体支持体にカップリングさせる技術は当該技術分野においてよく知られている（Weir et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10, 1986; Jacoby et al., Meth. Enzym. 34, Academic Press, N. Y., 1974）。本発明の固定化された抗体は、インビトロ、インビボ、およびインシトゥーアッセイに、ならびに免疫クロマトグラフィーに用いることができる。 40

【0228】

さらに、当業者は、合理的に設計された抗ペプチドペプチドを生成するために、現在利用 50

可能な方法、並びに抗体に関して本明細書に記載される技術、方法およびキットを容易に適合させて、特定のペプチド配列に結合しうるペプチドを容易に生成することができる（Hurby et al., "Application of Synthetic Peptides: Antisense Peptides", In Synthetic Peptides, A User's Guide, W. H. Freeman, NY, pp. 289-307, 1992; Kaspczak et al., Biochemistry 28: 9230-9238, 1989）。

【0229】

抗ペプチドペプチドは、本発明のプロテアーゼのペプチド配列中に見いだされる塩基性アミノ酸残基を、疎水性および非荷電極性基を維持しながら酸性残基で置き換えることにより生成することができる。例えば、リジン、アルギニン、および／またはヒスチジン残基をアスパラギン酸またはグルタミン酸で置き換え、およびグルタミン酸残基をリジン、アルギニンまたはヒスチジンで置き換える。

10

【0230】

本発明はまた、試料中においてプロテアーゼポリペプチドを検出する方法を包含する。該方法は、（a）試料を、免疫複合体が形成するような条件下で上述の抗体と接触させ、そして（b）ポリペプチドに結合した前記抗体の存在を検出する、ことを含む。詳細には、該方法は、試験試料を1またはそれ以上の本発明の抗体とインキュベートし、抗体が試験試料に結合するか否かをアッセイすることを含む。試料中で本発明のプロテアーゼのレベルが正常なレベルと比較して変化していることは疾病を示すかもしれない。

20

【0231】

抗体を試験試料とインキュベートする条件は様々である。インキュベーション条件は、アッセイにおいて用いられるフォーマット、用いられる検出方法、およびアッセイにおいて用いられる抗体のタイプおよび性質によって異なる。当業者は、慣用的に入手可能な免疫学的アッセイフォーマット（例えばラジオイムノアッセイ、酵素結合イムノソルベントアッセイ、拡散に基づくオクタロニー、またはロケット免疫蛍光アッセイ）の任意のものを、本発明の抗体を用いるために容易に適合させることができることを認識するであろう。そのようなアッセイの例は、Chard（"An Introduction to Radioimmunoassay and Related Techniques", Elsevier Science Publishers, Amsterdam, The Netherlands, 1986）、Bullock et al.（"Techniques in Immunocytochemistry", Academic Press, Orlando, FL Vol. 1, 1982; Vol. 2, 1983; Vol. 3, 1985）、Tijssen（"Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology", Elsevier Science Publishers, Amsterdam, The Netherlands, 1985）に見いだすことができる。

30

【0232】

本発明の免疫学的アッセイ試験試料には、細胞、蛋白質または細胞の膜抽出物、または体液、例えば血液、血清、血漿または尿が含まれる。上述の方法において用いられる試験試料は、アッセイフォーマット、検出方法の性質およびアッセイすべき試料として用いられる組織、細胞または抽出物により様々であろう。蛋白質抽出物または細胞の膜抽出物を製造する方法は当該技術分野においてよく知られており、用いるシステムにより試験しうる試料を得るために容易に適合させることができる。

40

【0233】

キットは、先に記載した検出方法を実施するために必要な全ての試薬を含む。キットは、（i）上述の抗体を含む第1の容器手段、および（ii）抗体の結合パートナーと標識を含むコンジュゲートを含む第2の容器手段を含むことができる。別の好ましい態様においては、キットは以下の1またはそれ以上を含む1またはそれ以上の他の容器をさらに含む

50

：洗浄試薬および結合した抗体の存在を検出する試薬。

【0234】

検出試薬の例には、限定されないが、標識二次抗体が含まれ、あるいは、一次抗体が標識されている場合には、標識された抗体と反応する発色団、酵素、または抗体結合試薬が含まれる。コンパートメント化キットは核酸プローブキットについて上述したようなものでもよい。当業者は、本発明に記載される抗体を、当該技術分野においてよく知られる確立されたキットフォーマットの1つに容易に取り込ませることができることを容易に認識するであろう。

【0235】

プロテアーゼと相互作用することができる化合物の単離

10

本発明はまた、本発明のプロテアーゼに結合する化合物を検出する方法に関する。該方法は、化合物を本発明のプロテアーゼとともにインキュベートし、プロテアーゼに結合した化合物の存在を検出することを含む。化合物は、複雑な混合物、例えば、血清、体液、または細胞抽出物中に存在していてもよい。

【0236】

本発明はまた、プロテアーゼの活性またはプロテアーゼの結合パートナーの活性のアゴニストまたはアンタゴニストを検出する方法に関する。該方法は、本発明のプロテアーゼを産生する細胞を化合物の存在下でインキュベートし、プロテアーゼ活性またはプロテアーゼ結合パートナーの活性のレベルの変化を検出することを含む。このようにして同定される化合物は、化合物の存在を示す活性の変化を生ずるであろう。化合物は、複雑な混合物、例えば、血清、体液、または細胞抽出物中に存在していてもよい。いったん化合物が同定されれば、当該技術分野においてよく知られる技術を用いてこれを単離することができる。

20

【0237】

本発明はまた、哺乳動物においてプロテアーゼ関連活性を調節する方法を包含する。該方法は、前記哺乳動物に本発明のプロテアーゼのアゴニストまたはアンタゴニストを前記調節を行うのに十分な量で投与することを含む。哺乳動物にプロテアーゼ関連機能をアゴナイズまたはアンタゴナイズするのに十分な量でアゴニストまたはアンタゴニストを投与することを含む、本発明のプロテアーゼの1つの活性のアゴニストまたはアンタゴニストを用いて哺乳動物において疾病を治療する方法もまた、本出願に包含される。

30

【0238】

疾病の新規な治療を発見することをめざして、生物医学研究者および化学者は、プロテアーゼの機能を阻害する分子を設計し、合成し、試験してきた。いくつかの小さい有機分子は蛋白質プロテアーゼの機能を調節する化合物の一群を形成する。

【0239】

蛋白質プロテアーゼの機能を阻害することが報告されている分子の例には、限定されないが、フッ化フェニルメチルスルホニル (PMSEF)、ジソプロピルフルオロホスフェート (DFP) (3章, Barrett et al., Handbook of Proteolytic Enzymes, 1998, Academic Press, San Diego), 3, 4-ジクロロイソクマリン (DCI) (同, 16章), セルピン (同, 37章), E-64 (トランス-エポキシスクシニル-L-ロイシルアミド (4-グアニジノ)ブタン) (同, 188章), ペプチジル-ジアゾメタン, ペプチジル-O-アシル-ヒドロキサメート, エポキシスクシニル-ペプチド (同, 210章), DAN, EPNP (1, 2-エポキシ-3 (p-ニトロフェノキシ)プロパン) (同, 298章), チオルファン (d1-3-メルカプト-2-ベンジルプロパノイル-グリシン) (同, 362章), CGS26303, PD069185 (同, 363章), およびCOT989-00 (N-4-ヒドロキシ-N1-[1-(s)-(4-アミノスルホニル)フェニルエチル-アミノカルボキシル-2-シクロヘキシルエチル]-2R-[4-メチル)フェニルプロピル]スクシンアミド) (同, 401章)が含まれる。他のプロテアーゼ阻害剤には、限定されないが、アプロチニン、アマスタチン、アンチペイン、カルシニューリン

40

50

自己阻害フラグメント、およびヒスタミン5（同）が含まれる。好ましくは、これらの阻害剤は、100-200ダルトン、200-300ダルトン、300-400ダルトン、400-600ダルトン、600-1000ダルトン、1000-2000ダルトン、2000-4000ダルトン、および4000-8000ダルトンの分子量を有する。

【0240】

細胞膜を横切ることができ、酸加水分解に耐性の化合物は、患者に経口投与された後に高度に生物利用性となることができるため、治療剤として利点を有する可能性がある。しかし、これらのプロテアーゼ阻害剤の多くは、プロテアーゼの機能を弱くしか阻害しない。さらに、多くは種々のプロテアーゼを阻害し、したがって、疾病の治療剤として多くの副作用を引き起こすであろう。

10

【0241】

トランスジェニック動物：

本発明に関連するトランスジェニック動物の製造には、種々の方法が利用可能である。DNAを受精可能卵の前核に注入した後、雄前核と雌前核を融合させるか、またはDNAを胚性細胞（例えば、2細胞胚の核）の核に注入した後、細胞分裂を開始させることができる（Brinster et al., Proc. Nat. Acad. Sci. USA 82:4438-4442, 1985）。本発明の無機イオンレセプターヌクレオチド配列を有するよう改変したウイルス、特にレトロウイルスを胚に感染させることができる。

【0242】

胚の内部細胞塊から誘導され培養中で安定化させた多能性幹細胞を、培養中で操作して本発明のヌクレオチド配列を取り込ませることができる。トランスジェニック動物は、そのような細胞を胚盤胞中に移植し、これを仮母に移植し、分娩させることにより製造することができる。トランスジェニック実験に適当な動物は、標準的な商業的供給源、例えばCharles River (Wilmington, MA), Taconic (Germantown, NY), Harlan Sprague Dawley (Indianapolis, IN) 等から入手することができる。

20

【0243】

齧歯類胚を操作する方法およびDNAを接合子の前核に注入する方法は、当業者によく知られている（Hogan et al., 上掲）。魚、両生類卵および鳥類のためのマイクロインジェクション法は、Houdebine and Chourrout (Experientia 47:897-905, 1991) に詳述されている。DNAを動物の組織中に導入する他の方法は、米国特許、4,945,050 (Sandford et al., 1990年7月30日) に記載されている。

30

【0244】

一例にすぎないが、トランスジェニックマウスを製造するためには、雌マウスを過剰排卵誘発する。雌を雄といっしょに置き、交配した雌をCO₂窒息または頸部脱臼により殺し、切除した卵管から胚を回収する。まわりの丘細胞を除去する。次に、前核胚を洗浄し、注入時まで保存する。ランダムな周期の成人雌を精管切除雄と対にする。レシピエント雌はドナー雌と同じ時に交配させる。次に胚を外科的に移す。トランスジェニックラットを製造する方法はマウスの場合と類似する方法である（Hammer et al., Cell 63:1099-1112, 1990）。

40

【0245】

胚性幹（ES）細胞を培養し、次にエレクトロポレーション、リン酸カルシウム／DNA沈澱および直接注入等の方法を用いてDNAをES細胞中に導入することによりトランスジェニック動物を製造する方法もまた当業者によく知られている（Teratocarcinomas and Embryonic Stem Cell, A Practical Approach, E. J. Robertson, ed., IRL Press, 1987）。

【0246】

ランダム遺伝子インテグレーションを含む場合には、本発明の配列を含むクローンを耐性

50

をコードする遺伝子とともにコトランスフェクトすることができる。あるいは、ネオマイシン耐性をコードする遺伝子を、本発明の配列に物理的に連結させることができる。所望のクローンのトランスフェクションおよび単離は、当業者によく知られるいくつかの方法のいずれかを用いて実施することができる (E. J. Robertson, 上掲)。

【0247】

ES細胞中に導入されたDNA分子はまた、相同組換えのプロセスにより染色体中にインテグレートされることができる (Capecci, Science 244:1288-1292, 1989)。組換え事象のポジティブ選択 (すなわち, neo耐性) および二重ポジティブネガティブ選択 (すなわち, neo耐性およびガンシクロビル耐性) の方法, および続くPCRによる所望のクローンの同定は, Capecci, 上掲および Joyner et al. (Nature 338:153-156, 1989) に記載されており, これらの教示は, 図面を含めその全体を本明細書の一部としてここに引用する。この方法の最後の段階は, 標的とするES細胞を胚盤胞中に注入し, 胚盤胞を擬妊娠雌に移すことである。得られるキメラ動物を繁殖させ, 子孫をサザンブロッティングにより分析して, トランスジンを有する個体を同定する。非齧歯類哺乳動物および他の動物の製造方法は別の者により議論されている (Houdebine and Chourout, 上掲; Pursell et al., Science 244:1281-1288, 1989; and Simms et al., Bio/Technology 6:179-183, 1988)。

【0248】

すなわち, 本発明は, 本発明のプロテアーゼをコードするトランスジンをまたはプロテアーゼの発現に影響する遺伝子を含むトランスジェニックの非ヒト哺乳動物を提供する。そのようなトランスジェニックの非ヒト哺乳動物は, プロテアーゼの導入の効果を研究するため, またはプロテアーゼの発現を制御 (すなわち, 追加の遺伝子, アンチセンス核酸, またはリボザイムの導入により) するためのインビボ試験系として特に有用である。

【0249】

”トランスジェニック動物”とは, 細胞中に人工的に挿入されたDNAを含む細胞を有する動物である。該DNAは, その細胞から発生した動物のゲノムの一部となる。好ましいトランスジェニック動物は, 霊長類, マウス, ラット, ウシ, ブタ, ウマ, ヤギ, ヒツジ, イヌおよびネコである。トランスジェニックDNAは, ヒトプロテアーゼをコードすることができる。動物における自然の発現は, レセプターの発現を減少させるのに有効な量のアンチセンスRNAまたはDNAを与えることにより減少させることができる。

【0250】

遺伝子治療:

本発明のプロテアーゼまたはその遺伝子配列は, 遺伝子治療においても有用である (総説としてMiller, Nature 357:455-460, 1992)。Millerは, 進歩により, ポジティブな初期の結果を示したヒト遺伝子治療に対する実用的なアプローチが得られたと述べている。遺伝子治療の基本的な科学はMulligan (Science 260:926-931, 1993) に記載されている。

【0251】

1つの好ましい態様においては, プロテアーゼのコーディング配列を含む発現ベクターを細胞に挿入し, 細胞をインビトロで成長させ, 次にこれを大量に患者に注入する。別の好ましい態様においては, 選択されたプロモーター (例えば, 強いプロモーター) を含むDNAセグメントを, 本発明のプロテアーゼをコードする内因性遺伝子を含む細胞中に, プロモーターセグメントが内因性プロテアーゼ遺伝子の発現を増強する様式で移送する (例えば, プロモーターセグメントをこれが内因性プロテアーゼ遺伝子に直接連結するように細胞内に移送する)。

【0252】

遺伝子治療は, 腫瘍を標的とするプロテアーゼcDNAを含むアデノウイルスの使用を含むことができる。全身プロテアーゼは, 遺伝子工学処理した細胞の移植, そのようなプロ

テアーゼをコードするウイルスの注入，または裸のプロテアーゼDNAの適当な組織への注入により増加する。

【0253】

そのような複合体の活性を調節するために，標的細胞集団を，蛋白質複合体の1またはそれ以上の変更された形の成分を導入することにより改変することができる。例えば，標的細胞中における複合体成分の活性を減少させるか阻害することにより，そのような状態につながる異常なシグナル伝達事象を減少させ，阻害し，または逆転させることができる。蛋白質複合体の他の成分と相互作用する能力を保持しているが，シグナル伝達において機能することができない成分の欠失またはミスセンス変異体を用いて，異常な，有害なシグナル伝達事象を阻害することができる。

10

【0254】

ウイルス，例えばレトロウイルス，ワクチニアウイルス，アデノウイルス，アデノ随伴ウイルス，ヘルペスウイルス，いくつかのRNAウイルス，またはウシパピローマウイルスに由来する発現ベクターを用いて，本発明の組換えプロテアーゼをコードするヌクレオチド配列（例えばcDNA）を標的細胞集団（例えば腫瘍細胞）に輸送することができる。当業者によく知られる方法を用いて，コーディング配列を含む組換えウイルスベクターを構築することができる（Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, N. Y., 1989; Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N. Y., 1989）。あるいは，蛋白質配列をコードする組換え核酸分子を裸のDNAとしてまたは再構築系，例えば標的細胞への輸送のためのリポソームまたは他の脂質系において用いることができる（例えば，Felgner et al., Nature 337: 387-8, 1989）。ヒト遺伝子治療において用いるための，プラスミドDNAを細胞内に直接輸送するいくつかの他の方法が存在し，これはプラスミドDNAを蛋白質に複合体化させることによりDNAを細胞上のレセプターにターゲティングすることを含む（Miller, 上掲）。

20

【0255】

最も簡単な形においては，遺伝子輸送は，単にマイクロインジェクション工程により細胞の核内に最小量のDNAを注入することにより行うことができる（Capecci, Cell 22: 479-88, 1980）。組換え遺伝子は，いったん細胞内に導入されると，転写および翻訳に関する細胞の正常なメカニズムにより認識されることができ，遺伝子産物が発現される。より多数の細胞にDNAを導入するための他の方法も試みられている。これらの方法には，DNAをリン酸カルシウムで沈澱させピノサイトーシスにより細胞中に取り込ませるトランスフェクション（Chen et al., Mol. Cell Biol. 7: 2745-52, 1987）；細胞を高圧パルスに暴露して膜に穴をあけるエレクトロポレーション（Chu et al., Nucleic Acids Res. 15: 1311-26, 1987）；DNAを親油性ベヒクル中に封入し，これを標的細胞と融合させるリポフェクチン／リポソーム融合（Felgner et al., Proc. Natl. Acad. Sci. USA. 84: 7413-7417, 1987）；および小さい発射体に結合させたDNAを用いる粒子衝撃（Yang et al., Proc. Natl. Acad. Sci. 87: 9568-9572, 1990）が含まれる。DNAを細胞内に導入する他の方法は，DNAを化学的に修飾した蛋白質にカップリングさせることである。

30

40

【0256】

また，アデノウイルス蛋白質がエンドソームを不安定化させ，DNAの細胞への取り込みを促進しうることが示されている。アデノウイルスをDNA複合体を含む溶液と混合するか，または蛋白質架橋試薬を用いてアデノウイルスに共有結合したポリリジンにDNAを結合させることにより，組換え遺伝子の取り込みおよび発現が実質的に改良される（Cu

50

riel et al., Am. J. Respir. Cell. Mol. Biol. 6 : 247-52, 1992)。

【0257】

本明細書において用いる場合，“遺伝子輸送”とは，外来核酸分子を細胞中に導入する工程を意味する。遺伝子輸送は，一般に，遺伝子によりコードされる特定の産物の発現を可能とするために行われる。産物には，蛋白質，ポリペプチド，アンチセンスDNAまたはRNA，または酵素的に活性なRNAが含まれる。遺伝子輸送は，培養細胞中で，または動物への直接投与により行うことができる。一般に，遺伝子輸送には，核酸を非特異的レセプター媒介性相互作用により標的細胞と接触させ，核酸を膜を通してまたはエンドサイトーシスにより細胞内に取り込ませ，核酸を原形質膜またはエンドソームから細胞質内に放出させる工程が含まれる。さらに，発現には，核酸が細胞の核に移動し，転写のための適当な核因子に結合することが必要である。 10

【0258】

本明細書において用いる場合，“遺伝子治療”とは，遺伝子輸送の1つの形であり，本明細書において用いられる遺伝子輸送の定義の中に含まれ，特にインビボでまたはインビトロで細胞から治療用産物を発現させるための遺伝子輸送を表す。遺伝子輸送は，細胞でエクスピボで行い，次に患者に移植することにより，または，核酸または核酸蛋白質複合体を患者に直接投与することにより，行うことができる。

【0259】

別の好ましい態様においては，プロテアーゼポリペプチドをコードする核酸配列を有するベクターが提供され，ここで，核酸配列は特定の組織においてのみ発現される。組織特異的遺伝子発現を行う方法は，国際公開WO93/09236（1992年11月3日出願，1993年5月13日公開）に記載される。 20

【0260】

上述したすべてのベクターにおいて，本発明のさらに別の観点は，ベクターに含まれる核酸配列が，核酸の配列の一部または全てについて，上で定義したような付加，欠失または修飾を含んでいてもよいことである。

【0261】

本発明のプロテアーゼポリペプチドの発現（過剰発現を含む）は，ポリペプチドをコードするmRNAに結合しその発現を阻害するアンチセンス分子を投与することにより阻害することができる。あるいは，mRNAを切断するリボザイムを用いて同様の様式で発現を阻害することができる。アンチセンスおよびリボザイム技術を用いて遺伝子発現を制御する一般的方法，またはこのようにして外因性遺伝子を発現させる遺伝子治療方法は当該技術分野においてよく知られている。これらの各方法は，例えば，本発明のプロテアーゼポリペプチドのアンチセンスまたはリボザイム転写産物のいずれかをコードするベクター等のシステムを利用する。 30

【0262】

“リボザイム”との用語は，触媒的特性を有する1またはそれ以上のRNAのRNA構造を表す。リボザイムは，一般に，エンドヌクレアーゼ，リガーゼまたはポリメラーゼ活性を示す。リボザイムは多数のRNA自己切断反応を媒介する構造化されたRNA分子である。異なる二次構造を有する種々のタイプのトランス作用性リボザイム，例えば“ハンマーヘッド”および“ヘアピン”タイプのものが同定されている。種々のリボザイムが特性決定されている。例えば，米国特許5,246,921, 5,225,347, 5,225,337および5,149,796を参照。デオキシリボオリゴヌクレオチドとリボオリゴヌクレオチドとを含む，触媒活性を有する混合リボザイムが記載されている（Perrault, et al., Nature, 344:565-567 (1990)）。 40

【0263】

本明細書において用いる場合，“アンチセンス”とは，ゲノムDNAおよび／または本発明のプロテアーゼポリペプチドをコードするmRNAと，特異的にハイブリダイズし，例えば細胞条件下で結合して，例えば，転写および／または翻訳を阻害することによりその 50

蛋白質の発現を阻害する、核酸分子またはその誘導体を表す。結合は一般的な塩基対相補性によるものでもよく、または、例えば、DNAデュプレックスへの結合の場合には、二重ヘリックスの主溝における特異的相互作用によるものでもよい。

【0264】

1つの観点においては、アンチセンス構築物は、エキスピボで生成され、細胞中に導入されたときに、限定されないが、本発明のプロテアーゼポリヌクレオチドのmRNAおよび／またはゲノム配列とハイブリダイズすることにより遺伝子発現を阻害しうる核酸である。

【0265】

アンチセンス方法は、プロテアーゼポリペプチドmRNAに相補的であり、本発明のプロ
テアーゼポリヌクレオチド、例えば、配列番号1、配列番号2、配列番号3、配列番号4
、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列
番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、
配列番号17、配列番号18、配列番号19、配列番号20、配列番号21、配列番号2
2、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、配列番
号28、配列番号29、配列番号30、配列番号31、配列番号32、配列番号33、配
列番号34、配列番号35、配列番号36、配列番号37、配列番号38、配列番号39
、配列番号40、配列番号41、配列番号42、配列番号43、配列番号44、配列番号
45、配列番号46、配列番号47、配列番号48、配列番号49、配列番号50、配列
番号51、配列番号52、配列番号53、配列番号54、配列番号55、配列番号56、
配列番号57、配列番号58、および配列番号59に基づくオリゴヌクレオチド（DNA
またはRNAのいずれか）の設計を含むことができる。アンチセンスオリゴヌクレオチド
は、プロテアーゼポリペプチドのmRNA転写産物に結合し、翻訳を妨害する。

【0266】

完全な相補性が好ましいが、必要ではない。本明細書において用いる場合、RNAの一部
に”相補的な”配列とは、RNAとハイブリダイズして安定なデュプレックスを形成す
るのに十分な相補性を有する配列を意味する。二本鎖アンチセンス核酸の場合には、デュ
プレックスDNAの一本鎖を試験するか、またはトリプレックス形成をアッセイするこ
とができる。ハイブリダイズする能力は、相補性の程度とアンチセンス核酸の長さの両方
に依存する。一般に、ハイブリダイズする核酸が長いほど、RNAとより多い塩基ミスマ
ッチを含むことができ、なお安定なデュプレックス（または場合によりトリプレックス
）を形成することができる。当業者は、ハイブリダイズした複合体の融点を決定する標準
的な方法を用いることにより、ミスマッチの許容可能な程度を確かめることができる。

【0267】

一般に、メッセージの5'末端、例えばAUG開始コドンまでおよびこれを含む5'非翻
訳配列に相補的なオリゴヌクレオチドは、翻訳の阻害において最も有効に作用するはずで
ある。しかし、mRNAの3'非翻訳配列に相補的な配列も同様にmRNAの翻訳の阻害
に有効であることが示されている（Wagner, R. (1994) Nature 372:333）。mRNAのコーディング領域に相補的なアンチセンスオリゴヌクレオチド
は比較的有效性の低い翻訳の阻害剤であるが、本発明にしたがって用いることが可能であ
る。プロテアーゼポリペプチドのmRNAの5'、3'またはコーディング領域のいずれ
にハイブリダイズするよう設計されているかにかかわらず、アンチセンス核酸は少なくと
も6ヌクレオチドの長さであるべきであり、好ましくは約100ヌクレオチドより短く、
より好ましくは約50または30ヌクレオチドより短い。典型的には、これらは10-2
5ヌクレオチドの長さであるべきである。このような原則は、現場の者に適当なオリゴヌ
クレオチドを選択する上での情報を与えるであろう。好ましい態様においては、アンチセ
ンス配列は、配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号
6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12
、配列番号13、配列番号14、配列番号15、配列番号16、配列番号17、配列番号
18、配列番号19、配列番号20、配列番号21、配列番号22、配列番号23、配列

番号 24, 配列番号 25, 配列番号 26, 配列番号 27, 配列番号 28, 配列番号 29, 配列番号 30, 配列番号 31, 配列番号 32, 配列番号 33, 配列番号 34, 配列番号 35, 配列番号 36, 配列番号 37, 配列番号 38, 配列番号 39, 配列番号 40, 配列番号 41, 配列番号 42, 配列番号 43, 配列番号 44, 配列番号 45, 配列番号 46, 配列番号 47, 配列番号 48, 配列番号 49, 配列番号 50, 配列番号 51, 配列番号 52, 配列番号 53, 配列番号 54, 配列番号 55, 配列番号 56, 配列番号 57, 配列番号 58, および配列番号 59 からなる群より選択される核酸配列またはこれらのドメインの, 約 10-30 個の, より好ましくは 15-25 個の連続するヌクレオチド塩基を含むか, これからなるか, または本質的にこれからなるオリゴヌクレオチド配列から選択される。

10

【0268】

別の好ましい態様においては, 本発明は, 配列番号 60, 配列番号 61, 配列番号 62, 配列番号 63, 配列番号 64, 配列番号 65, 配列番号 66, 配列番号 67, 配列番号 68, 配列番号 69, 配列番号 70, 配列番号 71, 配列番号 72, 配列番号 73, 配列番号 74, 配列番号 75, 配列番号 76, 配列番号 77, 配列番号 78, 配列番号 79, 配列番号 80, 配列番号 81, 配列番号 82, 配列番号 83, 配列番号 84, 配列番号 85, 配列番号 86, 配列番号 87, 配列番号 88, 配列番号 89, 配列番号 90, 配列番号 91, 配列番号 92, 配列番号 93, 配列番号 94, 配列番号 95, 配列番号 96, 配列番号 97, 配列番号 98, 配列番号 99, 配列番号 100, 配列番号 101, 配列番号 102, 配列番号 103, 配列番号 104, 配列番号 105, 配列番号 106, 配列番号 107, 配列番号 108, 配列番号 109, 配列番号 110, 配列番号 111, 配列番号 112, 配列番号 113, 配列番号 114, 配列番号 115, 配列番号 116, 配列番号 117 および配列番号 118 からなる群より選択されるポリペプチドをコードする核酸配列の, 約 10-30 個の, より好ましくは 15-25 個の連続するヌクレオチド塩基を含むか, これからなるか, 本質的にこれからなる, 単離された, 濃縮されたまたは精製された核酸分子を含む。

20

【0269】

本発明の配列を用いてアンチセンスオリゴヌクレオチドを設計することができる。このようなアンチセンスオリゴヌクレオチドを, 標的プロテアーゼを発現する細胞に投与して, 標的 RNA または蛋白質のレベルを内部対照 RNA または蛋白質のレベルと比較することができる。アンチセンスオリゴヌクレオチドを用いて得られる結果はまた, 適当な対照オリゴヌクレオチドを用いて得られる結果と比較することができる。好ましい対照オリゴヌクレオチドは, 試験オリゴヌクレオチドとほぼ同じ長さのオリゴヌクレオチドである。標的 RNA または蛋白質のレベルを減少させるアンチセンスオリゴヌクレオチドを選択する。

30

【0270】

オリゴヌクレオチドは, DNA または RNA, またはこれらのキメラ混合物または誘導体または改変体であることができ, 一本鎖でも二本鎖でもよい。オリゴヌクレオチドは, 塩基成分, 糖成分, またはホスフェート骨格において修飾して, 例えば, 分子の安定性またはハイブリダイゼーションを改良することができる。オリゴヌクレオチドは, 結合した他の基, 例えばペプチド (例えば, インビボで宿主細胞レセプターにターゲティングするため), または細胞膜 (例えば, Letsinger et al. (1989) Proc. Natl. Acad. Sci. U. S. A. 86:6553-6556; Lemaître et al. (1987) Proc. Natl. Acad. Sci. USA 84:648-652; PCT 国際公開 WO 88/09810 (1988 年 12 月 15 日公開) を参照), または血液脳関門 (例えば, PCT 国際公開 WO 89/10134 (1988 年 4 月 25 日公開) を参照) を越える輸送を容易にする試薬, ハイブリダイゼーション励起性切断試薬 (例えば, Krol et al. (1988) Bio Techniques 6:958-976 を参照) またはインターカレート剤 (例えば, Zon (1988) Pharm. Res. 5:539-549 を参照) を含むことができる。この目的

40

50

のために、オリゴヌクレオチドを他の分子、例えば、ペプチド、ハイブリダイゼーション
励起性架橋剤、輸送剤、ハイブリダイゼーション励起性切断剤等とコンジュゲートさせる
ことができる。

【0271】

アンチセンスオリゴヌクレオチドは、5-フルオロウラシル、5-ブロモウラシル、5-
クロロウラシル、5-ヨードウラシル、ヒポキサンチン、キサンチン、4-アセチルシト
シン、および5-(カルボキシヒドロキシエチル)ウラシルなどの成分から選択される少
なくとも1つの修飾塩基成分を含むことができる。アンチセンスオリゴヌクレオチドはま
た、限定されないが、アラビノース、2-フルオロアラビノース、キシロースおよびヘ
キソースを含む群から選択される少なくとも1つの修飾糖成分を含んでいてもよい。

10

【0272】

さらに別の態様においては、アンチセンスオリゴヌクレオチドは、ホスホロチオエート、
ホスホロジチオエート、ホスホルアミドチオエート、ホスホルアミデート、ホスホルジア
ミデート、メチルホスホネート、アルキルホスホトリエステル、およびホルムアセタール
またはこれらの類似体からなる群より選択される少なくとも1つの修飾ホスフェート骨格
を含む(米国特許5,176,996;5,264,564;および5,256,775
も参照)。

【0273】

さらに別の態様においては、アンチセンスオリゴヌクレオチドは α -アノマーオリゴヌク
レオチドである。 α -アノマーオリゴヌクレオチドは、通常の β -ユニットとは異なり、
相補的RNAと、鎖が互いに平行に走る特別の二本鎖ハイブリッドを形成する(Ga
u t i e r e t a l . (1 9 8 7) N u c l . A c i d s R e s . 1 5 : 6 6 2 5
- 6 6 4 1) 。 オリゴヌクレオチドは2'-O-メチルリボヌクレオチド(I n o u e
e t a l . (1 9 8 7) N u c l . A c i d s R e s . 1 5 : 6 1 3 1 - 6 1 4 8)
, またはキメラRNA-DNA類似体(I n o u e e t a l . (1 9 8 7) F E B S
L e t t . 2 1 5 : 3 2 7 - 3 3 0) である。

20

【0274】

また適しているものは、ポリセリン、ポリトレオニン等のポリペプチドであるペプチジル
核酸であり、これには側鎖の位置で核酸(T, A, G, C, U)で置換された種々のアミ
ノ酸を含むコポリマーが含まれる。このようなポリマーの鎖は、天然のDNA/RNAと
同じ様式で相補塩基を介してハイブリダイズすることができる。あるいは、本発明のアン
チセンス構築物は、例えば、細胞内で転写されたときに本発明のプロテアーゼポリペプチ
ドをコードする細胞mRNAの少なくとも独特の部分に相補的なRNAを生成する発現プ
ラスミドまたはベクターとして輸送することができる。

30

【0275】

領域配列をコードするプロテアーゼポリペプチドに相補的なアンチセンスヌクレオチドを
用いることができるが、転写された非翻訳領域に相補的なものが最も好ましい。

【0276】

別の好ましい態様においては、遺伝子置換の方法が記載される。本明細書において用いる
場合、"遺伝子置換"とは、インビボで発現しうる核酸配列を動物に供給し、このことに
よりその動物に欠失しているかまたは不完全な内因性遺伝子の機能を提供することを意味
する。

40

【0277】

医薬処方および投与経路

本明細書に記載される化合物、例えば本発明のプロテアーゼポリペプチド、アンチセンス
分子、リボザイム、および本発明のプロテアーゼポリペプチドの活性を調節する任意の他
の化合物は、それ自体で、または組み合わせ療法におけるように、化合物が他の活性成分
と混合されているか、または適当な担体または賦形剤と混合されている医薬組成物中でヒ
ト患者に投与することができる。本発明の化合物の処方および投与の手法は"Remin
g t o n ' s P h a r m a c e u t i c a l S c i e n c e s , " M a c k P u b

50

lishing Co., Easton, PAの最新版に見いだすことができる。

【0278】

A. 投与経路：

投与の適当な経路には、例えば、経口、直腸、経粘膜、または腸投与；非経口輸送、例えば筋肉内、皮下、静脈内、骨髄内注入、ならびに鞘内、直接心室内、腹膜内、鼻腔内、または眼内注射が含まれる。

【0279】

あるいは、化合物を全身ではなく局所的に投与してもよく、これには、例えば、化合物を、しばしばデポ製剤または徐放製剤として直接固体腫瘍に注射することが含まれる。

【0280】

さらに、薬物はターゲティングされたドラッグデリバリーシステムにおいて、例えば腫瘍特異的抗体により被覆されたリポソーム中で投与してもよい。リポソームは腫瘍にターゲティングされ、選択的に取り込まれるであろう。

【0281】

B. 組成物／処方：

本発明の医薬組成物は、当該技術分野においてよく知られる方法、例えば、限定されないが、慣用の混合、溶解、顆粒化、糖衣作成、研和、乳化、カプセル封入、捕捉、または凍結乾燥により製造することができる。

【0282】

すなわち、本発明にしたがって使用するための医薬組成物は、活性化合物を薬剤として使用することができる製品に加工することを容易にする賦形剤および補助剤を含む、1またはそれ以上の生理学的に許容しうる担体を用いて、慣用の方法で製剤することができる。適切な処方は、選択される投与経路に依存する。

【0283】

注射用には、本発明の薬剤を水性溶液、好ましくはハंकス溶液、リンゲル溶液、または生理的食塩緩衝液等の生理学的に適合性の緩衝液中で処方することができる。経粘膜投与用には、浸透すべき障壁に適した浸透剤が処方に用いられる。そのような浸透剤は当該技術分野において一般に知られている。

【0284】

経口投与のためには、化合物を当該技術分野においてよく知られる薬学的に許容しうる担体と混合することにより化合物を容易に処方することができる。そのような担体は、本発明の化合物を、治療すべき患者による経口摂取のための錠剤、丸薬、糖衣剤、カプセル、液体、ゲル、シロップ、スラリー、懸濁液等として処方することを可能とする。適当な担体には、特に、ラクトース、ショ糖、マンニトール、またはソルビトール等の糖類；トウモロコシデンプン、小麦デンプン、米デンプン、およびジャガイモデンプン等のセルロース製品、ゼラチン、トラガカントゴム、メチルセルロース、ヒドロキシプロピルメチルセルロース、カルボキシメチルセルロースナトリウム、および／またはポリビニルピロリドン（PVP）等の増量剤などの賦形剤が含まれる。所望の場合には、架橋されたポリビニルピロリドン、寒天、またはアルギン酸またはその塩、例えばアルギン酸ナトリウム等の崩壊剤を加えてもよい。

【0285】

糖衣剤のコアは、適当なコーティングとともに供給される。この目的のためには、濃縮された糖溶液を用いることができる。これは、アラビアゴム、タルク、ポリビニルピロリドン、カルボポールゲル、ポリエチレングリコール、および／または二酸化チタン、ラッカー溶液、および適当な有機溶媒または溶媒混合物を任意に含むことができる。識別のため、あるいは活性化合物の用量の異なる組合せを特徴づけるため、染料または色素を錠剤または糖衣剤コーティングに添加してもよい。

【0286】

経口で 사용할ことができる医薬製剤は、ゼラチンから作成されるプッシュフィットカプセル、ならびにゼラチンおよびグリセロール、ソルビトール等の可塑剤から作成される密

10

20

30

40

50

封軟カプセルを含む。プッシュフィットカプセルは、活性成分を、ラクトース等の増量剤、デンプン等の結合剤、および／またはタルクおよびステアリン酸マグネシウム等の潤滑剤、さらに任意に安定剤との混合物中に含むことができる。軟カプセルにおいては、活性化合物は脂肪油、流動パラフィン、または液体ポリエチレングリコール等の適当な液体中に溶解または懸濁することができる。さらに安定剤を添加してもよい。経口投与用のすべての処方は、そのような投与に適当な用量で調製すべきである。

【0287】

口内投与のためには、組成物は、慣用的な方法で錠剤またはトローチ剤の形にすることができる。

【0288】

吸入による投与用には、本発明に従って用いられる化合物は、噴射剤、例えば、ジクロロジフルオロメタン、トリクロロフルオロメタン、ジクロロテトラフルオロエタン、二酸化炭素または他の適当な気体を用いて、加圧されたパックまたはネブライザーからエアゾルスプレイの形状で便利に輸送される。加圧されたエアゾルの場合、用量単位は計量された量を送達するべく備えられたバルブにより調節することができる。例えば吸入器または注入器において使用するためのゼラチン製のカプセルおよびカートリッジは、化合物の粉末混合物と、ラクトースまたはデンプン等の適当な粉末基剤とを含むよう処方することができる。

10

【0289】

化合物は、例えばボーラス注射または連続注入による非経口投与用に処方することができる。注射用の処方は、単位用量にて、例えばアンプルにて、あるいは添加された保存料と共に多用量容器中で提供することができる。組成物は油性または水性のベヒクル中で、懸濁液、溶液、または乳濁液等の形状をとることができ、懸濁剤、安定剤および／または分散剤等の製剤物質を含んでいてもよい。

20

【0290】

非経口投与用の薬剤処方は、水溶性の形態の活性化合物の水性溶液を含む。さらに、活性化合物の懸濁液は、適当な油性の注入用懸濁液として調製することができる。適切な親油性溶媒またはベヒクルには、ゴマ油等の脂肪油、オレイン酸エチルまたはトリグリセリド等の合成脂肪酸エステル、またはリポソーム等を含む。水性の注射用懸濁液は、カルボキシメチルセルロースナトリウム、ソルビトール、またはデキストラン等の、懸濁液の粘度を増加させる物質を含んでいてもよい。任意に、懸濁液はまた、高度に濃縮された溶液の調製を可能にする、当該化合物の溶解性を増加させる適当な安定剤または薬剤を含んでいてもよい。

30

【0291】

あるいは、活性成分は粉体の形態であって、使用前に適当なベヒクル、例えば発熱物質を含まない滅菌水を用いて構成することができる。

【0292】

化合物はまた、例えばカカオバターまたは他のグリセリド等の慣用の坐剤基剤を用いて、坐剤または停留浣腸等の直腸用組成物に処方することができる。

【0293】

上述した処方に加えて、化合物はまたデポ製剤として処方することができる。そのような長時間作用性の処方、埋込み（例えば皮下または筋肉内への）によるか、または筋肉内注射により投与することができる。すなわち、例えば、化合物は、適当な高分子性または疎水性物質と共に（例えば許容される油剤中の乳濁液として）、イオン交換樹脂と共に、溶けにくい塩等の溶けにくい誘導体として、処方することができる。

40

【0294】

本発明の疎水性化合物のための薬学的担体は、ベンジルアルコール、非極性界面活性剤、水混和性有機ポリマーおよび水性相を含む共溶媒系であってもよい。共溶媒系はV P D共溶媒系であってもよい。V P Dは、3%（w/v）ベンジルアルコール、8%（w/v）非極性界面活性剤ポリソルベート80、および65%（w/v）ポリエチレングリコール

50

300を純粋エタノール中に作成した溶液である。VPD共溶媒系（VPD：D5W）は、VPDを5%デキストロースの水溶液中に1：1で希釈したものである。この共溶媒系は疎水性化合物をよく溶解し、それ自体、全身投与に際して低い毒性を示す。本来、共溶媒系の比率は、その溶解性および毒性特性を破壊することなく相当変化させることができる。さらに、共溶媒成分の種類も変化させることができる。例えば、他の低毒性非極性界面活性剤をポリソルベート80の代わりに用いることができ、ポリエチレングリコールの分画サイズは様々でありうる。他の生体適合性ポリマー、例えばポリビニルピロリドン、ポリエチレングリコールの代わりに用いることができ、他の糖または多糖類をデキストロースの代わりに用いることができる。

【0295】

あるいは、疎水的医薬化合物のための他の輸送系を用いてもよい。リポソームおよび乳剤は、疎水的薬剤のための輸送用ベヒクルまたは担体の例としてよく知られている。さらに、ある種の有機溶媒、例えばジメチルスルホキシドもまた用いることができるが、しばしば毒性がより高くなる。さらに、化合物は、持続放出系、例えば治療薬剤を含む固体疎水性ポリマーの準透過性マトリックスを用いて輸送することができる。種々の持続放出材料が当業者にはよく知られている。持続放出カプセルはその化学的性質に応じて、数週間から100日を越える期間、化合物を放出する。治療薬剤の化学的性質および生物学的安定性に応じて、さらに別の蛋白質安定化戦略を用いてもよい。

【0296】

医薬組成物はまた、適当な固体またはゲル相の担体または賦形剤を含んでいてもよい。そのような担体または賦形剤の例には、限定されないが、炭酸カルシウム、リン酸カルシウム、種々の糖、澱粉、セルロース誘導体、ゼラチン、およびポリエチレングリコール等の高分子が含まれる。

【0297】

本発明のプロテアーゼ調節化合物の多くは、薬学的に適合性のカウンターイオンとの塩として提供される。薬学的に適合性の塩は、多くの酸、例えば、限定されないが、塩酸、硫酸、酢酸、乳酸、酒石酸、リンゴ酸、クエン酸等を用いて形成することができる。塩は、水性または他のプロトン性溶媒において、対応する遊離塩基の形よりもより溶解性である傾向にある。

【0298】

C. 適切な投与量：

本発明において使用するのに適した医薬組成物には、活性成分がその意図される目的を達成するのに有効な量で含まれている組成物が含まれる。より詳細には、治療上有効量とは、疾病の症状を予防、緩和または改善するのに、または治療している被験者の生存を長くするのに有効な化合物の量を意味する。本発明の化合物の治療上有効量の決定は、特に本明細書に提供される詳細な開示に鑑みて、十分に当業者の能力の範囲内である。

【0299】

本発明の方法において用いられる任意の化合物について、治療上有効な用量は、最初は細胞培養アッセイから見積もることができる。例えば、動物モデルにおいて、培養細胞において決定された IC_{50} （すなわちプロテアーゼ活性の最大阻害の半分を達成する試験化合物の濃度）を含む循環濃度範囲を達成するような用量を処方することができる。そのような情報は、ヒトにおける有用な用量のさらに正確な決定のために用いることができる。

【0300】

本明細書に記載される化合物の毒性および治療有効性は、培養細胞または実験動物における標準的な薬理学的な方法、例えば LD_{50} （集団の50%に致死的な用量）および ED_{50} （集団の50%に治療上有効な用量）を決定することにより、決定することができる。毒性と治療上有効性の用量の比は治療指数であり、 LD_{50} と ED_{50} の比率として表すことができる。高い治療指数を示す化合物が好ましい。これらの培養細胞アッセイおよび動物研究から得られるデータは、ヒトにおいて用いるための範囲の投与量を処方するために用いることができる。このような化合物の投与量は、好ましくは、 ED_{50} を含み毒性

10

20

30

40

50

がほとんどまたは全くない循環濃度の範囲内にある。投与量は、用いる投与形態および用いる投与経路により、この範囲内で様々でありうる。正確な処方、投与経路、および投与量は、個々の医師が、患者の状態を考慮して選択することができる（例えば、Finglet et al. 1975, "The Pharmacological Basis of Therapeutics", Ch. 1, p. 1を参照）。

【0301】

投与量および間隔は、個々に、活性成分がプロテアーゼ調節効果を維持するのに十分な血漿レベル、すなわち最小有効濃度（MEC）を与えるよう調節することができる。MECは、各化合物について異なるが、インビトロのデータ、例えば、限定されないが、本明細書に記載されるアッセイを用いて、プロテアーゼの50－90%の阻害を達成するのに必要な濃度から見積もることができる。MECを達成するのに必要な投与量は、個々の特性および投与経路に依存するであろう。しかし、血漿濃度はHPLCアッセイまたはバイオアッセイを用いて決定することができる。

10

【0302】

投与間隔もまた、MEC値を用いて決定することができる。化合物は、10－90%の時間、好ましくは30－90%の時間、最も好ましくは50－90%の時間、MECより高い血漿レベルを維持する投与計画を用いて投与すべきである。

【0303】

局所投与または選択的取り込みの場合には、薬剤の有効な局所濃度は血漿濃度とは関係ないであろう。

20

【0304】

投与される特定の組成物の量は、もちろん、治療中の患者、患者の体重、苦痛の激しさ、投与方法、および担当医師の判断に依存するであろう。

【0305】

D. 包装：

組成物は、所望の場合には、活性成分を含む1またはそれ以上の単位用量形を含んでいてもよいパックまたはディスペンサー装置中で提供することができる。パックは、例えば、ブリストパックなどの金属またはプラスチック箔を含むことができる。パックまたはディスペンサー装置には、投与の指示が添付されていてもよい。パックまたはディスペンサー装置はまた、薬剤の製造、使用、または販売を規制する政府機関によって規定された形式の、容器に付随した注意書が添付されていてもよく、その注意書はヒトまたは獣医学的投与用のポリヌクレオチドの形状の当該機関による承認を反映するものである。そのような注意書は、例えば米国食品医薬品局により処方箋調剤薬として承認されたラベルによるものか、または承認された製品に差込まれたものでもよい。適合した薬学的担体中に処方された、本発明の化合物を含む組成物もまた製造され、適当な容器内に配置され、さらに指示された条件による処置のためにラベルを付すことができる。ラベル上に示される適切な状態としては、腫瘍の治療、新脈管形成の阻害、線維症、糖尿病等の治療が挙げられる。

30

【0306】

機能的誘導体

本発明はまた、本発明のポリペプチドまたは核酸の機能的誘導体を提供する。"機能的誘導体"は、本発明のポリペプチドまたは核酸の"化学的誘導体"、"フラグメント"または"変種"を意味し、これらの用語は以下に定義される。機能的誘導体は、蛋白質の機能の少なくとも一部、例えば蛋白質に特異的な抗体との反応性、非触媒ドメインにより媒介される酵素活性または結合活性を保持しており、これにより本発明にしたがう有用性を有する。遺伝コードの縮重のため、多くの異なる核酸配列が同じアミノ酸配列をコードすることができることは当該技術分野においてよく知られている。同じく、アミノ酸を保存的に変更して、元の機能性を保持する蛋白質またはポリペプチドを得ることができることも当該技術分野においてよく知られている。いずれの場合においても、すべての順列が本明細書の開示によりカバーされることが意図される。

40

50

【0307】

本明細書に記載される単離された核酸分子の機能的等価物も本発明の範囲内に含まれる。遺伝コードの縮重のため、あるコドンと同じアミノ酸を規定する他のコドンで置き換え、同じ蛋白質を生じさせることができる。既知のアミノ酸は、メチオニンおよびトリプトファンを除き、2以上のコドンによりコードすることができるため、核酸配列は実質的に様々でありうる。すなわち、本発明の遺伝子の一部または全部を合成して、配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、配列番号21、配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、配列番号28、配列番号29、配列番号30、配列番号31、配列番号32、配列番号33、配列番号34、配列番号35、配列番号36、配列番号37、配列番号38、配列番号39、配列番号40、配列番号41、配列番号42、配列番号43、配列番号44、配列番号45、配列番号46、配列番号47、配列番号48、配列番号49、配列番号50、配列番号51、配列番号52、配列番号53、配列番号54、配列番号55、配列番号56、配列番号57、配列番号58、および配列番号59に記載される配列からなる群より選択される核酸配列と有意に異なる核酸配列を得ることができる。しかし、これによりコードされるアミノ酸配列は保存される。

10

【0308】

さらに、核酸配列は、配列番号1、配列番号2、配列番号3、配列番号4、配列番号5、配列番号6、配列番号7、配列番号8、配列番号9、配列番号10、配列番号11、配列番号12、配列番号13、配列番号14、配列番号15、配列番号16、配列番号17、配列番号18、配列番号19、配列番号20、配列番号21、配列番号22、配列番号23、配列番号24、配列番号25、配列番号26、配列番号27、配列番号28、配列番号29、配列番号30、配列番号31、配列番号32、配列番号33、配列番号34、配列番号35、配列番号36、配列番号37、配列番号38、配列番号39、配列番号40、配列番号41、配列番号42、配列番号43、配列番号44、配列番号45、配列番号46、配列番号47、配列番号48、配列番号49、配列番号50、配列番号51、配列番号52、配列番号53、配列番号54、配列番号55、配列番号56、配列番号57、配列番号58、および配列番号59に記載される配列、またはその誘導体からなる群より選択される核酸の式またはその誘導体の5'-末端および／または3'-末端で少なくとも1つのヌクレオチドを付加、欠失または置換することにより得られるヌクレオチド配列を含んでいてもよい。その付加、欠失または置換が、ヌクレオチド配列によりコードされる、配列番号60、配列番号61、配列番号62、配列番号63、配列番号64、配列番号65、配列番号66、配列番号67、配列番号68、配列番号69、配列番号70、配列番号71、配列番号72、配列番号73、配列番号74、配列番号75、配列番号76、配列番号77、配列番号78、配列番号79、配列番号80、配列番号81、配列番号82、配列番号83、配列番号84、配列番号85、配列番号86、配列番号87、配列番号88、配列番号89、配列番号90、配列番号91、配列番号92、配列番号93、配列番号94、配列番号95、配列番号96、配列番号97、配列番号98、配列番号99、配列番号100、配列番号101、配列番号102、配列番号103、配列番号104、配列番号105、配列番号106、配列番号107、配列番号108、配列番号109、配列番号110、配列番号111、配列番号112、配列番号113、配列番号114、配列番号115、配列番号116、配列番号117および配列番号118に記載される配列からなる群より選択されるアミノ酸配列を変更しない限り、任意のヌクレオチドまたはポリヌクレオチドをこのように用いることができる。例えば、本発明は、本発明の核酸配列またはその誘導体の5'-末端に開始コドンとしてATGを付加することにより、または本発明のヌクレオチド配列またはその誘導体の3'-末端に終止コドンとしてTTA、TAGまたはTGAを付加することにより得られる任意の核酸配列を含むことを意図する。さらに、本発明の核酸分子は、必要に応じて、これらの5'-末端および／または

20

30

40

50

3'-末端に付加された制限エンドヌクレアーゼ認識部位を有することができる。

【0309】

所定の核酸配列のそのような機能的変更により、これに融合した外来核酸配列によりコードされる異種蛋白質の分泌および／またはプロセッシングが促進される機会が与えられる。したがって、遺伝コードにより許容される本発明のプロテアーゼ遺伝子およびそのフラグメントのヌクレオチド配列のすべての変種は本発明に含まれる。

【0310】

さらに、コドンを削除するかまたは1またはそれ以上のコドンで縮重コドン以外のコドンと置換して、構造的に改変されているが、未改変核酸分子により産生されるポリペプチドと実質的に同じ有用性または活性を有するポリペプチドを製造することが可能である。当該技術分野において認識されているように、2つのポリペプチドは機能的に同等であり、これらの核酸分子の間の相違が遺伝コードの縮重に関連しない場合であっても、これらの製造を生じさせる2つの核酸分子も機能的に同等である。

【0311】

複合体の“化学的誘導体”は、通常は蛋白質の一部ではない追加の化学的成分を含む。蛋白質またはペプチドの共有結合修飾は、本発明の範囲内に含まれる。そのような修飾は、以下に記載するように、ペプチドの標的アミノ酸残基を選択された側鎖または末端残基と反応しうる有機誘導化剤と反応させることにより、分子中に導入することができる。

【0312】

システイン残基は、最も一般的にはアルファーハロアセテート（および対応するアミン）、例えばクロロ酢酸またはクロロアセトアミドと反応させて、カルボキシメチルまたはカルボキシアミドメチル誘導体を得る。システイン残基は、プロモトリフルオロアセトン、クロロアセチルホスフェート、N-アルキルマレイミド、3-ニトロ-2-ピリジルジスルフィド、メチル2-ピリジルジスルフィド、p-クロロ水銀ベンゾエート、2-クロロ水銀-4-ニトロフェノール、またはクロロ7-ニトロベンゾ-2-オキサー1,3-ジアゾールと反応させることにより誘導化する。

【0313】

ヒスチジン残基は、pH 5.5-7.0でジエチルプロカーボネートと反応させることにより誘導化することができる。これは、この薬剤がヒスチジンの側鎖に比較的特異的であるためである。臭化パラプロモフェナシルもまた有用である。反応は、好ましくは、0.1 M カコジル酸ナトリウム中でpH 6.0で行う。

【0314】

リジンおよびアミノ末端残基はコハク酸または他のカルボン酸無水物と反応させる。これらの薬剤による誘導化は、リジン残基の電荷を逆転させる効果を有する。1級アミン含有残基を誘導化するための他の適当な試薬には、イミドエステル、例えば、メチルピコリンイミデート；ピリドキサルホスフェート；ピリドキサル；クロロホウ化水素；トリニトロベンゼンスルホン酸；O-メチルイソウレア；2,4-ペンタンジオン；およびトランスアミナーゼに触媒されるグリオキシレートとの反応が含まれる。

【0315】

アルギニン残基は慣用的な試薬の1つまたはいくつかと反応させることにより修飾する。これには、例えば、フェニルグリオキサール、2,3-ブタンジオン、1,2-シクロヘキサジオン、およびニンヒドリンがある。アルギニン残基の誘導化には、グアニジン官能基の高いpKaのため、反応をアルカリ条件中で行うことが必要である。さらに、これらの試薬はリジンの基ならびにアルギニンアルファアミノ基とも反応することができる。

【0316】

チロシン残基は、芳香族性ジアゾニウム化合物またはテトラニトロメタンと反応させることにより光学的標識を導入するための修飾のよく知られる標的である。最も一般的には、N-アセチルイミダゾールおよびテトラニトロメタンを用いて、それぞれO-アセチルチロシル種および3-ニトロ誘導体を生成する。

【0317】

カルボキシル側鎖（アスパラギン酸およびグルタミン酸）は、カルボジイミド（ $R'-N-C-N-R'$ ），例えば1-シクロヘキシル-3-（2-モルホリニル（4-エチル）カルボジイミドまたは1-エチル-3-（4-アゾニア-4，4-ジメチルペンチル）カルボジイミドと反応させることにより，選択的に修飾する。さらに，アスパラギン酸およびグルタミン酸残基は，アンモニウムイオンと反応させることにより，アスパラギンおよびグルタミン残基に変換する。

【0318】

グルタミンおよびアスパラギン残基は，しばしば脱アミド化して，対応するグルタミン酸およびアスパラギン酸残基に変換する。あるいは，これらの残基をより穏和な酸性条件下で脱アミド化する。これらの残基のいずれの形も本発明の範囲内である。

10

【0319】

二官能性薬剤による誘導化は，例えば，蛋白質の成分ペプチドを互いに，または複合体中の他の蛋白質と，または水不溶性支持体マトリクスと，または他の高分子担体と架橋するのに有用である。一般に用いられる架橋薬剤としては，例えば，1，1-ビス（ジアゾアセチル）-2-フェニルエタン，グルタルアルデヒド，N-ヒドロキシスクシンイミドエステル，例えば，4-アジドサリチル酸のエステル，ホモ二官能性イミドエステル（3，3'-ジチオビス（スクシンイミジルプロピオネート）等のジスクシンイミジルエステルを含む），および二官能性マレイミド，例えばビス-N-マレイミド-1，8-オクタンが挙げられる。メチル-3-[p-アジドフェニル]ジチオールプロピオイミデート等の誘導化薬剤により，光の存在下で架橋を形成しうる光活性化可能な中間体得られる。あるいは，反応性の水不溶性マトリクス，例えば臭化シアン活性化炭水化物および米国特許3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; および4,330,440に記載される反応性基質を蛋白質固定化のために用いることができる。

20

【0320】

他の修飾としては，プロリンおよびリジンのヒドロキシル化，セリンまたはトレオニン残基のヒドロキシル基のリン酸化，リジン，アルギニン，およびヒスチジン側鎖のアルファアミノ基のメチル化（Creighton, T. E., *Proteins: Structure and Molecular Properties*, W. H. Freeman & Co., San Francisco, pp. 79-86 (1983)），N末端アミンのアセチル化，および場合によりC末端カルボキシル基のアミド化が挙げられる。

30

【0321】

そのような誘導化された成分は，安定性，溶解性，吸収，生物学的半減期等を改良することができる。あるいは，これらの成分は，蛋白質複合体の望ましくない副作用を排除するかまたは緩和することができる。そのような効果を媒介しうる成分は，例えば，Remington's Pharmaceutical Sciences, 18th ed., Mack Publishing Co., Easton, PA (1990)に開示されている。

【0322】

”フラグメント”との用語は，これが由来する全長ポリペプチドより短い長さを有する蛋白質または複合体のアミノ酸配列に由来するポリペプチドを示すものとして用いられる。そのようなフラグメントは，例えば，全長蛋白質の蛋白質分解性切断により製造することができる。好ましくは，フラグメントは，蛋白質をコードするDNA配列を適当に改変して，C末端，N末端，および／または天然配列中の1またはそれ以上の部位において1またはそれ以上のアミノ酸を除去することにより，組換えにより得る。蛋白質のフラグメントは，本明細書に記載されるように，シグナル伝達を調節するように作用する物質のスクリーニングに有用である。そのようなフラグメントは，天然の複合体の1またはそれ以上の特徴的部分を保持しているであろうことが理解される。そのような保持される特徴の例としては，触媒的活性；基質特異性；無傷の細胞中における他の分子との相互作用；制御機能；または天然の複合体またはそのエピトープに対する抗体との結合が挙げられる。

40

50

【0323】

本発明の範囲内に入ることが意図される別の機能的誘導体は、天然のポリペプチドに対して、1またはそれ以上のアミノ酸を欠失しているか、追加のまたは置換されたアミノ酸を含む”変種”ポリペプチドである。変種は、蛋白質のDNAコーディング配列を適当に改変して、C末端、N末端、および／または天然配列中の1またはそれ以上の部位において1またはそれ以上のアミノ酸のためのコドンが付加、除去、および／または改変することにより、天然に生ずる複合体成分から誘導することができる。追加、置換および／または追加アミノ酸を有するそのような変種は、上述するように、天然の蛋白質の1またはそれ以上の特徴的部分を保持していることが理解される。

【0324】

アミノ酸残基が除去、挿入および／または置換されている蛋白質の機能的誘導体は、当業者によく知られる標準的な手法を用いて製造することができる。例えば、機能的誘導体の改変された成分は、改変されたコード配列を生成するように、DNAコード配列中のヌクレオチドを修飾する部位特異的突然変異手法 (Adelman et al., 1983, DNA 2:183に例示されている) を用いて、その後上述したような手法を用いてこの組換えDNAを原核生物または真核生物宿主細胞中で発現させることにより、製造することができる。あるいは、アミノ酸が削除、挿入および／または置換されている蛋白質は、当該技術分野においてよく知られる方法を用いて、直接化学合成により便利に製造することができる。蛋白質の機能的誘導体は、典型的には天然の蛋白質と同じ質の生物学的活性を示す。

【0325】

表およびその記載

本発明は、ゲノム配列のデータベースにおいて同定された新規プロテアーゼを記述する。結果は以下の4つの表に示される。

【0326】

表1は、各遺伝子の名前、各遺伝子の分類、配列中のオープンリーディングフレームの位置、および対応するペプチドの長さを表す。データは、左から右に、以下のものを表す: ”遺伝子名”, ”ID#na”, ”ID#aa”, ”FL/Cat”, ”スーパーファミリー”, ”グループ”, ”ファミリー”, ”NA長さ”, ”ORF開始”, ”ORF終了”, ”ORF長さ”, および ”AA長さ”。 ”遺伝子名” は、プロテアーゼ酵素をコードする配列に与えられた名前を表す。各遺伝子は、 ”SGPr” 名称およびその後の番号により表される。SGPr名は、通常は、1つの連続配列 (”コンティグ”) に構築された多数の重複配列を表す。 ”ID#na” および ”ID#aa” は、本明細書において各核酸およびアミノ酸配列に与えられた同定番号を表す。 ”FL/Cat” は、遺伝子の長さを表し、FLは全長を、 ”Cat” は触媒ドメインのみが示されていることを表す。このカラム中の ”Partial” とは、配列が部分的触媒ドメインをコードすることを示す。 ”スーパーファミリー” とは、遺伝子がプロテアーゼであるか否かを識別する。 ”グループ” および ”ファミリー” は、配列ホモロジーにより定義されるプロテアーゼの分類を表す。 ”NA長さ” は、対応する核酸配列の長さをヌクレオチドで表す。 ”ORF開始” は、オープンリーディングフレームの開始ヌクレオチドを表す。 ”ORF終了” は、オープンリーディングフレームの終止コドンを含む最後のヌクレオチドを表す。 ”ORF長さ” は、オープンリーディングフレームの長さ (終止コドンを含む) をヌクレオチドで表す。 ”AA長さ” は、対応する核酸配列においてコードされるペプチドの長さをアミノ酸で表す。

【0327】

【表1】

遺伝子名	ID# na	ID# aa	FL/ Cat	スーパーファミ リ	グループ	ファミリー	NA 長 さ	ORF 開始	ORF 終了	ORF 長さ	AA 長 さ
SGPr397	1	60	FL	プロテアーゼ	カルボキシペプチ ダーゼ	Zn カルボキシ ペプチダーゼ	948	1	948	948	315
SGPr413	2	61	FL	プロテアーゼ	カルボキシペプチ ダーゼ	Zn カルボキシ ペプチダーゼ	1125	1	1125	1125	374
SGPr404	3	62	FL	プロテアーゼ	カルボキシペプチ ダーゼ	Zn カルボキシ ペプチダーゼ	1590	1	1590	1590	529
SGPr536_1	4	63	FL	プロテアーゼ	システイン	パパイン	1404	1	1404	1404	467
SGPr414	5	64	FL	プロテアーゼ	システイン	UCH2b	10062	1	10062	10062	3353
SGPr430	6	65	FL	プロテアーゼ	システイン	UCH2b	2943	1	2943	2943	980
SGPr496_1	7	66	FL	プロテアーゼ	システイン	UCH2b	2862	1	2862	2882	953
SGPr495	8	67	FL	プロテアーゼ	システイン	UCH2b	2352	1	2352	2352	783
SGPr407	9	68	FL	プロテアーゼ	システイン	UCH2b	2259	1	2259	2259	752
SGPr453	10	69	FL	プロテアーゼ	システイン	UCH2b	2139	1	2139	2139	712
SGPr445	11	70	FL	プロテアーゼ	システイン	UCH2b	870	1	870	870	289
SGPr401_1	12	71	FL	プロテアーゼ	システイン	UCH2b	1101	1	1101	1101	366
SGPr408	13	72	FL	プロテアーゼ	システイン	UCH2b	3864	1	3864	3864	1287

10

20

30

40

【 0 3 2 8 】

【 表 2 】

SGPr480	14	73	FL	プロテアーゼ	システイン	UCH2b	4815	1	4815	4815	1604
SGPr431	15	74	FL	プロテアーゼ	システイン	UCH2b	3129	1	3129	3129	1042
SGPr429	16	75	FL	プロテアーゼ	システイン	UCH2b	3102	1	3102	3102	1033
SGPr503	17	76	FL	プロテアーゼ	システイン	UCH2b	1554	1	1554	1554	517
SGPr427	18	77	FL	プロテアーゼ	システイン	UCH2b	3372	1	3372	3372	1123
SGPrO92	19	78	FL	プロテアーゼ	メタロプロテアーゼ	PepM10	786	1	786	786	261
SGPr59	20	79	FL	プロテアーゼ	メタロプロテアーゼ	PepM10	1452	1	1452	1452	483
SGPr104_1	21	80	FL	プロテアーゼ	メタロプロテアーゼ	PepM13	2298	1	2298	2298	765
SGPr303	22	81	CAT	プロテアーゼ	メタロプロテアーゼ	PepM2	1257	1	1257	1257	418
SGPr402_1	23	82	FL	プロテアーゼ	セリン	ズブチラーゼ	2268	1	2268	2268	755
SGPr434	24	83	FL	プロテアーゼ	セリン	トリプシン	1176	1	1176	1178	391
SGPr446_1	25	84	CAT	プロテアーゼ	セリン	トリプシン	681	1	681	681	226
SGPr447	26	85	FL	プロテアーゼ	セリン	トリプシン	888	1	888	888	295
SGPr432_1	27	86	FL	プロテアーゼ	セリン	トリプシン	1887	1	1887	1887	628
SGPr529	28	87	FL	プロテアーゼ	セリン	トリプシン	831	1	831	831	276
SGPr4281	29	88	CAT	プロテアーゼ	セリン	トリプシン	858	1	858	858	285
SGPr425	30	89	FL	プロテアーゼ	セリン	トリプシン	1242	1	1242	1242	413
SGPr548	31	90	FL	プロテアーゼ	セリン	トリプシン	963	1	963	963	320

10

20

30

40

【 0 3 2 9 】

【 表 3 】

SGPr3g6	32	91	FL	プロテアーゼ	セリン	トリプシン	987	1	987	987	328
SGPr426	33	92	FL	プロテアーゼ	セリン	トリプシン	1278	1	1278	1278	425
SGPr552	34	93	CAT	プロテアーゼ	セリン	トリプシン	666	1	666	666	221
SGPr405	35	94	FL	プロテアーゼ	セリン	トリプシン	2847	1	2847	2847	948
SGPr485_1	36	95	FL	プロテアーゼ	セリン	トリプシン	1059	1	1059	1059	352
SGPr534	37	96	FL	プロテアーゼ	セリン	トリプシン	792	1	792	792	263
SGPr390	38	97	FL	プロテアーゼ	セリン	トリプシン	3387	1	3387	3387	1128
SGPr521	39	98	FL	プロテアーゼ	セリン	トリプシン	762	1	762	762	253
SGPr530_1	40	99	CAT	プロテアーゼ	セリン	トリプシン	816	1	816	816	271
SGPr520	41	100	FL	プロテアーゼ	セリン	トリプシン	1737	1	1737	1737	578
SGPr455	42	101	FL	プロテアーゼ	セリン	トリプシン	2913	1	2913	2913	970
SGPr507_2	43	102	FL	プロテアーゼ	セリン	トリプシン	798	1	798	798	265
SGPr559	44	103	FL	プロテアーゼ	セリン	トリプシン	1365	1	1365	1365	454
SGPr567	45	104	FL	プロテアーゼ	セリン	トリプシン	1614	1	1614	1614	537
SGPr479_1	46	105	FL	プロテアーゼ	セリン	トリプシン	961	1	981	981	326
SGPr489_1	47	106	CAT	プロテアーゼ	セリン	トリプシン	1671	1	1671	1671	556
SGPr465_1	48	107	CAT	プロテアーゼ	セリン	トリプシン	894	1	894	894	297
SGPr524_1	49	108	FL	プロテアーゼ	セリン	トリプシン	2553	1	2553	2553	850

10

20

30

40

【 0 3 3 0 】

【 表 4 】

SGPr422	50	109	FL	プロテアーゼ	セリン	トリプシン	1344	1	1344	1344	447
SGPr538	51	110	FL	プロテアーゼ	セリン	トリプシン	1374	1	1374	1374	457
SGPr527_1	52	111	FL	プロテアーゼ	セリン	トリプシン	2457	1	2457	2457	818
SGPr542	53	112	FL	プロテアーゼ	セリン	トリプシン	855	1	855	855	284
SGPr551	54	113	FL	プロテアーゼ	セリン	トリプシン	2409	1	2409	2409	802
SGPr451	55	114	FL	プロテアーゼ	セリン	トリプシン	1080	1	1080	1080	359
SGPr452_1	56	115	FL	プロテアーゼ	セリン	トリプシン	867	1	867	867	288
SGPr504	57	116	部分	プロテアーゼ	セリン	トリプシン	135	1	135	135	44
SGPr469	58	117	部分	プロテアーゼ	セリン	トリプシン	138	1	138	138	45
SGPr400	59	118	部分	プロテアーゼ	セリン	トリプシン	930	1	930	930	309

10

20

30

40

【0331】

表2は、本明細書に記載される遺伝子の以下の特徴を示す：染色体位置、一塩基多型（SNPs）、dbESTにおける表示、および反復領域。示されるデータは、左から右に、以下のとおりである：“遺伝子名”、“ID#na”、“ID#aa”、“FL/Cat”、“スーパーファミリー”、“グループ”、“ファミリー”、“染色体”、“SNPs

50

”, ”dbEST_ヒット”, および ”反復”。最初の7つのカラム（すなわち, ”遺伝子名”, ”ID#na”, ”ID#aa”, ”FL/Cat”, ”スーパーファミリー”, ”グループ”, ”ファミリー”）の内容は, 表1について上述したとおりである。 ”染色体” は, 遺伝子の細胞遺伝学的位置を表す。 ”SNPs” カラムの情報は, 候補一塩基多型 (SNPs) の核酸位置および縮重性を示す。これらのSNPsは, DNA配列をNCBI (<http://www.ncbi.nlm.nih.gov/SNP/snpblastByChr.html>) において維持されている一塩基多型のデータベースに対してblastnを行うことにより同定された。 ”dbESTヒット” は, 対応する遺伝子に対して少なくとも100bpの100%同一性を含む, ESTの公共のデータベース (dbEST, <http://www.ncbi.nlm.nih.gov/dbEST/index.html>) 中のエントリーの受託番号を示す。これらのESTはdbESTのblastnにより同定した。 ”反復” は, 複雑性が低く, いくつかの別々の遺伝子中に存在する約20bpの長さの短い配列の位置に関する情報を含む。

10

【0332】

【表5】

遺伝子名	ID #n a	ID# aa	FL/ Cat	スーパー ファミリー	グループ	ファミリー	染色体	SNPS	d6EST_ヒッ ト	反復
SGPr397	1	60	FL	プロテア ーゼ	カルボキ シペプチ ダーゼ	Znカルボキシ ペプチダーゼ	8q12		AV763490	
SGPr413	2	61	FL	プロテア ーゼ	カルボキ シペプチ ダーゼ	Znカルボキシ ペプチダーゼ	2q35		none	
SGPr404	3	62	FL	プロテア ーゼ	カルボキ シペプチ ダーゼ	Znカルボキシ ペプチダーゼ	10q26	ss1782198_allelePos=2 01,agaaggcctaygaagggg	AA045748 AA148684 AA047483	477ggagctgct gctgctgctggtg4 98
SGPr536 _1	4	63	FL	プロテア ーゼ	システイン	パパイン	1q35		AL542213 AL547246 AL552037	480gctgctgctg ctgctggtgcag5 01

10

20

30

40

【 0 3 3 3 】

【 表 6 】

SGPr414	5	64	FL	プロテア ーゼ	システイン	UCH2b	2p14	ss16542_allelePos=101, ctaccctagcygaggaaga	AU118237 AU131420 AU125083	2249accaccac caccaccaccatc accaccaccac2 280
SGPr430	6	65	FL	プロテア ーゼ	システイン	UCH2b	2q37	ss1534585_allelePos=5 1.tggaatarctcggac:rs105 5687_allelePos=51,tggta atccgkgtagagg	W87666 A10786108, BGS62864	
SGPr493 _1	7	66	FL	プロテア ーゼ	システイン	UCH2b	Xp11. 4	ss1029756_allelePos=1 01.agagaaataygagggtatt	AW851066, AW851065, AW851076	
SGPr495	8	67	FL	プロテア ーゼ	システイン	UCH2b	6q16		AL559960, AL530470, AL516184	
SGPr407	9	68	FL	プロテア ーゼ	システイン	UCH2b	2q37		none	
SGPr453	10	69	FL	プロテア ーゼ	システイン	UCH2b	12q23		BG722436 AI927881 BG771888	553gtagtaaaa agagaagtaaa5 72

10

20

30

40

【 0 3 3 4 】

【 表 7 】

SGPr445	11	70	FL	プロテア ーゼ	システイン	UCH2b	6q16	AL559960, AL530470, AL516184	
SGPr401 _1	12	71	FL	プロテア ーゼ	システイン	UCH2b	4q11	AU124898 AU134553 AI269069	
SGPr408	13	72	FL	プロテア ーゼ	システイン	UCH2b	11p15	BG741190 BF575498 BG170829	
SGPr480	14	73	FL	プロテア ーゼ	システイン	UCH2b	17q24	AU131748, AU120381, BG420766	
SGPr431	15	74	FL	プロテア ーゼ	システイン	UCH2b	4q31. 3	BG575871 BG113469 BG112979	
SGPr429	16	75	FL	プロテア ーゼ	システイン	UCH2b	1p36. 2	AL518266, BG681225, BG217186	

10

20

30

40

【 0 3 3 5 】

【 表 8 】

SGPr503	17	76	FL	プロテア ーゼ	システイン	UCH2b	12q24 .3	BG678894 BG476418 BE264732	1534gagtga gtctgaagaatg1 553
SGPr427	18	77	FL	プロテア ーゼ	システイン	UCH2b	17p13	BG831111 AW996553 BE614914	
SGPr092	19	78	FL	プロテア ーゼ	メタロプロ テアーゼ	PepM10	11p15	BS189720, AW966183, BG198356	
SGPr359	20	79	FL	プロテア ーゼ	メタロプロ テアーゼ	PepM10	11q22	BG187290	
SGPr704 _1	21	80	FL	プロテア ーゼ	メタロプロ テアーゼ	PepM13	3q27	BF511209, AW341249 AL119270	
SGPr303	22	81	CAT	プロテア ーゼ	メタロプロ テアーゼ	PepM2	17q11 .1	AU138954, BG251083, AW161660	
SGPr402 _1	23	82	FL	プロテア ーゼ	セリン	ズブチラーゼ	19q11	AL041695 AA454137 BG719638	

10

20

30

40

【 0 3 3 6 】

【 表 9 】

SGPr434	24	83	FL	プロテア ーゼ	セリン	トリプシン	3q21	AW137088 BF593342	
SGPr446 _1	25	84	CAT	プロテア ーゼ	セリン	トリプシン	3q21	AW243584	798ggtgggcat catcagctggg8 18
SGPr447	26	85	FL	プロテア ーゼ	セリン	トリプシン	16p13 .3	none	
SGPr432 _1	27	86	FL	プロテア ーゼ	セリン	トリプシン	未知	BE264142, BG474605, BF304202	
SGPr529	28	87	FL	プロテア ーゼ	セリン	トリプシン	19q13 .4	ss1550333_allelePos=5 1,taggggatgacacgtct;s s1546197_allelePos=51, gcgggacsactgc	
SGPr428 _1	29	88	CAT	プロテア ーゼ	セリン	トリプシン	8p23	none	473catgcacctg gaaaagctg491
SGPr425	30	89	FL	プロテア ーゼ	セリン	トリプシン	6q14	ss674620_allelePos=20 1,gagcatctgcVggagagag	1111tcaggcca ccagtgggtgga 1130

10

20

30

40

【 0 3 3 7 】

【 表 1 0 】

SGPr546	31	90	FL	プロテア ーゼ	セリン	トリプシン	19q13 .4	none	
SGPr396	32	91	FL	プロテア ーゼ	セリン	トリプシン	4q32	none	
SGPr428	33	92	FL	プロテア ーゼ	セリン	トリプシン	4q13	none	
SGPr552	34	93	CAT	プロテア ーゼ	セリン	トリプシン	4q13	none	
SGPr405	35	94	FL	プロテア ーゼ	セリン	トリプシン	16p13 .3	none	
SGPr485 _1	36	95	FL	プロテア ーゼ	セリン	トリプシン	8p23	ss1532791_allelePos=5 1.tggagakaagaacac	AA781356

【 0 3 3 8 】

【 表 1 1 】

10

20

30

40

SGPr534	37	96	FL	プロテア ーゼ	セリン	トリプシン	16q23	ss1522946_allelePos=5 1,gtctaccwccacgccc;ss 1522943_allelePos=51,c gcacctctcyaccaccac;ss 1522933_allelePos=51,c tgccagaaggayggagcctgg ;ss1522931_allelePos=5 1 全長 =101,gtctccaraaggagc; ss1522930_allelePos=5 1,gggtgactctgmgggcccc t;ss1522928_allelePos= 51,tgcatgggygactctgg	AW583018 AW582942 AW960025	172cactctgag ggggctccctcat c195
SGPr590	38	97	FL	プロテア ーゼ	セリン	トリプシン	19q11	ss82431_allelePos=99,g ccgtgarcaccactg;ss1320 361_allelePos=225,agcg gccascattggcgt	C16607	
SGPr521	39	98	FL	プロテア ーゼ	セリン	トリプシン	19q13 .4		APA542994 BE713379 W58737	646caaggctcg gtgtctctgggg66 5

10

20

30

40

【 0 3 3 9 】

【 表 1 2 】

SGPr530 _1	40	99	CAT	プロテア ーゼ	セリン	トリプシン	9q22		none	
SGPr520	41	100	FL	プロテア ーゼ	セリン	トリプシン	2q37		none	
SGPr455	42	101	FL	プロテア ーゼ	セリン	トリプシン	12p11 .2		AW450155. AW995496	
SGPr507 _2	43	102	FL	プロテア ーゼ	セリン	トリプシン	7q36		HB217724 BG219738 BG192709	
SGPr559	44	103	FL	プロテア ーゼ	セリン	トリプシン	21q22		AI978874 AI489095, BF435570	
SGPr567 _1	45	104	FL	プロテア ーゼ	セリン	トリプシン	11q23		BE732381, R78581, AW845106	
SGPr479 _1	46	105	FL	プロテア ーゼ	セリン	トリプシン	1q42		BG718703 AA401705 AA398170	780tgggaattgtg agctggggcg8 00

10

20

30

40

【 0 3 4 0 】

【 表 1 3 】

SGPr489 _1	47	106	CAT	プロテア ーゼ	セリン	トリプシン	11p15	AW271430, AW237893	
SGPr455 _1	48	107	CAT	プロテア ーゼ	セリン	トリプシン	未知	none	
SGPr524 _1	49	108	FL	プロテア ーゼ	セリン	トリプシン	未知	none	711aaaaaaaa agaaaaagaaag gaaaa734
SGPr422	50	109	FL	プロテア ーゼ	セリン	トリプシン	4q13		
SGPr538	51	110	FL	プロテア ーゼ	セリン	トリプシン	11q23	AL538140 BF934870	545tggaggcct cctggaggag56 4
SGPr527 _1	52	111	FL	プロテア ーゼ	セリン	トリプシン	未知	AW450407, AI190509, AI864473	
SGPr542	53	112	FL	プロテア ーゼ	セリン	トリプシン	19q13 .1	none	

10

20

30

40

【 0 3 4 1 】

【 表 1 4 】

SGPr551	54	113	FL	プロテア ーゼ	セリン	トリプシン	22q13	rs881144_allelePos=200 ,ctgcagccctatgcccagagg ;rs855791_allelePos=10 1,agcgaggctatgccta	AV693114 N70418 AA609066	
SGPr451	55	114	FL	プロテア ーゼ	セリン	トリプシン	12q23	ss1881349_allelePos=2 01,gggcgcctgcaragg;ss1 266911_allelePos=101,c cactgcactaaagacrcatg	BG722131 BG722203	
SGPr452 _1	56	115	FL	プロテア ーゼ	セリン	トリプシン	16p13 .3		none	
SGPr504	57	116	部分	プロテア ーゼ	セリン	トリプシン	未知		none	
SGPr469	58	117	部分	プロテア ーゼ	セリン	トリプシン	未知			
SGPr400	59	118	部分	プロテア ーゼ	セリン	トリプシン	4q32		none	

10

20

30

40

50

【0342】

表3は、プロテアーゼ触媒ドメインおよび他の蛋白質ドメインの程度および境界を示す。カラムの標題は以下のとおりである：“遺伝子名”，“ID#na”，“ID#aa”，“FL/Cat”，“プロファイル開始”，“プロファイル終了”，“ドメイン開始”，“ドメイン終了”および“プロファイル”。最初の7つのカラム（すなわち，“遺伝子名”，“ID#na”，“ID#aa”，“FL/Cat”，“スーパーファミリー”，“

グループ”， “ファミリー”）の内容は表1について上述したとおりである。 “プロファイル開始”， “プロファイル終了”， “ドメイン開始” および “ドメイン終了” は， 隠れマルコフモデルを用いて得た触媒的範囲の境界を規定するデータを表す。 全体の蛋白質中の触媒ドメインの境界は， “ドメイン開始” および “ドメイン終了” のカラムで示される。 “プロファイル” は， 蛋白質配列中の触媒ドメインおよび他のタイプのドメインを同定するために用いた隠れマルコフモデルの種類を示す。 完全（ “グローバル” ） またはスミス・ウォーターマン（ “ローカル” ） モデルのいずれを用いてHMMR検索を行ったかは以下に説明される。 触媒ドメインの多重配列アラインメントから出発して， 2つの隠れマルコフモデルを構築した。 その一方は， 触媒ドメインに対する部分的マッチを可能とする； これは “ローカル” HMMであり， 配列マッチングにおけるスミス・ウォーターマンアラインメントと類似する。 他方のモデルは， 完全触媒ドメインに対するマッチのみを可能とする； これは “グローバル” HMMであり， 配列マッチングにおけるNeedleman-Wunschアラインメントと類似する。 スミス・ウォーターマンのローカルモデルはより特異的であり， 触媒ドメインに対するフラグメント的なマッチを可能とする。 一方， グローバルな “完全” モデルはより感度が高く， 遠隔のホモログ同定を可能とする。 これらのドメインは， PFAM（<http://pfam.wustl.edu/hmmsearch.shtml>） モデルを用いて同定した。 これは， 多くの共通蛋白質ドメインをカバーする多重配列アラインメントおよび隠れマルコフモデルの大きな集積物である。 PFAMのバージョン5.5（2000年9月）は， 2478種類の蛋白質ファミリーのアラインメントおよびモデルを含む（<http://pfam.wustl.edu/faq.shtml>）。 PFAMアラインメントは<http://pfam.wustl.edu/hmmsearch.shtml>からダウンロードし， HMM検索はTimeLogicコンピュータ（Time Logic Corporation, Inclined Village, NV）で現場で行った。 多くの蛋白質はHMM検索により認識される2以上のドメインを有する。 これらの蛋白質については， ドメインは別の行に示されている。

10

20

【0343】

【表15】

遺伝子名	ID# na	ID#aa	FL/Cat	プロフ ファイル _開始	プロフ ファイル _終了	ドメイン _開始	ドメイン _終了	プロファイル
SGPr397	1	60	FL	1	46	739	280	Zn カルボキシペプチダーゼ(PF00246)
SGPr397	1	60	FL	1	82	41	120	Zn カルボキシペプチダーゼ 活性化ペプチド
SGPr413	2	61	FL	1	246	50	291	Zn カルボキシペプチダーゼ (PF00246)
SGPr404	3	62	FL	1	248	91	466	Zn カルボキシペプチダーゼ (PF00246)
SGPr536_1	4	63	FL	1	337	203	456	パパイン (PF00112)
SGPr414	5	64	FL	1	72	1951	2045	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr414	5	64	FL	1	32	1707	7732	UCH2b (PF00442)
SGPr430	6	65	FL	1	72	886	951	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr430	6	65	FL	1	32	342	373	UCH2b (PF00442)
SGPr496_1	7	66	FL	1	72	875	935	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr496-1	7	66	FL	1	32	593	6241	UCH2b (PF00442)

10

20

30

40

【 0 3 4 4 】

【 表 1 6 】

SGPr496_1	7	66	PL	1	82	465	534	Zn-フィンガー, ユビキチンヒドロラーゼ中 (PF02148)
SGPr495	8	67	FL	1	72	895	781	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr495	8	67	FL	1	32	190	221	UCH2b (PF00442)
SGPr495	8	67	FL	7	82	78	148	Zn-フィンガー, ユビキチンヒドロラーゼ中 (PF02148)
SGPr407	9	68	FL	80	90	481	491	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr453	10	69	FL	1	72	615	677	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr453	10	69	FL	1	32	273	304	UCH2b (PF00442)
SGPr453	10	69	FL	1	82	29	99	Zn-フィンガー, ユビキチンヒドロラーゼ中 (PF02148)
SGPr445	11	70	FL	1	32	190	221	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr445	11	70	FL	7	82	78	148	Zn-フィンガー, ユビキチンヒドロラーゼ中 (PF02148)

10

20

30

40

【 0 3 4 5 】

【 表 1 7 】

SGPr401_1	12	71	FL	1	72	292	364	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr401_1	12	71	FL	1	32	35	66	UCH2b (PF00442)
SGPr408	13	72	FL	1	72	395	475	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr408	13	72	FL	1	32	100	1311	UCH2b (PF00442)
SGPr480	14	73	FL	1	72	1506	1566	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr480	14	73	FL	i	32	734	765	UCH2b (PF00442)
SGPr480	14	73	FL	i	29	26r	296	EF ハンド (PF00036)
SGPr480	14	73	FL	1	29	232	260	EF ハンド (PF00036)
SGPr431	15	74	FL	1	72	636	948	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr431	15	74	FL	1	32	445	476	UCH2b (PF00442)
SGPr429	16	75	FL	1	72	332	419	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)
SGPr429	16	75	PL	1	32	89	120	UCH2b (PF00442)
SGPr503	17	76	FL	1	72	432	501	ユビキチンカルボキシ末端ヒドロラーゼファミリー -2b (PF00443)

10

20

30

40

【 0 3 4 6 】

【 表 1 8 】

SGPr503	17	76	FL	1	32	69	99	UCH2b (PF00442)
SGPr427	16	77	FL	i	72	648	709	ユビキチンカルボキシ末端ヒドロラーゼファミリ ー2b (PF00443)
SGPr427	18	77	FL	1	29	101	129	UCH2b (PF00442)
SGPr092	19	78	FL	49	168	75	194	ペプチダーゼ M10 (PF00413)
SGPr092	19	78	FL	168	179	207	218	ADAM (PF00413)
SGPr359	20	79	FL	1	168	44	212	ペプチダーゼ M10 (PF00413)
SGPr359	20	79	FL	1	50	302	443	3x ヘモペキシン (PF00045)
SGPr104_1	21	80	FL	1	222	561	764	ペプチダーゼ M13 (PF01431)
SGPr303	22	81	CAT	1	416	10	397	ペプチダーゼ M1 (PF01433)
SGPr402_1	23	82	FL	1	360	118	437	ズブチラーゼ (PF00082)
SGPr434	24	83	FL	129	138	39	46	p20.1CE (PF00656)
SGPr446_1	25	84	CAT	1	242	13	227	トリプシン (PF00089)
SGPr447	26	85	FL	1	259	33	270	トリプシン (PF00089)
SGPr432_1	27	86	FL	6	259	117	343	トリプシン (PF00089)
SGPr529	28	87	FL	413	416	184	187	トリプシン (PF00089)
SGPr428_1	29	88	CAT	7	259	24	246	トリプシン (PF00089)
SGPr425	30	89	FL	387	408	287	305	トリプシン (PF00089)
SGPr548	31	90	FL	1	259	88	313	トリプシン (PF00089)

10

20

30

40

【 0 3 4 7 】

【 表 1 9 】

SGPr396	32	91	FL	1	259	28	282	トリプシン (PF00089)
SGPr426	33	92	FL	1	259	194	419	トリプシン (PF00089)
SGPr552	34	93	CAT	1	255	2	222	トリプシン (PF00089)
SGPr405	35	94	FL	60	259	218	406	トリプシン (PF00089)
SGPr405	35	94	FL	128	209	419	496	トリプシン (PF00089)
SGPr405	35	94	FL	122	251	836	761	トリプシン (PF00089)
SGPr485_1	36	95	FL	1	259	68	295	トリプシン (PF00089)
SGPr534	37	96	FL	1	259	34	258	トリプシン (PF00089)
SGPr390	38	97	FL	1	259	896	1122	トリプシン (PF00089)
SGPr390	38	97	FL	1	259	284	500	トリプシン (PF00089)
SGPr390	38	97	FL	1	259	573	800	トリプシン (PF00089)
SGPr521	39	98	FL	1	259	30	245	トリプシン (PF00089)
SGPr530_1	40	99	CAT	1	259	14	255	トリプシン (PF00089)
SGPr520	41	100	FL	1	259	73	306	トリプシン (PF00089)
SGPr455	42	101	FL	1	259	433	674	トリプシン (PF00089)
SGPr455	42	101	FL	109	259	4	156	トリプシン (PF00089)
SGPr455	42	101	FL	2	116	175	8121	3 x CUB ドメイン (PF00431)
SGPr507_2	43	102	FL	35	148	42	135	トリプシン (PF00089)
SGPr507_2	43	102	FL	247	259	246	258	トリプシン (PF00089)
SGPr559	44	103	FL	1	259	217	444	トリプシン (PF00089)

10

20

30

40

【 0 3 4 8 】

【 表 2 0 】

SGP1559	44	103	FL	1	43	71	109	低密度リポ蛋白質セプタードメインクラス A (PF00057)
SGP1567_1	45	104	FL	1	259	296	524	トリプシン (PF00089)
SGP1479_1	46	105	FL	1	259	60	288	トリプシン (PF00089)
SGP1489_1	47	106	CAT	1	227	56	257	トリプシン (PF00089)
SGP1489_1	47	106	CAT	1	116	304	633	2 x CUB ドメイン (PF00431)
SGP1465_1	48	107	CAT	12	259	2	240	トリプシン (PF00089)
SGP1524_1	49	108	FL	1	259	613	842	トリプシン (PF00089)
SGP1524_1	49	108	FL	1	43	489	6031	3 x 低密度リポ蛋白質セプタードメインクラス A (PF00057)
SGP1422	50	109	FL	1	259	216	441	トリプシン (PF00089)
SGP1538	51	110	FL	9	259	298	448	トリプシン (PF00089)
SGP1527_1	52	111	FL	1	259	47	266	トリプシン (PF00089)
SGP1527_1	52	111	FL	1	156	323	454	トリプシン (PF00089)
SGP1527_1	52	111	FL	12	149	564	679	トリプシン (PF00089)
SGP1542	53	112	FL	1	259	35	259	トリプシン (PF00089)
SGP1551	54	113	FL	1	259	568	797	トリプシン (PF00089)
SGP1551	54	113	FL	1	43	447	559	3 x 低密度リポ蛋白質セプタードメインクラス A (PF00057)
SGP1451	55	114	FL	1	259	89	324	トリプシン (PF00089)

10

20

30

40

【 0 3 4 9 】

【 表 2 1 】

SGPr452_1	56	115	FL	1	259	73	280	トリプシン (PF000089)
SGPr504	57	116	部分	7	52	1	45	トリプシン (PF000089)
SGPr469	58	117	部分	210	259	1	48	トリプシン (PF000089)
SGPr400	59	1118	部分	1	198	133	281	トリプシン (PF000089)

10

20

30

40

50

【0350】

表4は、非重複蛋白質配列のNCBIデータベース (<http://www.ncbi.nlm.nih.gov/Entrez/protein.html>) に対するアミノ酸配列のスミス・ウォーターマン類似性検索 (マトリックス: Paml00; ギャップオープン/イクステンションペナルティー12/2) の結果を示す。カラム標題は以下のとおりである: "遺伝子名", "ID#na", "ID#aa", "FL/Cat", "スー

パーファミリー”， “グループ”， “ファミリー”， “Pスコア”， “aa長さ”， “aa IDマッチ”， “%同一性”， “%類似性”， “ACC#nraaマッチ”， および “説明”。最初の7つのカラム（すなわち，“遺伝子名”， “ID#na”， “ID#aa”， “FL/Cat”， “スーパーファミリー”， “グループ”， および “ファミリー”）の内容は表1について上述したとおりである。“Pスコア”はスミス・ウォーターマン確率スコアを表す。この数値はアラインメントが偶然により生ずるおよその確率を表す。すなわち，非常に低い数値，例えば 2.10×10^{-64} は，クエリーとデータベース標的との間に非常に顕著なマッチがあることを示す。“aa長さ”は，蛋白質の長さをアミノ酸で表す。“aa IDマッチ”は，アラインメント中で同一であるアミノ酸の数を示す。“%同一性”は，アラインメントした領域にわたって同一であるヌクレオチドのパーセントを示す。“%類似性”は，アラインメント全体にわたって類似するアミノ酸のパーセントを示す。“ACC#nraaマッチ”は，非重複蛋白質のNCBIデータベース中で最も類似する蛋白質の受託番号を示す。“説明”は，非重複蛋白質NCBIデータベース中で最も類似する蛋白質の名前を含む。

10

【0351】

【表22】

遺伝子 名	ID #n a	ID#a	FL/C at	スーパー ファミリー	グループ	ファミリー	Pスコア	aa 長さ	aal Dマ ッチ	%同 一性	%類 似性	ACCIf nrea マッチ	説明
SGPr3 97	1	60	FL	プロテ アーゼ	カルボキ シペプチ ダーゼ	Znカルボ キシペプ チダーゼ	3.10E-22 0	315	315	100	100	NP_065 094.1	カルボキシペプチダー ゼB前駆体[Homo sapiens]
SGPr4 13	2	61	FL	プロテ アーゼ	カルボキ シペプチ ダーゼ	Znカルボ キシペプ チダーゼ	5.90E-93	374	148	49	68	AAF073 44.1	(AF190274)カルボキ シペプチダーゼホモロ グ[Bothrops jararaca]
SGPr4 04	3	62	FL	プロテ アーゼ	カルボキ シペプチ ダーゼ	Znカルボ キシペプ チダーゼ	0	528	502	94	98	NP_061 355.7	カルボキシペプチダー ゼ X2[Mus musculus]
SGPr5 36_1	4	63	FL	プロテ アーゼ	システイン	パパイン	1.10E-27 8	467	467	100	100	NP_071 447.1	P3ECSL[Homo sapiens]
SGPr4 14	5	64	FL	プロテ アーゼ	システイン	UCH2b	0	335 3	125 9	99	100	NP_055 524.1	KIAA0570 遺伝子産物 [Homo sapiens]

10

20

30

40

SGPr4 30	6	65	FL	プロテ アーゼ	システイン	UCH2b	0	980	930	99	99	BAB134 20.1	(AB046814)KIAA159 4 蛋白質[Homo sapiens]
SGPr4 961	7	66	FL	プロテ アーゼ	システイン	UCH2b	2.00E-19 0	953	498	95	98	AAF669 53.1	(AF229643)ユビキチ ン特異的プロテアーゼ [Mus musculus]
SGPr4 85	8	67	FL	プロテ アーゼ	システイン	UCH2b	2.40E-17 6	783	282	700	100	AAH059 91.1	(BC005991)未知 (MGC 用蛋白 質:14793)[Homo sapiens]
SGPr4 07	9	68	FL	プロテ アーゼ	システイン	UCH2b	2.60E-40	753	80	78	84	NP_036 807.1	ユビキチン特異的プロ テアーゼ 23;NEDD8- 特異的プロテアーゼ [Homo sapiens]
SGPr4 53	10	69	FL	プロテ アーゼ	システイン	UCH2b	0	712	712	100	100	NP_115 523.7	仮定蛋白質 DKFZp434D0127[Ho mo sapiens]

10

20

30

40

【 0 3 5 3 】

【 表 2 4 】

SGPr4 45	11	70	FL	プロテ アーゼ	システイン	UCH2b	3.60E-18 5	289	289	100	AAH058 97.1	(BC005991)未知 (MGC 用蛋白 質:14793)[Homo sapiens]
SGPr4 01_1	12	71	FL	プロテ アーゼ	システイン	UCH2b	7.30E-25 4	366	366	100	NP_073 743.7	仮定蛋白質 FLJ12552[Homo sapiens]
SGPr4 08	13	72	FL	プロテ アーゼ	システイン	UCH2b	0	128 7	128 7	100	BAB550 63.1	(AK027362)命名され ていない蛋白質産物 [Homo sapiens]
SGPr4 60	14	73	FL	プロテ アーゼ	システイン	UCH2b	0	160 4	127 2	99	NP_115 971.1	ユビキチン特異的プロ テアーゼ[Homo sapiens]
SGPr4 31	15	74	FL	プロテ アーゼ	システイン	UCH2b	2.40E-25 1	109 2	397	100	NP_115 946.7	HP43.8KD 蛋白質 [Homo sapiens]
SGPr4 29	16	75	FL	プロテ アーゼ	システイン	UCH2b	1.50E-25 0	103 3	368	100	NP_115 812.1	仮定蛋白質 FLJ23277[Homo sapiens]

【 0 3 5 4 】

【 表 2 5 】

10

20

30

40

SGPr5 03	17	76	FL	プロテ アーゼ	システイン	UCH2b	0	517	508	100	700	AAH048 68.1	(BC004868)未知 (MGC 用蛋白 質:10702)[Homo sapiens]
SGPr4 27	18	77	FL	プロテ アーゼ	システイン	UCH2b	1.80E-92	112 3	269	36	51	AAF472 60.1	(AE003465)CG3872 遺伝子産物 [Drosophila melanoga ster]
SGPr0 92	19	78	FL	プロテ アーゼ	メタロプロ テアーゼ	PepM10	4.70E-17 1	261	261	100	100	XP_011 971.1	マトリクスメタロプロテ イナーゼ 26[Homo sapiens]
SGPr3 59	20	79	FL	プロテ アーゼ	メタロプロ テアーゼ	PepM10	0	483	483	100	100	NP_004 767	マトリクスメタロプロテ イナーゼ20プレプロ蛋 白質:エナメルシン [Homo sapiens]
SGPr1 04_1	21	80	FL	プロテ アーゼ	メタロプロ テアーゼ	PepM13	0	765	765	100	100	NP_055 508.1	KIAA0604 遺伝子産物 [Homo sapiens]

10

20

30

40

【 0 3 5 5 】

【 表 2 6 】

SGPr3 03	22	81	CAT	プロテ アーゼ	メタロプロ テアーゼ	PepM2	2.20E-28 4	419	407	97	98	CAA107 09.1	(AJ132583)ピューロマ イン感受性アミノペプ チダーゼ[Homo sapiens]
SGPr4 02	23	82	FL	プロテ アーゼ	セリン	ズブチラー ゼ	0	755	513	82	89	P29121	神経内分泌コンベルタ ーゼ3前駆体[Mus musculus]
SGPr4 34	24	83	FL	プロテ アーゼ	セリン	トリプシン	6.20E-43	397	104	42	59	NP_036 164.7	胃膜トリプターゼ[Mus musculus]
SGPr4 46_1	25	84	CAT	プロテ アーゼ	セリン	トリプシン	2.50E-40	227	107	45	57	NP_038 949.1	遠位腸セリンプロテア ーゼ[Mus musculus]
SGPr4 47	26	85	FL	プロテ アーゼ	セリン	トリプシン	1.00E-97	296	167	60	77	BAB302 77.1	(AK016509)予測[Mus musculus]
SGPr4 32_1	27	86	FL	プロテ アーゼ	セリン	トリプシン	3.70E-56	628	95	100	100	NP_076 869.1	仮定蛋白質 IMAGE3455200[Hom o sapiens]
SGPr5 29	28	87	FL	プロテ アーゼ	セリン	トリプシン	1.70E-18 4	276	276	100	100	NP_002 767.7	カリクレイン10;プロテ アーゼ, セリン 様, 1[Homo sapiens]

10

20

30

40

【 0 3 5 6 】

【 表 2 7 】

SGPr4 28_1	29	88	CAT	プロテ アーゼ	セリン	トリプシン	1.90E-58	285	92	53	73	BAA242 15.1	(AK005740)予測[Mus musculus]
SGPr4 25_1	30	89	FL	プロテ アーゼ	セリン	トリプシン	5.80E-26 8	413	472	99	99	CAC35 071.1	(AL121939)dJ223E3. 1(予測分泌蛋白質 ZSIG13)[Homo sapiens]
SGPr5 48	31	90	FL	プロテ アーゼ	セリン	トリプシン	2.60E-16 8	320	256	100	100	AAG09 469.1	(AF242195)KLK15[H omo sapiens]
SGPr3 96	32	91	FL	プロテ アーゼ	セリン	トリプシン	1.60E-56	328	111	44	61	BAA849 41.1	(AB018694)表皮特異 的セリンプロテアーゼ [Xenopus laevis]
SGPr4 26	33	92	FL	プロテ アーゼ	セリン	トリプシン	7.70E-93	425	181	43	61	NP_054 777.7	DESC1 蛋白質[Homo sapiens]
SGPr5 52	34	93	CAT	プロテ アーゼ	セリン	トリプシン	1.20E-45	222	96	42	59	NP_054 777.1	DESC1 蛋白質[Homo sapiens]
SGPr4 05	35	94	FL	プロテ アーゼ	セリン	トリプシン	1.10E-30	848	111	54	65	P19236	肥満細胞腫プロテアー ゼ前駆体 [Canis familiaris]

10

20

30

40

【 0 3 5 7 】

【 表 2 8 】

SGPr4 85_1	36	95	FL	プロテ アーゼ	セリン	トリプシン	7.20E-13 3	352	223	94	96	BAB035 69.1	(AB046651)仮定蛋白 質[Macaca fascicularis]
SGPr5 34	37	96	FL	プロテ アーゼ	セリン	トリプシン	3.60E-16 5	263	253	96	98	NP_001 897.1	キモトリプシンノーゲン B1[Homo sapiens]
SGPr3 90	38	97	FL	プロテ アーゼ	セリン	トリプシン	2.60E-53	112 8	135	46	59	BAB231 684.1	(AK004939)予測[Mus musculus]
SGPr5 21	39	98	FL	プロテ アーゼ	セリン	トリプシン	2.30E-15 5	253	253	100	100	NP_005 037.1	カリクレイン7(キモトリ プシン, 角質層);プロ テアーゼ, セリン, 6(キ モトリプシン, 角質 層)[Homo sapiens]
SGPr5 30_1	40	99	CAT	プロテ アーゼ	セリン	トリプシン	1.10E-95	272	142	100	100	CAC12 709.1	(AL136097)hA62C3.1 (精巢セリンプロテアー ゼと類似)[Homo sapiens]
SGPr5 20	41	100	FL	プロテ アーゼ	セリン	トリプシン	1.50E-83	578	158	73	83	BAB245 87.1	(AK006434)予測[Mus musculus]

10

20

30

40

【 0 3 5 8 】

【 表 2 9 】

SGPr9 55	42	101	FL	プロテ アーゼ	セリン	トリプシン	5.90E-17 9	970	386	41	58	T30337	ポリ蛋白質-アフリカ カギツメカエル
SGPr5 97_2	43	102	FL	プロテ アーゼ	セリン	トリプシン	2.40E-12 1	265	195	73	81	NP_080 593.1	リケン cDNA1700016G05 遺 伝子[Mus musculus]
SGPr5 59	44	103	FL	プロテ アーゼ	セリン	トリプシン	1.40E-28 8	454	454	100	100	NP_076 927.1	貴膜プロテアーゼ, セ リン 3[Homo sapiens]
SGPr5 67_1	45	104	FL	プロテ アーゼ	セリン	トリプシン	1.70E-13 5	537	534	99	99	NP_114 435.1	モザイクセリンプロテア ーゼ[Homo sapiens]
SGPr4 79_1	46	105	FL	プロテ アーゼ	セリン	トリプシン	1.70E-39	326	107	42	57	NP_114 154.1	マラブシン[Homo sapiens]
SGPr4 89_1	47	106	CAT	プロテ アーゼ	セリン	トリプシン	2.70E- 90	556	194	37	54	T30338	オビダクチン[Xenopus laevis]
SGPr4 65_1	48	107	CAT	プロテ アーゼ	セリン	トリプシン	2.70E-76	296	144	48	66	NP_033 381.1	精巢セリンプロテアー ゼ 1[Mus musculus]
SGPr5 24_1	49	108	FL	プロテ アーゼ	セリン	トリプシン	1.30E-79	850	193	41	55	BAB236 84.1	(AK004939)予測[Mus musculus]

10

20

30

40

【 0 3 5 9 】

【 表 3 0 】

SGPr4 22	50	109	FL	プロテ アーゼ	セリン	トリプシン	4.90E-80	447	173	39	59	NP_054 77.1	DESC1 蛋白質[Homo sapiens]
SGPr5 38	51	110	FL	プロテ アーゼ	セリン	トリプシン	9.1e-315	457	457	100	100	NP_110 397.1	スピネシン[Homo sapiens]
SGPr3 27_1	52	111	FL	プロテ アーゼ	セリン	トリプシン	1.30E-52	818	114	42	59	AAH038 51.1	(BC003851)プロテア ーゼ, セリン, 8 (プロス タシン)に類似[Mus musculus]
SGPr5 42	53	112	FL	プロテ アーゼ	セリン	トリプシン	2.70E-41	284	110	43	58	NP_005 308.1	グランザイムM前駆 体;リンパ球 met-ase1[Homo sapiens]
SGPr5 51	54	113	FL	プロテ アーゼ	セリン	トリプシン	0	802	675	84	90	BAB236 84.1	(AK004939)予測[Mus musculus]
SGPr4 51	55	114	FL	プロテ アーゼ	セリン	トリプシン	9.90E-41	359	101	39	59	NP_072 152.1	副腎分泌性セリンプロ テアーゼ前駆体 [Rattus norvegicus]

10

20

30

40

【 0 3 6 0 】

【 表 3 1 】

SGPr4 52_1	56	115	FL	プロテ アーゼ	セリン	トリブシン	1.40E-81	288	142	57	72	AAK152 64.1	(AF305425)移植セリ ンプロテイナーゼ 2[Mus musculus]
SGPr5 04	57	116	部分	プロテ アーゼ	セリン	トリブシン	2.40E-13	45	26	61	88	NP_002 095.1	グランザイムK前駆体; グランザイム3; グラン ザイムK(セリンプロテ アーゼ, グランザイム 3); トリブターゼ II[Homo sapiens]
SGPr4 69	58	117	部分	プロテ アーゼ	セリン	トリブシン	2.20E-17	46	32	69	84	BAB302 77.1	(AK016509)予測[Mus musculus]
SGPr4 00	59	118	部分	プロテ アーゼ	セリン	トリブシン	2.30E-16	309	72	38	48	NP_036 164.1	貫膜トリブターゼ[Mus musculus]

10

20

30

40

【0361】

実施例

以下の実施例は限定ではなく、本発明の種々の観点および特徴の代表的なものにすぎない。以下の実施例は、本発明のプロテアーゼの単離および特徴付けを示す。

【0362】

実施例1：蛋白質プロテアーゼをコードするゲノムフラグメントの同定

新規プロテアーゼは、Celera ヒトゲノム配列データベースから、および公共のHu 50

man Genome Sequencing project (<http://www.ncbi.nlm.nih.gov/>) から、隠れマルコフモデル (HMMR) を用いて同定した。ゲノムデータベースエントリーは、6つのオープンリーディングフレームに翻訳し、HMMR 2.1のFieldプログラム可能アレイ (FPGA) 加速版を有するTimelogic Decypherボックスを用いて、モデルに対してこれらを検索した。HMMRプロファイルに対してアラインメントした、予測された蛋白質配列をコードするDNA配列を、元のゲノムデータベースから抽出した。次に、核酸配列をPangea Clusteringツールを用いてクラスター化し、反復エントリーを除外した。次に推定プロテアーゼ配列を一連のクエリーおよびフィルターを通して順番に走らせて、新規プロテアーゼ配列を同定した。特に、HMMRで同定された配列は、BLASTN およびBLASTXを用いて、既知のヒトプロテアーゼおよびすべてのその後の新規プロテアーゼ配列をそれが同定されるにつれて含有するヌクレオチドおよびアミノ酸レポジトリに対して検索した。出力はスプレッドシートで表示して、手動検査で既知の遺伝子を除外することを容易にした。2つのモデル、すなわち、“完全”モデルおよび“部分”またはスミス・ウォーターマンモデルを開発した。部分モデルを用いて準触媒的ドメインを同定し、完全モデルを用いて完全触媒ドメインを同定した。次に選択したヒットをBLASTNを用いて公共NRNAおよびESTデータベースに対してクエリーを行い、これらが実際にユニークであることを確認した。

10

【0363】

部分DNA配列の延長による、より長い配列、例えば全長オープンリーディングフレームの包含は、いくつかの方法により行った。表5に示されるcDNAデータベースを繰り返しblastn検索して、ゲノム配列を延長したcDNAを見いだした。“LifeGold”データベースはIncyte Genomics, Inc (<http://www.incyte.com/>) から入手した。NCBIデータベースはNational Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) から入手した。すべてのblastn検索は、ヌクレオチドミスマッチについてのペナルティー3、およびヌクレオチドマッチのリワード1を用いて行った。ギャップ付きblastアルゴリズムは以下に記載されている (Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”, Nucleic Acids Res. 25: 3389-3402)。

20

30

【0364】

部分DNA配列の延長による全長オープンリーディングフレームの包含はまた、ゲノムデータベースの検索の繰り返しにより行った。第1の方法は、スミス・ウォーターマンアルゴリズムを用いて、最も近いホモログまたはオルログを部分配列に対して蛋白質-蛋白質検索を行った。標的データベースは、Genescan [Chris Burge and Sam Karlin, “Prediction of Complete Gene Structures in Human Genomic DNA”, JMB (1997), 268 (1): 78-94] およびヒトゲノムプロジェクト (HGP) ならびにCelerから得られた全ヒトゲノム配列のオープンリーディングフレーム (ORF) 予測からなる。検索したゲノムデータベースの完全なセットは以下の表6に示される。潜在的な延長をコードするゲノム配列は、NCBI非重複データベースに対するblastp分析を行ってヒットの新規性を確認することによりさらに評価した。延長するゲノム配列は、DNASTarのSeqmanプログラムを用いて潜在的イントロンを削除した後cDNA配列中に組み込んだ。スミス・ウォーターマン検索について用いたデフォルトのパラメータは以下のとおりである。マトリクス: PAM100; ギャップオープ

40

50

ニングペナルティ：12；ギャップ延長ペナルティ：2。Genescan予測は，Genescanプログラムを，Chris Burge and Sam Karlin（"Prediction of Complete Gene Structure in Human Genomic DNA"، JMB（1997）268（1）：78－94）に詳細が記載されるように用いて行った。ゲノムDNAからのORFの予測は，標準的な6フレーム翻訳を用いて行った。

【0365】

ゲノム配列からDNA延長を規定する別の方法は，Genescanプログラムを用いてゲノムデータベースを繰り返し検索してエクソンスプライシングを予測した[Burge and Karlin, JMB（1997），268（1）：78－94]。次にこれらの予測された遺伝子を，関連するプロテアーゼに対するホモロジーに基づいてこれらが部分遺伝子の"本物の"延長であるかを見ることにより評価した。

10

【0366】

別の方法は，Genewiseプログラム（<http://www.sanger.ac.uk/Software/Wise2/>）を使用して，最も近いオルトログ／ホモログに対するホモロジーに基づいて潜在的ORFを予測することを含む。Genewiseは，2つの入力，すなわち相同の蛋白質と目的とする遺伝子を含むゲノムDNAを必要とする。ゲノムDNAはCeleraおよびヒトゲノムプロジェクトのデータベースのblastn検索により同定した。オルトログは，NCBI非重複蛋白質データベース（NR）のblastp検索により同定した。Genewiseは，蛋白質配列をゲノムDNA配列と比較し，イントロンおよびフレームシフトエラーを得ることができる。

20

【0367】

表5－cDNAに基づく配列延長に用いたデータベース

【表32】

データベース	データベース日付
LifeGold templates	2001年5月
LifeGold compseqs	2001年5月
LifeGold compseqs	2001年5月
LifeGold compseqs	2001年5月
LifeGold fl	2001年5月
LifeGold flft	2001年5月
NCBI ヒト Ests	2001年5月
NCBI ネズミ Ests	2001年5月
NCBI 非重複	2001年5月

30

【0368】

表6－ゲノムに基づく配列延長に用いたデータベース

【表33】

40

データベース	登録物の数	データベース日付
Celera v. 1-5	5,306,158	2000年1月
Celera v. 6-10	4,209,980	2000年5月
Celera v. 11-14	7,22,425	2000年4月
Celera v. 15	243,044	2000年4月
Celera v. 16-17	25,885	2000年4月
Celera Assembly 5 (release 25h)	479,986	2001年5月
HGP Phase 0	3,189	2000年11月1日
HGP Phase 1	20,447	2001年1月1日
HGP Phase 2	1,619	2001年1月1日
HGP Phase 3	9,224	2001年5月
HGP Chromosomal assemblies	2759	2001年5月

10

【0369】

結果

特許仮出願において遺伝子を同定するために用いた配列情報源を以下に示す。Genewiseを用いて延長した遺伝子については、蛋白質オルトログおよびゲノムDNAの受託番号が記載されている（Genewiseはゲノム配列から標的遺伝子のコーディング配列を組み立てるためにオルトログを用いる）。オルトログのアミノ酸配列は、蛋白質のNCBI非重複データベースから得た（<http://www.ncbi.nlm.nih.gov/Entrez/protein.html>）。ゲノムDNAは2つの起源，すなわち，CeleraおよびNCBI-NRNAからのものであり，これは以下に示される。cDNA起源もまた以下に示される。Genscan予測の入力としてすべてのゲノム配列を用いてスプライシング部位を予測した[Burge and Karlin, JMB (1997) 268(1):78-94]。略号：HGP：ヒトゲノムプロジェクト；NCBI, National Center for Biotechnology Information

30

【0370】

SGPr397, 配列番号1, 配列番号60

Genewiseオルトログ：BAB25826.1, XP_005284.2, NP_065094.1

ゲノムDNA起源：Celera_asm5h181000001172043cDNA

起源：公共(gi|9966830|ref|NM_020361.1)

【0371】

SGPr413, 配列番号2, 配列番号61

Genewiseオルトログ：gi|6013463, XP_003009.1, P15086

ゲノムDNA起源：Celera_asm5h300475633

40

【0372】

SGPr404, 配列番号3, 配列番号62

Genewiseオルトログ：BAB31768.1, NP_061355.1, AAH03713

ゲノムDNA起源：Celera_asm5h90000641768196

【0373】

SGPr536_1, 配列番号4, 配列番号63

Genewiseオルトログ：BAB18637

ゲノムDNA起源：90000642234172

50

【0374】

SGPr414, 配列番号5, 配列番号64

Genewise オルトログ: AAF50752.1

ゲノムDNA 起源: 90000628114448

cDNA 起源: AK023845.1 | AK023845 Homo sapiens

cDNA FLJ13783fis; Incyte399773.5

【0375】

SGPr430, 配列番号6, 配列番号65

Genewise オルトログ: NP_065954

ゲノムDNA 起源: 301015601

cDNA 起源: AB046814 Homo sapiens mRNA, KIAA1594

10

【0376】

SGPr496_1, 配列番号7, 配列番号66

Genewise オルトログ: AAH07196, AAF66953

ゲノムDNA 起源: 90000627702299

【0377】

SGPr495, 配列番号8, 配列番号67

Genewise オルトログ: NP_006438

ゲノムDNA 起源: 90000627041101

20

【0378】

SGPr407, 配列番号9, 配列番号68

Genewise オルトログ: BAB27431, AAH03130, NP_057656

ゲノムDNA 起源: 92000003986525

【0379】

SGPr453, 配列番号10, 配列番号69

Genewise オルトログ: NP_006528

ゲノムDNA 起源: 90000640175777

cDNA 起源: AL136825.1 | HSM801793 Homo sapiens mRNA; Incyte428428.1

30

【0380】

SGPr445, 配列番号11, 配列番号70

Genewise オルトログ: NP_006438

ゲノムDNA 起源: 90000627041101

cDNA 起源: 9863487 328bp ユビキチンカルボキシル末端ヒドロラーゼ; Incyte4802789CA2

【0381】

SGPr401_1, 配列番号12, 配列番号71

Genewise オルトログ: BAB14881, NP_073743, BAB24720

40

ゲノムDNA 起源: 92000004473288

cDNA 起源: NM022832.1 | Homo sapiens 仮定蛋白質 FLJ12552 (FLJ12552)

【0382】

SGPr408, 配列番号13, 配列番号72

Genewise オルトログ: Q24574, AAF50752

ゲノムDNA 起源: 90000628565543

cDNA 起源: AK027362

【0383】

50

SGPr480, 配列番号14, 配列番号73

Genewiseオルトログ:AAF49100, T29010

ゲノムDNA起源:90000640697688

cDNA起源:EF-hand;CAAX:NP_115971

【0384】

SGPr431, 配列番号15, 配列番号74

Genewiseオルトログ:AAK26248, BAA92610, Q92353

ゲノムDNA起源:90000642340202

【0385】

SGPr429, 配列番号16, 配列番号75

Genewiseオルトログ:BAB15591, AAG42764, gi_11358453

ゲノムDNA起源:90000642540891

cDNA起源:AK026930.1|AK026930 Homo sapiens

cDNA:FLJ23277

【0386】

SGPr503, 配列番号17, 配列番号76

Genewiseオルトログ:AAF40451, AAF46096, AAH04868

ゲノムDNA起源:90000642658172

cDNA起源:BC004868.1|BC004868 Homo sapiens, 20

クローンMGC:10702;IncYTE5432879CB1

【0387】

SGPr427, 配列番号18, 配列番号77

Genewiseオルトログ:XP_003288, AAC27356, BAA86517

ゲノムDNA起源:181000001646773

cDNA起源:IncYTE7485896CB1

【0388】

SGPr092, 配列番号19, 配列番号78

Genewiseオルトログ:XP_011971, 1, NP_068573.1, AAF80180.1 30

ゲノムDNA起源:clera_asm5h300261795

cDNA起源:gi|12736016

【0389】

SGPr359, 配列番号20, 配列番号79

Genewiseオルトログ:1.) gi|11545845|ref 2.) gi|1200634|gb 3.) gi|3511149|gb|A

ゲノムDNA起源:Celera_asm5h90000642045264

cDNA起源:gi|13639688

【0390】

SGPr104_1, 配列番号21, 配列番号80

Genewiseオルトログ:1.) gi|76762200|ref

ゲノムDNA起源:HGP_sgi|12039078|4

cDNA起源:NP_055508.1

【0391】

SGPr303, 配列番号22, 配列番号81

Genewiseオルトログ:CAA10709.1

ゲノムDNA起源:HGP_sgi|8082389_31

【0392】

SGPr402_1, 配列番号23, 配列番号82

10

20

30

40

50

Genewise オルトログ: A54306, I77530, A45357

ゲノムDNA 起源: Cclera_asm5h92000004018126

【0393】

SGPr434, 配列番号24, 配列番号83

Genewise オルトログ: gi|6755819, gi|6912728, gi|8570164

ゲノムDNA 起源: 90000628646128, 160000117588372, 165000100269164, 90000628646080

cDNA 起源: gi|6141221, gi|3754092, Incyte18565892CB1

10

【0394】

SGPr446_1, 配列番号25, 配列番号84

Genewise オルトログ: gi_11055972, gi_12839280, gi_13633971

ゲノムDNA 起源: 90000628646080

【0395】

SGPr447, 配列番号26, 配列番号85

Genewise オルトログ: gi|12855280, gi|11055972, gi|8777606,

ゲノムDNA 起源: 90000628729589

20

【0396】

SGPr432_1, 配列番号27, 配列番号86

Genewise オルトログ: gi_11181573, gi_12832944, gi_13124769, gi_13277969, gi_13632973

ゲノムDNA 起源: 90000631961624

cDNA 起源: IncyteEST474674.1

【0397】

SGPr529, 配列番号28, 配列番号87

Genewise オルトログ: NP_002767, AAH02100

ゲノムDNA 起源: Cclera_asm5h92000003497776

30

cDNA 起源: gil4506157

【0398】

SGPr428_1, 配列番号29, 配列番号88

Genewise オルトログ: gi|12838473, gi|12839985, gi|9651113, gi|4165315

ゲノムDNA 起源: 90000627342893

【0399】

SGPr425, 配列番号30, 配列番号89

Genewise オルトログ: gi_12844896. gi_6005882 ゲノムDNA 起源: 181000004221955

40

cDNA 起源: Incyte 配列400833.1

【0400】

SGPr548, 配列番号31, 配列番号90

Genewise オルトログ: gi|9957760, gi|5803199, gi|6681654

ゲノムDNA 起源: 92000003497776, gi|11178143

cDNA 起源: gi|9957759

【0401】

SGPr396, 配列番号32, 配列番号91

Genewise オルトログ: gi_11055972, gi_12839280, gi 50

__6680267, gi__8393560, gi__9757698

ゲノムDNA起源: 90000632590917

cDNA起源: Incyte 配列7480224CB1

【0402】

SGPr426, 配列番号33, 配列番号92

Genewise オルトログ: gi__13640890, gi__13646365, gi__7661558

ゲノムDNA起源: 90000641479138

cDNA起源: Incyte 配列7481056CB1

【0403】

10

SGPr552, 配列番号34, 配列番号93

Genewise オルトログ: gi|7661558, gi|4758508, ゲノムDNA起源: 90000641479138

【0404】

SGPr405, 配列番号35, 配列番号94

Genewise オルトログ: gi__7415931, gi__126839, gi__136423, gi__13183572,

ゲノムDNA起源: gi|13509126

cDNA起源: Incyte 配列7474351CB1 および134360.1

【0405】

20

SGPr485__1, 配列番号36, 配列番号95

Genewise オルトログ: gi|9651113

ゲノムDNA起源: 90000627342893

cDNA起源: Incyte 配列6026494CA2

【0406】

SGPr534, 配列番号37, 配列番号96

Genewise オルトログ: gi|4503135

ゲノムDNA起源: 92000004436076, 165000101932709, 92000004433469

cDNA起源: Incyte ESTs: 1383391.20, 1383391.10, 1383391.13, 7691434H1, 2070278CB1, 741522CA2; NCBI ESTs: gi|7260671, gi|7260006, gi|7260642, gi|7259962, gi|2018619, gi|7260655, gi|2019751

30

【0407】

SGPr390, 配列番号38, 配列番号97

Genewise オルトログ: BAB23684

ゲノムDNA起源: hCG22693

【0408】

SGPr521, 配列番号39, 配列番号98

40

Genewise オルトログ: BAB55604, AAF01139, AAF01139

ゲノムDNA起源: HGP__sg|11178143__10

cDNA起源: gi|4826949

【0409】

SGPr530__1, 配列番号40, 配列番号99

Genewise オルトログ: gi__12314133, NP__033381.13, NP__033382.1

ゲノムDNA起源: Celerasm5h181000001848433

【0410】

SGPr520, 配列番号41, 配列番号100

50

Genewise オルトログ: gi | 12839535, gi | 1352368, gi | 4506151
ゲノムDNA 起源: 90000640807190
cDNA 起源: ESTs gi | 13745759, 7472044CB1, 7474338CB1, gi | 13703426, gi | 5392427, gi | 2142177, gi | 2103202, LIB4218-103-R1-K1-H5, LIB4218-085-Q1-K1-C6, LIB4752-019-R1-K1-H4
【0411】
SGPr455, 配列番号42, 配列番号101
Genewise オルトログ: gi | 7512178, gi | 7512176 10
ゲノムDNA 起源: 90000641321557
cDNA 起源: Incyte テンプレート987279. 1
【0412】
SGPr507__2, 配列番号43, 配列番号102
Genewise オルトログ: gi | 13385812, gi | 12854692, gi | 2499862
ゲノムDNA 起源: 90000642611957
【0413】
SGPr559, 配列番号44, 配列番号103
Genewise オルトログ: XP__016993, BAB20079 20
ゲノムDNA 起源: Celerasm5h335001064013332 cDNA
起源: gi | 13173471
【0414】
SGPr567__1, 配列番号45, 配列番号104
Genewise オルトログ: NP__114435, Q9JIQ8
ゲノムDNA 起源: Celerasm5h90000642045213
cDNA 起源: gi | 1404983 | ref | NM__032046. 1
【0415】
SGPr479__1, 配列番号46, 配列番号105
Genewise オルトログ: NP__114154, NP__038949, NP__033 30
382
ゲノムDNA 起源: 90000624931837
cDNA 起源: EST gi | 13997890 および Incyte EST 7480124 CB1,
【0416】
SGPr489__1, 配列番号47, 配列番号106
Genewise オルトログ: gi | 7512176, gi | 7512178, gi | 757698.,
ゲノムDNA 起源: 90000628565500
【0417】 40
SGPr465__1, 配列番号48, 配列番号107
Genewise オルトログ: gi | 6678293, gi | 6678295, ゲノムDNA
NA 起源: gi | 13431162
【0418】
SGPr524__1, 配列番号49, 配列番号108
Genewise オルトログ: gi | 12836503, gi | 10257390, gi | 11415040
ゲノムDNA 起源: 90000626428259
【0419】
SGPr422, 配列番号50, 配列番号109 50

Genewise オルトログ: gi | 7661558, gi | 4758508
ゲノムDNA 起源: 90000641479138

【0420】

SGPr538, 配列番号51, 配列番号110

Genewise オルトログ: NP_110397, Q9ER04, NP_109634
ゲノムDNA 起源: Celerasm5h90000642044035 および 90000642045412

cDNA 起源: gi | 13540535

【0421】

SGPr527_1, 配列番号52, 配列番号111

Genewise オルトログ: gi | 1111573, gi | 13277969, gi | 10441463

ゲノムDNA 起源: 90000631961624

cDNA 起源: Incyte2751509CB1

【0422】

SGPr542, 配列番号53, 配列番号112

Genewise オルトログ: gi | 1705760, gi | 4885369

ゲノムDNA 起源: 92000004018116, gi | 2896799, 92000004013323, 92000004013330, 165000100427031

【0423】

SGPr551, 配列番号54, 配列番号113

Genewise オルトログ: BAB23684.1, NP_035306.2, BAB03502.1

ゲノムDNA 起源: Celerasm5h90000643090998

【0424】

SGPr451, 配列番号55, 配列番号114

Genewise オルトログ: gi_5002340, gi | 12018322, gi | 1480413,

ゲノムDNA 起源: 181000000828193

【0425】

SGPr452_1, 配列番号56, 配列番号115

Genewise オルトログ: gi | 13183572, gi | 339983, gi | 7415931

ゲノムDNA 起源: 92000004034678

【0426】

SGPr504, 配列番号57, 配列番号116

Genewise オルトログ: gi | 1633237

ゲノムDNA 起源: Celerasm5h92000004018137

【0427】

SGPr469, 配列番号58, 配列番号117

Genewise オルトログ: BAB30277, CAB41988, XP_016204

ゲノムDNA 起源: GA_x2HTBKPYSW7D

【0428】

SGPr400, 配列番号59, 配列番号118

Genewise オルトログ: gi | 6755819, gi | 6912728

ゲノムDNA 起源: 90000632590917

【0429】

新規プロテアーゼポリヌクレオチドの説明

SGPr397, 配列番号1, 配列番号60は948ヌクレオチドの長さである。オーブ 50

ンリーディングフレームは位置1で開始し、位置948で終了し、948ヌクレオチドの長さのORFを与える。予測蛋白質は315アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、カルボキシペプチダーゼ、Znカルボキシペプチダーゼ。この遺伝子の細胞遺伝学的位置は8q12である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：AV763490。

【0430】

SGPr413, 配列番号2, 配列番号61は1125ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置1125で終了し、1125ヌクレオチドの長さのORFを与える。予測蛋白質は374アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、カルボキシペプチダーゼ、Znカルボキシペプチダーゼ。この遺伝子の細胞遺伝学的位置は2q35である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：なし。

10

【0431】

SGPr404, 配列番号3, 配列番号62は1590ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置1590で終了し、1590ヌクレオチドの長さのORFを与える。予測蛋白質は529アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、カルボキシペプチダーゼ、Znカルボキシペプチダーゼ。この遺伝子の細胞遺伝学的位置は10q26である。このヌクレオチド配列は以下の一塩基多型を含む（SNPの受託番号がアレル位置とともに与えられており、次に遺伝子中のSNPのまわりの配列が与えられる）：ss1782198__alleles=201, agaaaggoc t a y g a a g g g g。SNPss1782198はORFのヌクレオチド612（aa58）で生ずる（CまたはT＝サイレント；AA204＝いずれのヌクレオチドについてもチロシン）。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：AA045748, AA148684, AA047483。核酸は短い反復配列を含む（反復の位置および配列）：477ggagc t g c t g c t g c t g c t g g t g 498。

20

【0432】

SGPr536__1, 配列番号4, 配列番号63は1404ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置1404で終了し、1404ヌクレオチドの長さのORFを与える。予測蛋白質は467アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、システイン、パパイン。この遺伝子の細胞遺伝学的位置は1p35である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：AL542213, AL547246, AL552037。核酸は短い反復配列を含む（反復の位置および配列）：480g c t g c t g c t g c t g c t g g t g c a g 501。

30

【0433】

SGPr414, 配列番号5, 配列番号64は10062ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置10062で終了し、10062ヌクレオチドの長さのORFを与える。予測蛋白質は3353アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、システイン、UCH2b。この遺伝子の細胞遺伝学的位置は2p14である。このヌクレオチド配列は以下の一塩基多型を含む（SNPの受託番号が与えられており、次に遺伝子中のSNPのまわりの配列が与えられる）：ss16542__alleles=101, c t a c c c t a g c y g a g g a a g a。SNPss16542はORFのヌクレオチド9807（aa3269, アラニン）で生ずる。SNPはサイレントである。この配列は公共ESTsのデータベース（dbEST

40

50

）において以下のE S T sにより表される：A U 1 1 8 2 3 7，A U 1 3 1 4 2 0，A U 1 2 5 0 8 3。核酸は短い反復配列を含む（反復の位置および配列）：2 2 4 9 a c c a c c a c c a c c a c c a c c a t c a c c a c c a c c a c c a c 2 2 8 0。

【0434】

S G P r 4 3 0，配列番号6，配列番号65は2943ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置2943で終了し，2943ヌクレオチドの長さのORFを与える。予測蛋白質は980アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置は2 q 3 7である。このヌクレオチド配列は以下の一塩基多型を含む（S N Pの受託番号がアレル位置とともに与えられており，次に遺伝子中のS N Pのまわりの配列が与えられる）：s s 1 5 3 4 5 8 5 _ a l l e l e P o s = 5 1，t g g a a t a r c t c g g a c；r s 1 0 5 5 6 8 7 _ a l l e l e P o s = 5 1，t g g t a a t c c g k g t a g a g g。S N P s s 1 5 3 4 5 8 5はORFのヌクレオチド538（a a 1 8 0）で生ずる（AまたはG）。S N P s s 1 5 3 4 5 8 5はアミノ酸180を変更する。ヌクレオチド538がアデニンである場合，アミノ酸180はトレオニンであり，グアニンである場合，アミノ酸180はアラニンである。第2のS N P，r s 1 0 5 5 6 8 7はヌクレオチド499でGまたはTをコードする。r s 1 0 5 5 6 8 7は遺伝子のアミノ酸配列を変更する。ヌクレオチド499がチミジンである場合アミノ酸167はシステインであり，ヌクレオチド499がグアニンである場合アミノ酸167はグリシンである。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：W 8 7 6 6 6，A I 0 7 6 1 0 8，B G 6 1 2 8 6 4。

【0435】

S G P r 4 9 6 _ 1，配列番号7，配列番号66は2862ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置2862で終了し，2862ヌクレオチドの長さのORFを与える。予測蛋白質は953アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置はX p 1 1. 4である。このヌクレオチド配列は以下の一塩基多型を含む（S N Pの受託番号がアレル位置とともに与えられており，次に遺伝子中のS N Pのまわりの配列が与えられる）：s s 1 0 2 9 7 5 6 _ a l l e l e P o s = 1 0 1，a g a g a a a t a y g a g g g t a t t。S N P s s 1 0 2 9 7 5 6はヌクレオチド351においてCまたはTをコードする。いずれのヌクレオチドについてもアミノ酸117はチロシンであり，したがってS N Pはサイレントである。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：A W 8 5 1 0 6 6，A W 8 5 1 0 6 5，A W 8 5 1 0 7 6。

【0436】

S G P r 4 9 5，配列番号8，配列番号67は2352ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置2352で終了し，2352ヌクレオチドの長さのORFを与える。予測蛋白質は783アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置は6 q 1 6である。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：A L 5 5 9 9 6 0，A L 5 3 0 4 7 0，A L 5 1 6 1 8 4。

【0437】

S G P r 4 0 7，配列番号9，配列番号68は2259ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置2259で終了し，2259ヌクレオチドの長さのORFを与える。予測蛋白質は752アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ

／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置は2 q 3 7である。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：なし。

【0 4 3 8】

S G P r 4 5 3，配列番号1 0，配列番号6 9は2 1 3 9ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置2 1 3 9で終了し，2 1 3 9ヌクレオチドの長さのO R Fを与える。予測蛋白質は7 1 2アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置は1 2 q 2 3である。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：B G 7 2 2 4 3 6，A I 9 2 7 8 8 1，B G 7 7 1 8 8 8。核酸は短い反復配列を含む（反復の位置および配列）：5 5 3 g t a g t a a a a g a g a a g t a a a 5 7 2。

【0 4 3 9】

S G P r 4 4 5，配列番号1 1，配列番号7 0は8 7 0ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置8 7 0で終了し，8 7 0ヌクレオチドの長さのO R Fを与える。予測蛋白質は2 8 9アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置は6 q 1 6である。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：A L 5 5 9 9 6 0，A L 5 3 0 4 7 0，A L 5 1 6 1 8 4。

【0 4 4 0】

S G P r 4 0 1 __ 1，配列番号1 2，配列番号7 1は1 1 0 1ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置1 1 0 1で終了し，1 1 0 1ヌクレオチドの長さのO R Fを与える。予測蛋白質は3 6 6アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置は4 q 1 1である。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：A U 1 2 4 8 9 8，A U 1 3 4 5 5 3，A I 2 6 9 0 6 9。

【0 4 4 1】

S G P r 4 0 8，配列番号1 3，配列番号7 2は3 8 6 4ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置3 8 6 4で終了し，3 8 6 4ヌクレオチドの長さのO R Fを与える。予測蛋白質は1 2 8 7アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置は1 1 p 1 5である。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：B G 7 4 1 1 9 0，B F 5 7 5 4 9 8，B G 1 7 0 8 2 9。

【0 4 4 2】

S G P r 4 8 0，配列番号1 4，配列番号7 3は4 8 1 5ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置4 8 1 5で終了し，4 8 1 5ヌクレオチドの長さのO R Fを与える。予測蛋白質は1 6 0 4アミノ酸の長さである。この配列は部分蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，システイン，U C H 2 b。この遺伝子の細胞遺伝学的位置は1 7 q 2 4である。この配列は公共E S T sのデータベース（d b E S T）において以下のE S T sにより表される：A U 1 3 1 7 4 8，A U 1 2 0 3 8 1，B G 4 2 0 7 6 6。

【0 4 4 3】

S G P r 4 3 1，配列番号1 5，配列番号7 4は3 1 2 9ヌクレオチドの長さである。オ

オープンリーディングフレームは位置1で開始し、位置3129で終了し、3129ヌクレオチドの長さのORFを与える。予測蛋白質は1042アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、システイン、UCH2b。この遺伝子の細胞遺伝学的位置は4q31.3である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：BG575871, BG113469, BG112979。

【0444】

SGPr429, 配列番号16, 配列番号75は3102ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置3102で終了し、3102ヌクレオチドの長さのORFを与える。予測蛋白質は1033アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、システイン、UCH2b。この遺伝子の細胞遺伝学的位置は1p36.2である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：AL518266, BG681225, BG217186。

10

【0445】

SGPr503, 配列番号17, 配列番号76は1554ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置1554で終了し、1554ヌクレオチドの長さのORFを与える。予測蛋白質は517アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、システイン、UCH2b。この遺伝子の細胞遺伝学的位置は12q24.3である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：BG678894, BG476418, BE264732。核酸は短い反復配列を含む（反復の位置および配列）：1534gagtgcaagtcctgaagaatg1553。

20

【0446】

SGPr427, 配列番号18, 配列番号77は3372ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置3372で終了し、3372ヌクレオチドの長さのORFを与える。予測蛋白質は1123アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、システイン、UCH2b。この遺伝子の細胞遺伝学的位置は17p13である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：BG831111, AW996553, BE614914。

30

【0447】

SGPr092, 配列番号19, 配列番号78は786ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置786で終了し、786ヌクレオチドの長さのORFを与える。予測蛋白質は261アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、メタロプロテアーゼ, PepM10。この遺伝子の細胞遺伝学的位置は11p15である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：BG189720, AW966183, BG198356。

40

【0448】

SGPr359, 配列番号20, 配列番号79は1452ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し、位置1452で終了し、1452ヌクレオチドの長さのORFを与える。予測蛋白質は483アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ、メタロプロテアーゼ, PepM10。この遺伝子の細胞

50

遺伝学的位置は11q22である。この配列は公共ESTsのデータベース(d b E S T)において以下のESTsにより表される: BG187290。

【0449】

SGPr104__1, 配列番号21, 配列番号80は2298ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置2298で終了し, 2298ヌクレオチドの長さのORFを与える。予測蛋白質は765アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, メタロプロテアーゼ, PepM13。この遺伝子の細胞遺伝学的位置は3q27である。この配列は公共ESTsのデータベース(d b E S T)において以下のESTsにより表される: BF511209, AW341249, AL119270。

10

【0450】

SGPr303, 配列番号22, 配列番号81は1257ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置1257で終了し, 1257ヌクレオチドの長さのORFを与える。予測蛋白質は418アミノ酸の長さである。この配列は全長触媒ドメインをコードする。これは次のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, メタロプロテアーゼ, PepM2。この遺伝子の細胞遺伝学的位置は17q11.1である。この配列は公共ESTsのデータベース(d b E S T)において以下のESTsにより表される: AU138954, BG251083, AW161660。

20

【0451】

SGPr402__1, 配列番号23, 配列番号82は2268ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置2268で終了し, 2268ヌクレオチドの長さのORFを与える。予測蛋白質は755アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, ズブチラーゼ。この遺伝子の細胞遺伝学的位置は19q11である。この配列は公共ESTsのデータベース(d b E S T)において以下のESTsにより表される: AL041695, AA454137, BG719638。

30

【0452】

SGPr434, 配列番号24, 配列番号83は1176ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置1176で終了し, 1176ヌクレオチドの長さのORFを与える。予測蛋白質は391アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は3p21である。この配列は公共ESTsのデータベース(d b E S T)において以下のESTsにより表される: AW137088, BF593342。

【0453】

SGPr446__1, 配列番号25, 配列番号84は681ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置681で終了し, 681ヌクレオチドの長さのORFを与える。予測蛋白質は226アミノ酸の長さである。この配列は全長触媒ドメインをコードする。これは以下のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は3p21である。この配列は公共ESTsのデータベース(d b E S T)において以下のESTsにより表される: AW243584。核酸は短い反復配列を含む(反復の位置および配列): 798gggtgggcattcagctgggg818。

40

【0454】

SGPr447, 配列番号26, 配列番号85は888ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置888で終了し, 888ヌクレオチドの長さのORFを与える。予測蛋白質は295アミノ酸の長さである。この配列は部分蛋

50

白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は16p13.3である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：なし。

【0455】

SGPr432__1，配列番号27，配列番号86は1887ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置1887で終了し，1887ヌクレオチドの長さのORFを与える。予測蛋白質は628アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は不明である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：BE264142，BG474605，BF304202。 10

【0456】

SGPr529，配列番号28，配列番号87は831ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置831で終了し，831ヌクレオチドの長さのORFを与える。予測蛋白質は276アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は19q13.4である。このヌクレオチド配列は以下の一塩基多型を含む（SNPの受託番号がアレル位置とともに与えられており，次に遺伝子中のSNPのまわりの配列が与えられる）：ss1550333__allele1Pos=51，taggggatgaycaccctgct；ss1546197__allele1Pos=51，gccggacsactcgc。SNPss1550333はヌクレオチド297においてCまたはTをコードする。いずれのヌクレオチドの場合についてもアミノ酸99はアスパラギン酸であり，したがってSNPはサイレントである。別のSNPss1546197が存在し，これは位置336においてCまたはGをコードする。いずれのヌクレオチドが存在する場合にもアミノ酸112はトレオニンであり，したがって，このSNPはサイレントである。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：BE898352，BG469321。 20

【0457】

SGPr428__1，配列番号29，配列番号88は858ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置858で終了し，858ヌクレオチドの長さのORFを与える。予測蛋白質は285アミノ酸の長さである。この配列は全長触媒ドメインをコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は8p23である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：なし。核酸は短い反復配列を含む（反復の位置および配列）：473catgcacctggaaaagctg491。 30

【0458】

SGPr425，配列番号30，配列番号89は1242ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置1242で終了し，1242ヌクレオチドの長さのORFを与える。予測蛋白質は413アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は6q14である。このヌクレオチド配列は以下の一塩基多型を含む（SNPの受託番号がアレル位置とともに与えられており，次に遺伝子中のSNPのまわりの配列が与えられる）：ss674620__allele1Pos=201，gagcatctgcVggaggagg。SNPss674620はヌクレオチド671においてGまたはAまたはCをコードする。ヌクレオチドがグアニンである場合，アミノ酸224はアルギニンである。ヌクレオチドがアデニンである場合，アミノ酸224はグルタミンである。ヌクレオチ 40 50

ドがシトシンである場合、アミノ酸 2 2 4 はプロリンである。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: A L 5 5 1 2 8 6, A A 4 4 5 9 4 8, A A 4 2 4 0 7 3。核酸は短い反復配列を含む (反復の位置および配列): 1 1 1 1 t c a g g g c a c c a g t g g g t g g a 1 1 3 0。

【0 4 5 9】

S G P r 5 4 8, 配列番号 3 1, 配列番号 9 0 は 9 6 3 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 9 6 3 で終了し, 9 6 3 ヌクレオチドの長さの O R F を与える。予測蛋白質は 3 2 0 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 1 9 q 1 3. 4 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: なし。

10

【0 4 6 0】

S G P r 3 9 6, 配列番号 3 2, 配列番号 9 1 は 9 8 7 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 9 8 7 で終了し, 9 8 7 ヌクレオチドの長さの O R F を与える。予測蛋白質は 3 2 8 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 4 q 3 2 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: なし。

20

【0 4 6 1】

S G P r 4 2 6, 配列番号 3 3, 配列番号 9 2 は 1 2 7 8 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 1 2 7 8 で終了し, 1 2 7 8 ヌクレオチドの長さの O R F を与える。予測蛋白質は 4 2 5 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 4 q 1 3 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: なし。

【0 4 6 2】

S G P r 5 5 2, 配列番号 3 4, 配列番号 9 3 は 6 6 6 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 6 6 6 で終了し, 6 6 6 ヌクレオチドの長さの O R F を与える。予測蛋白質は 2 2 1 アミノ酸の長さである。この配列は全長触媒ドメインをコードする。これは以下のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 4 q 1 3 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: なし。

30

【0 4 6 3】

S G P r 4 0 5, 配列番号 3 5, 配列番号 9 4 は 2 8 4 7 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 2 8 4 7 で終了し, 2 8 4 7 ヌクレオチドの長さの O R F を与える。予測蛋白質は 9 4 8 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 1 6 p 1 3. 3 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: なし。

40

【0 4 6 4】

S G P r 4 8 5 _ 1, 配列番号 3 6, 配列番号 9 5 は 1 0 5 9 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 1 0 5 9 で終了し, 1 0 5 9 ヌクレオチドの長さの O R F を与える。予測蛋白質は 3 5 2 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的

50

位置は8 p 2 3である。このヌクレオチド配列は以下の一塩基多型を含む（SNPの受託番号がアレル位置とともに与えられており，次に遺伝子中のSNPのまわりの配列が与えられる）：ss1532791__allelePos=51, tggagakaagaaacc。ss1532791は位置834においてGまたはTをコードする。この多型はアミノ酸278を変更する。834のヌクレオチドがグアニンである場合，アミノ酸278はグルタミン酸（E）である。ヌクレオチドがチミンである場合，アミノ酸278はアスパラギン酸（D）である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：AA781356。

【0465】

SGPr534, 配列番号37, 配列番号96は792ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置792で終了し，792ヌクレオチドの長さのORFを与える。予測蛋白質は263アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は16 q 2 3である。このヌクレオチド配列は以下の6個の一塩基多型を含む（SNPの受託番号がアレル位置とともに与えられており，次に遺伝子中のSNPのまわりの配列が与えられる）：ss1522946__allelePos=51, gctctaccwccacgcc; ss1522943__allelePos=51, cgcaacctgctcyaccaccac; ss1522933__allelePos=51, ctgccagagagayggagcctgg; ss1522931__allelePos=51, 全長=101, gctctgccaraaggacg; ss1522930__allelePos=51, gggctgactctggmggcccccct; ss1522928__allelePos=51, tgcattgggygactctgg。SNPss1522946は位置721においてAまたはTをコードする。721がアデニンである場合，アミノ酸241はトレオニン（T）である。721がチミンである場合，アミノ酸241はセリン（S）である。SNPss1522943は位置717においてCまたはTをコードする。このSNPはサイレントである（239＝セリン）。SNPss1522933は666においてCまたはTをコードする。このSNPはサイレントである（222＝アスパラギン酸）。SNPss1522931は位置660においてAまたはGをコードする。このSNPはサイレントである（220＝グルタミン）。SNPss1522930は位置642においてAまたはCをコードする。このSNPはサイレントである（214＝グリシン）。SNPss1522928は位置633においてCまたはTをコードする。このSNPはサイレントである（211＝グリシン）。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：AW583018, AW582942, AW960025。核酸は短い反復配列を含む（反復の位置および配列）：172cacttctgcgggggctccctcatc195。

【0466】

SGPr390, 配列番号38, 配列番号97は3387ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置3387で終了し，3387ヌクレオチドの長さのORFを与える。予測蛋白質は1128アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は19 q 1 1である。このヌクレオチド配列は以下の一塩基多型を含む（SNPの受託番号がアレル位置とともに与えられており，次に遺伝子中のSNPのまわりの配列が与えられる）：ss82431__allelePos=99, gccgtgarcacactg; ss1320361__allelePos=225, agcggccascattggcgt。ss82431は位置2585においてAまたはGをコードする。このヌクレオチドがアデニンである場合，アミノ酸862はアスパラギン（N）である。このヌクレオチドがグアニンである場合，アミノ酸862はセリンである。SNPss1320361は位置89においてCまたはGをコードする。位置89がシトシンである場合，アミ

ノ酸 30 はトレオニン (T) である。位置 89 がグアニンである場合、アミノ酸 30 はセリンである。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: C 1 6 6 0 7。

【0467】

S G P r 5 2 1, 配列番号 39, 配列番号 98 は 762 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 762 で終了し, 762 ヌクレオチドの長さの O R F を与える。予測蛋白質は 253 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 19 q 13.4 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: A A 5 4 2 9 9 4, B E 7 1 3 3 7 9, W 5 8 7 3 7。核酸は短い反復配列を含む (反復の位置および配列): 6 4 6 c a a g g t c t g g t g t c c t g g g g 6 6 5。 10

【0468】

S G P r 5 3 0 _ 1, 配列番号 40, 配列番号 99 は 816 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 816 で終了し, 816 ヌクレオチドの長さの O R F を与える。予測蛋白質は 271 アミノ酸の長さである。この配列は全長触媒ドメインをコードする。これは以下のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 9 q 22 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: なし。 20

【0469】

S G P r 5 2 0, 配列番号 41, 配列番号 100 は 1737 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 1737 で終了し, 1737 ヌクレオチドの長さの O R F を与える。予測蛋白質は 578 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 2 q 37 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: なし。

【0470】

S G P r 4 5 5, 配列番号 42, 配列番号 101 は 2913 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 2913 で終了し, 2913 ヌクレオチドの長さの O R F を与える。予測蛋白質は 970 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 12 p 11.2 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: A W 4 5 0 1 5 5, A W 9 9 5 4 9 6。 30

【0471】

S G P r 5 0 7 _ 2, 配列番号 43, 配列番号 102 は 798 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 798 で終了し, 798 ヌクレオチドの長さの O R F を与える。予測蛋白質は 265 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される (スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は 7 q 36 である。この配列は公共 E S T s のデータベース (d b E S T) において以下の E S T s により表される: B G 2 1 7 7 2 4, B G 2 1 9 7 3 8, B G 1 9 2 7 0 9。 40

【0472】

S G P r 5 5 9, 配列番号 44, 配列番号 103 は 1365 ヌクレオチドの長さである。オープンリーディングフレームは位置 1 で開始し, 位置 1365 で終了し, 1365 ヌクレオチドの長さの O R F を与える。予測蛋白質は 454 アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される (スーパーファミリー/グル 50

ープ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は2 1 q 2 2である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：A I 9 7 8 8 7 4，A I 4 6 9 0 9 5，B F 4 3 5 6 7 0。

【0 4 7 3】

SGPr 5 6 7 __ 1，配列番号4 5，配列番号1 0 4は1 6 1 4ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置1 6 1 4で終了し，1 6 1 4ヌクレオチドの長さのORFを与える。予測蛋白質は5 3 7アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は1 1 q 2 3である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：B E 7 3 2 3 8 1，R 7 8 5 8 1，A W 8 4 5 1 0 6。

【0 4 7 4】

SGPr 4 7 9 __ 1，配列番号4 6，配列番号1 0 5は9 8 1ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置9 8 1で終了し，9 8 1ヌクレオチドの長さのORFを与える。予測蛋白質は3 2 6アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は1 q 4 2である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：B G 7 1 8 7 0 3，A A 4 0 1 7 0 5，A A 3 9 8 1 7 0。核酸は短い反復配列を含む（反復の位置および配列）：7 8 0 t g g a a t t g t g a g c t g g g g c c g 8 0 0。

【0 4 7 5】

SGPr 4 8 9 __ 1，配列番号4 7，配列番号1 0 6は1 6 7 1ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置1 6 7 1で終了し，1 6 7 1ヌクレオチドの長さのORFを与える。予測蛋白質は5 5 6アミノ酸の長さである。この配列は全長触媒ドメインをコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は1 1 p 1 5である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：A W 2 7 1 4 3 0，A W 2 3 7 8 9 3。

【0 4 7 6】

SGPr 4 6 5 __ 1，配列番号4 8，配列番号1 0 7は8 9 4ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置8 9 4で終了し，8 9 4ヌクレオチドの長さのORFを与える。予測蛋白質は2 9 7アミノ酸の長さである。この配列は全長触媒ドメインをコードする。これは以下のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は不明である。この配列は公共ESTsのデータベース（dbEST）において以下のESTsにより表される：なし。

【0 4 7 7】

SGPr 5 2 4 __ 1，配列番号4 9，配列番号1 0 8は2 5 5 3ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し，位置2 5 5 3で終了し，2 5 5 3ヌクレオチドの長さのORFを与える。予測蛋白質は8 5 0アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される（スーパーファミリー／グループ／ファミリー）：プロテアーゼ，セリン，トリプシン。この遺伝子の細胞遺伝学的位置は不明である。このヌクレオチド配列は以下の一塩基多型を含む（SNPの受託番号がアレル位置とともに与えられており，次に遺伝子中のSNPのまわりの配列が与えられる）：s s 2 0 1 3 5 5 8 __ a l l e l e P o s = 2 0 1，g a c a t g g a w g t g g a c g a c；s s 2 0 1 4 1 2 8 __ a l l e l e P o s = 3 5 8，a c a a t t t t t y g a g t g c c c a。s s 2 0 1 3 5 5 8は位置6 7 5においてTまたはCをコードする

10

20

30

40

50

。これはサイレントSNPである。ss2014128は1369においてCまたはTをコードする。ヌクレオチドがシトシンである場合、アミノ酸457はアルギニンである。ヌクレオチド1369がチミンである場合、停止コドンが導入され、蛋白質が456アミノ酸に切断される。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される：なし。核酸は短い反復配列を含む(反復の位置および配列)：711aaaaaaagaaaaagaaaaa734。

【0478】

SGPr422, 配列番号50, 配列番号109は1344ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置1344で終了し, 1344ヌクレオチドの長さのORFを与える。予測蛋白質は447アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される(スーパーファミリー/グループ/ファミリー)：プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は4q13である。このヌクレオチド配列は以下の一塩基多型を含む(SNPの受託番号がアレル位置とともに与えられており, 次に遺伝子中のSNPのまわりの配列が与えられる)：ss1091793__allelePos=101, acatagccrgatttggttg; ss448607__allelePos=101, tgggagccrggtccctgct。SNPss1091793は位置956においてアデニンまたはグアニンをコードする。956がグアニンである場合, アミノ酸319はアルギニン(R)である。ヌクレオチド956がアデニンである場合, アミノ酸319はグルタミン(Q)である。SNPss448607は位置552においてAまたはGをコードする。これはサイレントである(アミノ酸184=アラニン)。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される：なし。

【0479】

SGPr538, 配列番号51, 配列番号110は1374ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置1374で終了し, 1374ヌクレオチドの長さのORFを与える。予測蛋白質は457アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される(スーパーファミリー/グループ/ファミリー)：プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は11q23である。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される：AL538140, BF934870。核酸は短い反復配列を含む(反復の位置および配列)：545tgggaggcttcctggaggagg564。

【0480】

SGPr527__1, 配列番号52, 配列番号111は2457ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置2457で終了し, 2457ヌクレオチドの長さのORFを与える。予測蛋白質は818アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される(スーパーファミリー/グループ/ファミリー)：プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は不明である。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される：AW450407, AI190509, AI864473。

【0481】

SGPr542, 配列番号53, 配列番号112は855ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置855で終了し, 855ヌクレオチドの長さのORFを与える。予測蛋白質は284アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される(スーパーファミリー/グループ/ファミリー)：プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は19q13.1である。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される：なし。

【0482】

SGPr 551, 配列番号54, 配列番号113は2409ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置2409で終了し, 2409ヌクレオチドの長さのORFを与える。予測蛋白質は802アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は22q13である。このヌクレオチド配列は以下の一塩基多型を含む(SNPの受託番号がアレル位置とともに与えられており, 次に遺伝子中のSNPのまわりの配列が与えられる): rs881144__allelePos=200, ctgcagccctayg ccgagagg; rs855791__allelePos=101, agcgaaggyctatcgcta。SNPrs881144は位置1227においてCまたはTをコードする。これはサイレントSNPである(409=チロシン)。SNPrs855791は位置2180においてCまたはTをコードする。2180のヌクレオチドがシトシンである場合, アミノ酸727はアラニンである。ヌクレオチドがチミンである場合, アミノ酸727はバリンである。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される: AV693114, N70418, AA609066。

【0483】

SGPr 451, 配列番号55, 配列番号114は1080ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置1080で終了し, 1080ヌクレオチドの長さのORFを与える。予測蛋白質は359アミノ酸の長さである。この配列は全長蛋白質をコードする。これは次のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は12q23である。このヌクレオチド配列は以下の一塩基多型を含む(SNPの受託番号がアレル位置とともに与えられており, 次に遺伝子中のSNPのまわりの配列が与えられる): ss1881349__allelePos=201, gggcgcatagcaraagg; ss1266911__allelePos=101, ccactgcactaaagacrcctag。SNPss1881349は位置217においてAまたはGをコードする。217のヌクレオチドがアデニンである場合, アミノ酸73はリシン(K)である。ヌクレオチドがグアニンである場合, アミノ酸73はグルタミン酸(E)である。SNPss1266911は位置412においてAまたはGをコードする。412がグアニンである場合, アミノ酸138はアラニン(A)である。412がアデニンである場合, アミノ酸138はトレオニン(T)である。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される: BG722131, BG722203。

【0484】

SGPr 452__1, 配列番号56, 配列番号115は867ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置867で終了し, 867ヌクレオチドの長さのORFを与える。予測蛋白質は288アミノ酸の長さである。この配列は全長蛋白質をコードする。これは以下のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は16p13.3である。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される: なし。

【0485】

SGPr 504, 配列番号57, 配列番号116は135ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置135で終了し, 135ヌクレオチドの長さのORFを与える。予測蛋白質は44アミノ酸の長さである。この配列は部分長蛋白質をコードする。これは次のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は不明である。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される: なし。

10

20

30

40

50

【0486】

SGPr 469, 配列番号58, 配列番号117は138ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置138で終了し, 138ヌクレオチドの長さのORFを与える。予測蛋白質は45アミノ酸の長さである。この配列は部分長蛋白質をコードする。これは以下のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は不明である。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される:

AW753029, Z19070。核酸は短い反復配列を含む(反復の位置および配列): 55gggattgtgagctggggc72。

10

【0487】

SGPr 400, 配列番号59, 配列番号118は930ヌクレオチドの長さである。オープンリーディングフレームは位置1で開始し, 位置930で終了し, 930ヌクレオチドの長さのORFを与える。予測蛋白質は309アミノ酸の長さである。この配列は部分長蛋白質をコードする。これは以下のように分類される(スーパーファミリー/グループ/ファミリー): プロテアーゼ, セリン, トリプシン。この遺伝子の細胞遺伝学的位置は4q32である。この配列は公共ESTsのデータベース(dbEST)において以下のESTsにより表される: なし。

【0488】

新規プロテアーゼポリペプチドの説明

20

SGPr 397, 配列番号1, 配列番号60は, 315アミノ酸の長さの蛋白質をコードする。これはZnカルボキシペプチダーゼファミリーのカルボキシペプチダーゼプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸139からアミノ酸280までZnカルボキシペプチダーゼ(PF00246)ドメインについての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置146である。この蛋白質中で同定された他のドメインは以下のとおりである: カルボキシペプチダーゼ活性化ペプチド(PF02244), アミノ酸41-120。プロセグメント成分(活性化ペプチド)はプロ酵素の折り畳みおよび活性の調節を担う(http://pfam.wustl.edu/cgi-bin/getdesc?name=Propep_M14を参照)。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=3.10E-220; 同一のアミノ酸の数=315; 同一性のパーセント=100%; 類似性のパーセント=100%; NRAA中の最も類似するエントリーの受託番号はNP_065094.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: カルボキシペプチダーゼB前駆体[*Homosapiens*]。

30

【0489】

SGPr 413, 配列番号2, 配列番号61は, 374アミノ酸の長さの蛋白質をコードする。これはZnカルボキシペプチダーゼファミリーのカルボキシペプチダーゼプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸50からアミノ酸291までZnカルボキシペプチダーゼ(PF00246)についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置248である。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=5.90E-93; 同一のアミノ酸の数=146; 同一性のパーセント=49%; 類似性のパーセント=68%; NRAA中の最も類似するエントリーの受託番号はAAF01344.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: (AF190274) カルボキシペプチダーゼホモログ[*Bothrops*

40

50

j a r a r a c a] 。

【0490】

S G P r 4 0 4 , 配列番号 3 , 配列番号 6 2 は , 5 2 9 アミノ酸の長さの蛋白質をコードする。これは Z n カルボキシペプチダーゼファミリーのカルボキシペプチダーゼプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 9 1 からアミノ酸 4 6 6 まで Z n カルボキシペプチダーゼ (P F 0 0 2 4 6) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 2 4 8 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0 , ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである : P スコア = 0 ; 同一のアミノ酸の数 = 5 0 2 ; 同一性のパーセント = 9 4 % ; 類似性のパーセント = 9 8 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 0 6 1 3 5 5 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである : カルボキシペプチダーゼ X 2 [M u s m u s c u l u s] 。

10

【0491】

S G P r 5 3 6 _ 1 , 配列番号 4 , 配列番号 6 3 は , 4 6 7 アミノ酸の長さの蛋白質をコードする。これはパインファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 2 0 3 からアミノ酸 4 5 6 までパイン (P F 0 0 1 1 2) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 3 3 7 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0 , ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである : P スコア = 1 . 1 0 E - 2 7 6 ; 同一のアミノ酸の数 = 4 6 7 ; 同一性のパーセント = 1 0 0 % ; 類似性のパーセント = 1 0 0 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 0 7 1 4 4 7 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである : P 3 E C S L [H o m o s a p i e n s] 。

20

【0492】

S G P r 4 1 4 , 配列番号 5 , 配列番号 6 4 は , 3 3 5 3 アミノ酸の長さの蛋白質をコードする。これは U C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 1 9 5 1 からアミノ酸 2 0 4 5 までユビキチンカルボキシル末端ヒドロラーゼファミリー 2 b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 7 2 である。この蛋白質中で同定された他のドメインは以下のとおりである : ユビキチンカルボキシル末端ヒドロラーゼファミリー 2 (U C H 2 b , P F 0 0 4 4 2) , アミノ酸 1 7 0 1 - 1 7 3 1 。ユビキチンカルボキシル末端ヒドロラーゼ (E C 3 . 1 . 2 . 1 5) (U C H) (脱ユビキチン化酵素) はユビキチンの C - 末端グリシンのペプチド結合を認識し加水分解するチオールプロテアーゼである。これらの酵素はポリユビキチン前駆体ならびにユビキチン化蛋白質のプロセッシングに関与している。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0 , ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである : P スコア = 0 ; 同一のアミノ酸の数 = 1 2 5 9 ; 同一性のパーセント = 9 9 % ; 類似性のパーセント = 1 0 0 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 0 5 5 5 2 4 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである : K I A A 0 5 7 0 遺伝子産物 [H o m o s a p i e n s] 。

30

40

【0493】

S G P r 4 3 0 , 配列番号 6 , 配列番号 6 5 は , 9 8 0 アミノ酸の長さの蛋白質をコードする。これは U C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 8 8 6 からアミノ酸 9 5 1 までユビキチンカル

50

ボキシル末端ヒドロラーゼファミリー 2 b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置 1 からプロファイル位置 7 2 である。この蛋白質中で同定された他のドメインは以下のとおりである: U C H 2 b (P F 0 0 4 4 2), アミノ酸 3 4 2 - 3 7 3。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 0; 同一のアミノ酸の数 = 9 3 0; 同一性のパーセント = 9 9 %; 類似性のパーセント = 9 9 %; N R A A 中の最も類似するエントリーの受託番号は B A B 1 3 4 2 0. 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (A B 0 4 6 8 1 4) K I A A 1 5 9 4 蛋白質 [H o m o s a p i e n s]。 10

【0494】

S G P r 4 9 6 __ 1, 配列番号 7, 配列番号 6 6 は, 9 5 3 アミノ酸の蛋白質をコードする。これは U C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 8 7 5 からアミノ酸 9 3 5 までユビキチンカルボキシル末端ヒドロラーゼファミリー 2 b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置 1 からプロファイル位置 7 2 である。この蛋白質中で同定された他のドメインは以下のとおりである: U C H 2 b (P F 0 0 4 4 2), 5 9 3 - 6 9 4; および Z n - フィンガードメイン (P F 0 2 1 4 8), ユビキチン-ヒドロラーゼ中に見いだされる, 4 6 5 - 5 3 4。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 2. 0 0 E - 1 9 0; 同一のアミノ酸の数 = 4 9 6; 同一性のパーセント = 9 5 %; 類似性のパーセント = 9 8 %; N R A A 中の最も類似するエントリーの受託番号は A A F 6 6 9 5 3. 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (A F 2 2 9 6 4 3) ユビキチン特異的プロテアーゼ [M u s m u s c u l u s]。 20

【0495】

S G P r 4 9 5, 配列番号 8, 配列番号 6 7 は, 7 8 3 アミノ酸の長さの蛋白質をコードする。これは U C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 6 9 5 からアミノ酸 7 8 1 までユビキチンカルボキシル末端ヒドロラーゼファミリー 2 b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置 1 からプロファイル位置 7 2 である。この蛋白質中で同定された他のドメインは以下のとおりである: U C H 2 b (P F 0 0 4 4 2), 1 9 0 - 2 2 1; およびユビキチン-ヒドロラーゼ (P F 0 2 1 4 8) 中の Z n - フィンガー, 4 6 5 - 5 3 4。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 2. 4 0 E - 1 7 6; 同一のアミノ酸の数 = 2 8 2; 同一性のパーセント = 1 0 0 %; 類似性のパーセント = 1 0 0 %; N R A A 中の最も類似するエントリーの受託番号は A A H 0 5 9 9 1. 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (B C 0 0 5 9 9 1) 未知 (M G C 用蛋白質: 1 4 7 9 3) [H o m o s a p i e n s]。 30 40

【0496】

S G P r 4 0 7, 配列番号 9, 配列番号 6 8 は, 7 5 2 アミノ酸の長さの蛋白質をコードする。これは U C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 4 8 1 からアミノ酸 4 9 1 までユビキチンカルボキシル末端ヒドロラーゼファミリー 2 b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置 8 0 からプロファイル位置 9 0 である。この蛋白質配列を用いたアミノ酸配 50

列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである: Pスコア = 2. 6 0 E - 4 0 ; 同一のアミノ酸の数 = 8 0 ; 同一性のパーセント = 7 6 % ; 類似性のパーセント = 8 4 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 0 3 6 6 0 7 . 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: ユビキチン特異的プロテアーゼ 2 3 ; N E D D 8 - 特異的プロテアーゼ [H o m o s a p i e n s] 。

【0497】

S G P r 4 5 3, 配列番号10, 配列番号69は, 712アミノ酸の長さの蛋白質をコードする。これはU C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸615からアミノ酸677までユビキチンカルボキシル末端ヒドロラーゼファミリー2b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置72である。この蛋白質中で同定された他のドメインは以下のとおりである: U C H 2 b (P F 0 0 4 4 2), 273-304; およびユビキチン-ヒドロラーゼ (P F 0 2 1 4 8) 中のZn-フィンガー, アミノ酸29-99。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである: Pスコア = 0 ; 同一のアミノ酸の数 = 7 1 2 ; 同一性のパーセント = 1 0 0 % ; 類似性のパーセント = 1 0 0 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 1 1 5 5 2 3 . 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: 仮定蛋白質 D K F Z p 4 3 4 D 0 1 2 7 [H o m o s a p i e n s] 。

【0498】

S G P r 4 4 5, 配列番号11, 配列番号70は, 289アミノ酸の長さの蛋白質をコードする。これはU C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸190からアミノ酸221までユビキチンカルボキシル末端ヒドロラーゼファミリー2b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置32である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである: Pスコア = 3. 6 0 E - 1 8 5 ; 同一のアミノ酸の数 = 2 8 9 ; 同一性のパーセント = 1 0 0 % ; 類似性のパーセント = 1 0 0 % ; N R A A 中の最も類似するエントリーの受託番号は A A H 0 5 9 9 1 . 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (B C 0 0 5 9 9 1) 未知 (M G C 用蛋白質: 1 4 7 9 3) [H o m o s a p i e n s] 。

【0499】

S G P r 4 0 1 _ 1, 配列番号12, 配列番号71は, 366アミノ酸の長さの蛋白質をコードする。これはU C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸292からアミノ酸364までユビキチンカルボキシル末端ヒドロラーゼファミリー2b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置72である。この蛋白質中で同定された他のドメインは以下のとおりである: U C H 2 b (P F 0 0 4 4 2), アミノ酸35-66。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである: Pスコア = 7. 3 0 E - 2 5 4 ; 同一のアミノ酸の数 = 3 6 6 ; 同一性のパーセント = 1 0 0 % ; 類似性のパーセント = 1 0 0 % ; N R A A 中の最も類似するエントリーの受託番号は N P 0 7 3 7 4 3 . 1 である; N R A A 中の最も類似する

10

20

30

40

50

蛋白質の名前または説明および種は次のとおりである：仮定蛋白質 F L J 1 2 5 5 2 [H o m o s a p i e n s]。

【0500】

S G P r 4 0 8, 配列番号 1 3, 配列番号 7 2 は, 1 2 8 7 アミノ酸の長さの蛋白質をコードする。これは U C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 3 9 5 からアミノ酸 4 7 5 までユビキチンカルボキシル末端ヒドロラーゼファミリー 2 b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 7 2 である。この蛋白質中で同定された他のドメインは以下のとおりである：U C H 2 b (P F 0 0 4 4 2), アミノ酸 1 0 0 - 1 3 1。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) スミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである：P スコア = 0 ; 同一のアミノ酸の数 = 1 2 8 7 ; 同一性のパーセント = 1 0 0 % ; 類似性のパーセント = 1 0 0 % ; N R A A 中の最も類似するエントリーの受託番号は B A B 5 5 0 6 3 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである：(A K 0 2 7 3 6 2) 命名されていない蛋白質産物 [H o m o s a p i e n s]。

10

【0501】

S G P r 4 8 0, 配列番号 1 4, 配列番号 7 3 は, 1 6 0 4 アミノ酸の長さの蛋白質をコードする。これは U C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 1 5 0 6 からアミノ酸 1 5 6 6 までユビキチンカルボキシル末端ヒドロラーゼファミリー 2 b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 7 2 である。この蛋白質中で同定された他のドメインは以下のとおりである：U C H 2 b (P F 0 0 4 4 2), 7 3 4 - 7 6 5 ; および 2 つの E F ハンド (P F 0 0 0 3 6), 2 3 2 - 2 6 0, および 2 6 8 - 2 9 6。多くのカルシウム結合蛋白質は同じ進化的ファミリーに属し, いずれも E F - ハンドとして知られるタイプのカルシウム結合ドメインを有する (<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00018> を参照)。このタイプのドメインは, 両側を 1 2 残基のアルファヘリックスドメインに挟まれた 1 2 残基のループから構成される。E F - ハンドループにおいては, カルシウムイオンは五角形の二錐体コンフィギュレーションで配位される。この蛋白質は予測 C A A X モチーフ (C V L Q) を有し, これはこの蛋白質を膜画分に向かわせるかもしれない。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである：P スコア = 0 ; 同一のアミノ酸の数 = 1 2 7 2 ; 同一性のパーセント = 9 9 % ; 類似性のパーセント = 9 9 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 1 1 5 9 7 1 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである：ユビキチン特異的プロテアーゼ [H o m o s a p i e n s]。

20

30

40

【0502】

S G P r 4 3 1, 配列番号 1 5, 配列番号 7 4 は, 1 0 4 2 アミノ酸の長さの蛋白質をコードする。これは U C H 2 b ファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 8 3 6 からアミノ酸 9 4 8 までユビキチンカルボキシル末端ヒドロラーゼファミリー 2 b (P F 0 0 4 4 3) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 7 2 である。この蛋白質中で同定された他のドメインは以下のとおりである：U C H 2 b (P F 0 0 4 4 2), 4 4 5 - 4 7 6。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以

50

下のとおりである：Pスコア＝2.40E-251；同一のアミノ酸の数＝397；同一性のパーセント＝100%；類似性のパーセント＝100%；NRAA中の最も類似するエントリーの受託番号はNP_115946.1である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：HP43.8KD蛋白質 [Homo sapiens]。

【0503】

SGPr429, 配列番号16, 配列番号75は, 1033アミノ酸の長さの蛋白質をコードする。これはUCH2bファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸332からアミノ酸419までユビキチンカルボキシル末端ヒドロラーゼファミリー2b (PF00443) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置72である。この蛋白質中で同定された他のドメインは以下のとおりである：UCH2b (PF00442), 89-120。この蛋白質配列を用いたアミノ酸配列の公共データベース (NRAA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝1.50E-250；同一のアミノ酸の数＝368；同一性のパーセント＝100%；類似性のパーセント＝100%；NRAA中の最も類似するエントリーの受託番号はNP_115612.1である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：假定蛋白質FLJ23277 [Homo sapiens]。この蛋白質はアミノ酸87-アミノ酸109に貫膜ドメインを有する。

10

20

【0504】

SGPr503, 配列番号17, 配列番号76は, 517アミノ酸の長さの蛋白質をコードする。これはUCH2bファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸432からアミノ酸501までユビキチンカルボキシル末端ヒドロラーゼファミリー2b (PF00443) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置72である。この蛋白質中で同定された他のドメインは以下のとおりである：UCH2b (PF00442), 68-99。この蛋白質配列を用いたアミノ酸配列の公共データベース (NRAA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝0；同一のアミノ酸の数＝508；同一性のパーセント＝100%；類似性のパーセント＝100%；NRAA中の最も類似するエントリーの受託番号はAAH04868.1である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：(BC004868) 未知 (MGC用蛋白質：10702) [Homo sapiens]。この蛋白質はアミノ酸35-アミノ酸57に貫膜ドメインを有する。

30

【0505】

SGPr427, 配列番号18, 配列番号77は, 1123アミノ酸の長さの蛋白質をコードする。これはUCH2bファミリーのシステインプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸648からアミノ酸709までユビキチンカルボキシル末端ヒドロラーゼファミリー2b (PF00443) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置72である。この蛋白質中で同定された他のドメインは以下のとおりである：UCH2b (PF00442), 101-129。この蛋白質配列を用いたアミノ酸配列の公共データベース (NRAA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝1.80E-92；同一のアミノ酸の数＝269；同一性のパーセント＝36%；類似性のパーセント＝53%；NRAA中の最も類似するエントリーの受託番号はAAF47260.1である；NRAA中の最も類似する蛋白質の名前

40

50

または説明および種は次のとおりである：(A E 0 0 3 4 6 5) C G 3 8 7 2 遺伝子産物 [Drosophila melanogaster]。

【0506】

SGPr092, 配列番号19, 配列番号78は, 261アミノ酸の長さの蛋白質をコードする。これはPepM10ファミリーのメタロプロテアーゼプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸75からアミノ酸194までペプチダーゼ__M10 (PF00413) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置49からプロファイル位置168である。この蛋白質中で同定された他のドメインは以下のとおりである：ADAMドメイン, アミノ酸207-218。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア=4.70E-171; 同一のアミノ酸の数=261; 同一性のパーセント=100%; 類似性のパーセント=100%; NR A A中の最も類似するエントリーの受託番号はX P __011971.1である; NR A A中の最も類似する蛋白質の名前または説明および種は次のとおりである：マトリクスメタロプロテイナーゼ26 [Homo sapiens]。

10

【0507】

SGPr359, 配列番号20, 配列番号79は, 483アミノ酸の長さの蛋白質をコードする。これはPepM10ファミリーのメタロプロテアーゼプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸44からアミノ酸212までペプチダーゼ__M10 (PF00413) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置168である。この蛋白質中で同定された他のドメインは以下のとおりである：3個のヘモペキシン (PF00045) ドメイン, 302-403。ヘモペキシンはヘムに結合しこれを分解および鉄回収のために肝臓に輸送し, その後, 遊離のヘモペキシンが循環中に戻る血清糖蛋白質である。ヘモペキシン様ドメインは2つのタイプの蛋白質において見いだされている：ビトロネクチン (血漿および組織に見いだされる細胞接着および核酸因子) およびマトリクスメタロプロテイナーゼファミリーのほとんどのメンバー (マトリクキシン), 例えば, MMP-1, MMP-2, MMP-3, MMP-8, MMP-9, MMP-10, MMP-11, MMP-12, MMP-13, MMP-14, MMP-15, MMP-16, MMP-17, MMP-18, MMP-19, MMP-20, MMP-24, およびMMP-25 (<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00023>を参照)。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア=0; 同一のアミノ酸の数=483; 同一性のパーセント=100%; 類似性のパーセント=100%; NR A A中の最も類似するエントリーの受託番号はN P __004762.1である; NR A A中の最も類似する蛋白質の名前または説明および種は次のとおりである：マトリクスメタロプロテイナーゼ20プレプロ蛋白質; エナメリシン [Homo sapiens]。この蛋白質はアミノ酸7-アミノ酸29に貫膜ドメインを有する。これはシグナルペプチドとして機能するかもしれない。

20

30

40

【0508】

SGPr104__1, 配列番号21, 配列番号80は, 765アミノ酸の長さの蛋白質をコードする。これはPepM13ファミリーのメタロプロテアーゼプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸561からアミノ酸764までペプチダーゼ__M13 (PF01431) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置222である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12

50

および延長ペナルティー2)の結果は以下のとおりである：Pスコア=0；同一のアミノ酸の数=765；同一性のパーセント=100%；類似性のパーセント=100%；NR AA中の最も類似するエントリーの受託番号はNP_055508.1である；NR AA中の最も類似する蛋白質の名前または説明および種は次のとおりである：KIAA0604 遺伝子産物 [Homo sapiens]。この蛋白質はアミノ酸61ーアミノ酸83に貫膜ドメインを有する。

【0509】

SGPr303, 配列番号22, 配列番号81は, 418アミノ酸の長さの蛋白質をコードする。これはPepM2ファミリーのメタロプロテアーゼプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸10からアミノ酸397までペプチダーゼ_M1 (PF01433) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置416である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR AA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア=2.20E-284；同一のアミノ酸の数=407；同一性のパーセント=97%；類似性のパーセント=98%；NR AA中の最も類似するエントリーの受託番号はCAA10709.1である；NR AA中の最も類似する蛋白質の名前または説明および種は次のとおりである：(AJ132583) ピューロマイシン感受性アミノペプチダーゼ [Homo sapiens]。

10

【0510】

SGPr402_1, 配列番号23, 配列番号82は, 755アミノ酸の長さの蛋白質をコードする。これはズブチラーゼファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸118からアミノ酸437までズブチラーゼ (PF00082) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置360である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR AA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア=0；同一のアミノ酸の数=513；同一性のパーセント=82%；類似性のパーセント=89%；NR AA中の最も類似するエントリーの受託番号はP29121である；NR AA中の最も類似する蛋白質の名前または説明および種は次のとおりである：神経内分泌コンベルターゼ3前駆体 [Mus musculus]。

20

30

【0511】

SGPr434, 配列番号24, 配列番号83は, 391アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸39からアミノ酸46までp20-ICE (PF00656) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置129からプロファイル位置136である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR AA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア=6.20E-43；同一のアミノ酸の数=104；同一性のパーセント=42%；類似性のパーセント=59%；NR AA中の最も類似するエントリーの受託番号はNP_036164.1である；NR AA中の最も類似する蛋白質の名前または説明および種は次のとおりである：貫膜トリプターゼ [Mus musculus]。

40

【0512】

SGPr446_1, 配列番号25, 配列番号84は, 226アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸13からアミノ酸227までトリプシン (PF00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチ

50

するHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置242である。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=2.50E-40; 同一のアミノ酸の数=107; 同一性のパーセント=45%; 類似性のパーセント=57%; NRAA中の最も類似するエントリーの受託番号はNP_038949.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: 遠位腸セリンプロテアーゼ [Mus musculus]。

【0513】

SGPr447, 配列番号26, 配列番号85は, 295アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸33からアミノ酸270までトリプシン(PF00089)についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=1.00E-97; 同一のアミノ酸の数=167; 同一性のパーセント=60%; 類似性のパーセント=77%; NRAA中の最も類似するエントリーの受託番号はBAB30277.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: (AK016509) 予測 [Mus musculus]。

【0514】

SGPr432_1, 配列番号27, 配列番号86は, 628アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸117からアミノ酸343までトリプシン(PF00089)についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置6からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=3.70E-56; 同一のアミノ酸の数=95; 同一性のパーセント=100%; 類似性のパーセント=100%; NRAA中の最も類似するエントリーの受託番号はNP_076869.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: 仮定蛋白質IMAGE3455200 [Homo sapiens]。この蛋白質は2つの貫膜ドメインを有する, アミノ酸10-アミノ酸29, および82-99。アミノ酸10-29の領域はシグナルペプチドとして機能するかもしれない。

【0515】

SGPr529, 配列番号28, 配列番号87は, 276アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸184からアミノ酸187までトリプシン(PF00089)についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置413からプロファイル位置416である。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=1.70E-184; 同一のアミノ酸の数=276; 同一性のパーセント=100%; 類似性のパーセント=100%; NRAA中の最も類似するエントリーの受託番号はNP_002767.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: カリクレイン10; プロテアーゼ, セリン様, 1 [Homo sapiens]。

【0516】

S G P r 4 2 8 _ 1, 配列番号 2 9, 配列番号 8 8 は, 2 8 5 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 2 4 からアミノ酸 2 4 6 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする HMMR プロファイル中の位置はプロファイル位置 7 からプロファイル位置 2 5 9 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 1. 9 0 E - 5 8 ; 同一のアミノ酸の数 = 9 2 ; 同一性のパーセント = 5 3 % ; 類似性のパーセント = 7 3 % ; N R A A 中の最も類似するエントリーの受託番号は B A B 2 4 2 1 5. 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (A K 0 0 5 7 4 0) 予測 [M u s m u s c u l u s]。この蛋白質はアミノ酸 2 6 2 - アミノ酸 2 8 4 に貫膜ドメインを有する。

【0517】

S G P r 4 2 5, 配列番号 3 0, 配列番号 8 9 は, 4 1 3 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 2 8 7 からアミノ酸 3 0 6 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする HMMR プロファイル中の位置はプロファイル位置 3 8 7 からプロファイル位置 4 0 6 である。この蛋白質は予測 C A A X モチーフ (C A Y G) を有し, これはこの蛋白質を原形質膜に向けるかもしれない。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 5. 8 0 E - 2 6 8 ; 同一のアミノ酸の数 = 4 1 2 ; 同一性のパーセント = 9 9 % ; 類似性のパーセント = 9 9 % ; N R A A 中の最も類似するエントリーの受託番号は C A C 3 5 0 7 1. 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (A L 1 2 1 9 3 9) d J 2 2 3 E 3. 1 (予測分泌蛋白質 Z S I G 1 3) [H o m o s a p i e n s]。

【0518】

S G P r 5 4 8, 配列番号 3 1, 配列番号 9 0 は, 3 2 0 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 8 6 からアミノ酸 3 1 3 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする HMMR プロファイル中の位置はプロファイル位置 1 からプロファイル位置 2 5 9 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 2. 6 0 E - 1 6 8 ; 同一のアミノ酸の数 = 2 5 6 ; 同一性のパーセント = 1 0 0 % ; 類似性のパーセント = 1 0 0 % ; N R A A 中の最も類似するエントリーの受託番号は A A G 0 9 4 6 9. 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (A F 2 4 2 1 9 5) K L K 1 5 [H o m o s a p i e n s]。

【0519】

S G P r 3 9 6, 配列番号 3 2, 配列番号 9 1 は, 3 2 8 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 2 8 からアミノ酸 2 6 2 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする HMMR プロファイル中の位置はプロファイル位置 1 からプロファイル位置 2 5 9 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 1. 6 0 E - 5 6 ; 同一のアミノ酸の数 = 1 1 1 ;

同一性のパーセント＝44%；類似性のパーセント＝61%；NR A A中の最も類似するエントリーの受託番号はB A A 8 4 9 4 1. 1である；NR A A中の最も類似する蛋白質の名前または説明および種は次のとおりである：(A B 0 1 8 6 9 4) 表皮特異的セリンプロテアーゼ [*Xenopus laevis*]。

【0520】

S G P r 4 2 6, 配列番号33, 配列番号92は, 425アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸194からアミノ酸419までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝7. 70E－93；同一のアミノ酸の数＝181；同一性のパーセント＝43%；類似性のパーセント＝61%；NR A A中の最も類似するエントリーの受託番号はN P _ 0 5 4 7 7 7. 1である；NR A A中の最も類似する蛋白質の名前または説明および種は次のとおりである：D E S C 1 蛋白質 [*Homo sapiens*]。この蛋白質はアミノ酸30－アミノ酸52に貫膜ドメインを有する。この領域はシグナルペプチドとして機能するかもしれない。

10

【0521】

S G P r 5 5 2, 配列番号34, 配列番号93は, 221アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸2からアミノ酸222までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置255である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝1. 20E－45；同一のアミノ酸の数＝96；同一性のパーセント＝42%；類似性のパーセント＝59%；NR A A中の最も類似するエントリーの受託番号は0 5 4 7 7 7. 1である；NR A A中の最も類似する蛋白質の名前または説明および種は次のとおりである：D E S C 1 蛋白質 [*Homo sapiens*]

20

30

【0522】

S G P r 4 0 5, 配列番号35, 配列番号94は, 948アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸218からアミノ酸406までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置60からプロファイル位置259である。この蛋白質中で同定された他のドメインは以下のとおりである：2つの追加のトリプシンドメイン, アミノ酸419－496, およびアミノ酸636－761。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝1. 10E－30；同一のアミノ酸の数＝111；同一性のパーセント＝54%；類似性のパーセント＝65%；NR A A中の最も類似するエントリーの受託番号はP 1 9 2 3 6である；NR A A中の最も類似する蛋白質の名前または説明および種は次のとおりである：肥満細胞腫プロテアーゼ前駆体 [*Canis familiaris*]。

40

【0523】

S G P r 4 8 5 _ 1, 配列番号36, 配列番号95は, 352アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸68からアミノ酸295までトリプシン (P

50

F00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=7.20E-133; 同一のアミノ酸の数=223; 同一性のパーセント=94%; 類似性のパーセント=96%; NRAA中の最も類似するエントリーの受託番号はBAB03569.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: (AB046651) 仮定蛋白質 [Macaca fascicularis]。

【0524】

SGPr534, 配列番号37, 配列番号96は, 263アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸34からアミノ酸256までトリプシン(PF00089)についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=3.60E-165; 同一のアミノ酸の数=253; 同一性のパーセント=96%; 類似性のパーセント=98%; NRAA中の最も類似するエントリーの受託番号はNP_001897.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: キモトリプシノーゲンB1 [Homo sapiens]。この蛋白質はアミノ酸2-アミノ酸24に貫膜ドメインを有する。この領域はシグナルペプチドとして機能するかもしれない。

【0525】

SGPr390, 配列番号38, 配列番号97は, 1128アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸896からアミノ酸1122までトリプシン(PF00089)についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質中で同定された他のドメインは以下のとおりである: 2つのトリプシンドメイン, アミノ酸264-500, およびアミノ酸573-800。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=2.60E-53; 同一のアミノ酸の数=135; 同一性のパーセント=46%; 類似性のパーセント=59%; NRAA中の最も類似するエントリーの受託番号はBAB23684.1である; NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである: (AK004939) 予測 [Mus musculus]。この蛋白質はアミノ酸28-アミノ酸50に貫膜ドメインを有する。この領域はシグナルペプチドとして機能するかもしれない。

【0526】

SGPr521, 配列番号39, 配列番号98は, 253アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸30からアミノ酸245までトリプシン(PF00089)についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース(NRAA)のスミス・ウォーターマン検索(PAM100, ギャップオープン12および延長ペナルティー2)の結果は以下のとおりである: Pスコア=2.30E-155; 同一のアミノ酸の数=253; 同一性のパーセント=100%; 類似性のパーセント=100%; NRAA中の最も類似するエントリーの受託番号はNP_005037.1である; NRAA中の最も類似す

10

20

30

40

50

る蛋白質の名前または説明および種は次のとおりである：カリクレイン7（キモトリプシン，角質層）；プロテアーゼ，セリン，6（キモトリプシン，角質層）[Homo sapiens]。

【0527】

SGPr530__1，配列番号40，配列番号99は，271アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸14からアミノ酸255までトリプシン（PF00089）についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース（NRAA）のスマス・ウォーターマン検索（PAM100，ギャップオープン12および延長ペナルティー2）の結果は以下のとおりである：Pスコア=1.10E-95；同一のアミノ酸の数=142；同一性のパーセント=100%；類似性のパーセント=100%；NRAA中の最も類似するエントリーの受託番号はCAC12709.1である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：(AL136097) bA62C3.1（精巢セリンプロテアーゼと類似）[Homo sapiens]。

【0528】

SGPr520，配列番号41，配列番号100は，578アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸73からアミノ酸306までトリプシン（PF00089）についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース（NRAA）のスマス・ウォーターマン検索（PAM100，ギャップオープン12および延長ペナルティー2）の結果は以下のとおりである：Pスコア=1.50E-83；同一のアミノ酸の数=158；同一性のパーセント=73%；類似性のパーセント=83%；NRAA中の最も類似するエントリーの受託番号はBAB24587.1である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：(AK006434) 予測 [Mus musculus]。

【0529】

SGPr455，配列番号42，配列番号101は，970アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸433からアミノ酸674までトリプシン（PF00089）についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質中で同定された他のドメインは以下のとおりである：トリプシン，アミノ酸4-156；および3つの3xCUBドメイン（PF00431），アミノ酸175-812。この蛋白質配列を用いたアミノ酸配列の公共データベース（NRAA）のスマス・ウォーターマン検索（PAM100，ギャップオープン12および延長ペナルティー2）の結果は以下のとおりである：Pスコア=5.90E-179；同一のアミノ酸の数=386；同一性のパーセント=41%；類似性のパーセント=58%；NRAA中の最も類似するエントリーの受託番号はT30337である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：ポリ蛋白質-アフリカカギツメカエル。

【0530】

SGPr507__2，配列番号43，配列番号102は，265アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸42からアミノ酸135までトリプシン（PF00089）についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置35からプロファイル位置14

8である。この蛋白質中で同定された他のドメインは以下のとおりである：トリプシンドメイン，アミノ酸247-258。この蛋白質配列を用いたアミノ酸配列の公共データベース（NR AA）のスマス・ウォーターマン検索（PAM100，ギャップオープン12および延長ペナルティー2）の結果は以下のとおりである：Pスコア=2.40E-121；同一のアミノ酸の数=195；同一性のパーセント=73%；類似性のパーセント=81%；NR AA中の最も類似するエントリーの受託番号はNP_080593.1である；NR AA中の最も類似する蛋白質の名前または説明および種は次のとおりである：リケン cDNA 1700016G05 遺伝子 [Mus musculus]。

【0531】

SGPr559，配列番号44，配列番号103は，454アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸217からアミノ酸444までトリプシン（PF00089）についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質中で同定された他のドメインは以下のとおりである：低密度リポ蛋白質レセプタードメインクラスA（PF00057），アミノ酸71-109。LDLレセプターのクラスAドメインはLDLおよびカルシウムの結合部位を形成する。4番目と6番目のシステインの間の酸性残基はLDLRのリガンド中の正に荷電した配列の高親和性結合に重要である。反復はベータヘアピン構造および続く一連のベータターンから構成されることが示されている（<http://www.expasy.ch/cgi-bin/get-prodoc-entry?PDOC00929>を参照）。この蛋白質配列を用いたアミノ酸配列の公共データベース（NR AA）のスマス・ウォーターマン検索（PAM100，ギャップオープン12および延長ペナルティー2）の結果は以下のとおりである：Pスコア=1.40E-288；同一のアミノ酸の数=454；同一性のパーセント=100%；類似性のパーセント=100%；NR AA中の最も類似するエントリーの受託番号はNP_076927.1である；NR AA中の最も類似する蛋白質の名前または説明および種は次のとおりである：貫膜プロテアーゼ，セリン3 [Homo sapiens]。この蛋白質はアミノ酸49-アミノ酸71に貫膜ドメインを有する。

【0532】

SGPr567__1，配列番号45，配列番号104は，537アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸296からアミノ酸524までトリプシン（PF00089）についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース（NR AA）のスマス・ウォーターマン検索（PAM100，ギャップオープン12および延長ペナルティー2）の結果は以下のとおりである：Pスコア=1.70E-135；同一のアミノ酸の数=534；同一性のパーセント=99%；類似性のパーセント=99%；NR AA中の最も類似するエントリーの受託番号はNP_114435.1である；NR AA中の最も類似する蛋白質の名前または説明および種は次のとおりである：モザイクセリンプロテアーゼ [Homo sapiens]。

【0533】

SGPr479__1，配列番号46，配列番号105は，326アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸からアミノ酸288までトリプシン（PF00089）についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置である。この蛋白質配列を用いたアミノ酸配列の公共データベース（NR AA）のスマス・ウォーターマン検索（PAM100，ギャップオープン12および延長ペナルティー2）の結果は以下のとおりである：Pスコア=1.70E-39；同一のアミノ酸の数=107；同一

10

20

30

40

50

性のパーセント＝42%；類似性のパーセント＝57%；NRAA中の最も類似するエントリーの受託番号はNP_114154.1である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：マラプシン [Homo sapiens]。

【0534】

SGPr489__1, 配列番号47, 配列番号106は, 556アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸56からアミノ酸257までトリプシン (PF00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置227である。この蛋白質中で同定された他のドメインは以下のとおりである：2xCUBドメイン (PF00431), アミノ酸304-503。この蛋白質配列を用いたアミノ酸配列の公共データベース (NRAA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝2.70E90；同一のアミノ酸の数＝194；同一性のパーセント＝37%；類似性のパーセント＝54%；NRAA中の最も類似するエントリーの受託番号はT30338である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：オビダクチン [Xenopus laevis]。

10

【0535】

SGPr465__1, 配列番号48, 配列番号107は, 297アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸2からアミノ酸240までトリプシン (PF00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置12からプロファイル位置259である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NRAA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝2.70E-76；同一のアミノ酸の数＝144；同一性のパーセント＝48%；類似性のパーセント＝66%；NRAA中の最も類似するエントリーの受託番号はNP_033381.1である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：精巢セリンプロテアーゼ1 [Mus musculus]。

20

30

【0536】

SGPr524__1, 配列番号49, 配列番号108は, 850アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸613からアミノ酸842までトリプシン (PF00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチするHMMRプロファイル中の位置はプロファイル位置1からプロファイル位置259である。この蛋白質中で同定された他のドメインは以下のとおりである：3つの低密度リポ蛋白質レセプタードメインクラスAドメイン (PF00057), 489-603。LDLレセプターのクラスAドメインはLDLおよびカルシウムの結合部位を形成する。4番目と6番目のシステインの間の酸性残基はLDLRのリガンド中の正に荷電した配列の高親和性結合に重要である。反復はベータヘアピン構造および続く一連のベータターンから構成されることが示されている (<http://www.expasy.ch/cgi-bin/get-prodoc-entry?PDOCC00929>を参照)。この蛋白質配列を用いたアミノ酸配列の公共データベース (NRAA) のスミス・ウォーターマン検索 (PAM100, ギャップオープン12および延長ペナルティー2) の結果は以下のとおりである：Pスコア＝1.30E-79；同一のアミノ酸の数＝193；同一性のパーセント＝41%；類似性のパーセント＝55%；NRAA中の最も類似するエントリーの受託番号はBAB23684.1である；NRAA中の最も類似する蛋白質の名前または説明および種は次のとおりである：(AK004939) 予測 [Mus musc

40

50

u l u s]。この蛋白質はアミノ酸 77-アミノ酸 99 に貫膜ドメインを有する。

【0537】

SGPr 422, 配列番号 50, 配列番号 109 は, 447 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 216 からアミノ酸 441 までトリプシン (PF00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする HMMR プロファイル中の位置はプロファイル位置 1 からプロファイル位置 259 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 12 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 4. 90 E - 80; 同一のアミノ酸の数 = 173; 同一性のパーセント = 39%; 類似性のパーセント = 59%; NR A A 中の最も類似するエントリーの受託番号は NP - 054777. 1 である; NR A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: D E S C 1 蛋白質 [H o m o s a p i e n s]。この蛋白質はアミノ酸 32 からアミノ酸 54 に貫膜ドメインを有する。この領域はシグナルペプチドとして機能するかもしれない。

10

【0538】

SGPr 538, 配列番号 51, 配列番号 110 は, 457 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 218 からアミノ酸 448 までトリプシン (PF00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする HMMR プロファイル中の位置はプロファイル位置 1 からプロファイル位置 259 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 12 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 9. 1 e - 315; 同一のアミノ酸の数 = 457; 同一性のパーセント = 100%; 類似性のパーセント = 100%; NR A A 中の最も類似するエントリーの受託番号は NP _ 110397. 1 である; NR A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: スピネシン [H o m o s a p i e n s]。この蛋白質はアミノ酸 48-アミノ酸 70 に貫膜ドメインを有する。この領域はシグナルペプチドとして機能するかもしれない。

20

【0539】

SGPr 527_1, 配列番号 52, 配列番号 111 は, 818 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 47 からアミノ酸 286 までトリプシン (PF00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする HMMR プロファイル中の位置はプロファイル位置 1 からプロファイル位置 259 である。この蛋白質中で同定された他のドメインは以下のとおりである: 2つの追加のトリプシンドメイン, 323-454, および 564-679。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 12 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 1. 30 E - 52; 同一のアミノ酸の数 = 114; 同一性のパーセント = 42%; 類似性のパーセント = 59%; NR A A 中の最も類似するエントリーの受託番号は A A H 03851. 1 である; NR A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (B C 003851) プロテアーゼ, セリン, 8 (プロスタシン) に類似 [M u s m u s c u l u s]。

30

40

【0540】

SGPr 542, 配列番号 53, 配列番号 112 は, 284 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 35 からアミノ酸 259 までトリプシン (PF00089) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする HMMR プロファイル中の位置はプロファイル位置 1 からプロファイル位置 259 であ

50

る。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 2. 7 0 E - 4 1 ; 同一のアミノ酸の数 = 1 1 0 ; 同一性のパーセント = 4 3 % ; 類似性のパーセント = 5 8 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 0 0 5 3 0 8 . 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: グランザイム M 前駆体; リンパ球 m e t - a s e 1 [H o m o s a p i e n s] 。

【 0 5 4 1 】

S G P r 5 5 1, 配列番号 5 4, 配列番号 1 1 3 は, 8 0 2 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 5 6 8 からアミノ酸 7 9 7 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 2 5 9 である。この蛋白質中で同定された他のドメインは以下のとおりである: 3 つの低密度リポ蛋白質レセプタードメインクラス A ドメイン (P F 0 0 0 5 7), 4 4 7 - 5 5 9。L D L - レセプターのクラス A ドメインは L D L およびカルシウムの結合部位を形成する。4 番目と 6 番目のシステインの間の酸性残基は L D L R のリガンド中の正に荷電した配列の高親和性結合に重要である。反復はベータヘアピン構造および続く一連のベータターンから構成されることが示されている (<http://www.expasy.ch/cgi-bin/get-prodoc-entry?PDOC00929> を参照)。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 0 ; 同一のアミノ酸の数 = 6 7 5 ; 同一性のパーセント = 8 4 % ; 類似性のパーセント = 9 0 % ; N R A A 中の最も類似するエントリーの受託番号は B A B 2 3 6 8 4 . 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: (A K 0 0 4 9 3 9) 予測 [M u s m u s c u l u s]。この蛋白質はアミノ酸 4 4 - アミノ酸 6 6 に貫膜ドメインを有する。この領域はシグナルペプチドとして機能するかもしれない。

【 0 5 4 2 】

S G P r 4 5 1, 配列番号 5 5, 配列番号 1 1 4 は, 3 5 9 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 8 9 からアミノ酸 3 2 4 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 2 5 9 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 9. 9 0 E - 4 1 ; 同一のアミノ酸の数 = 1 0 1 ; 同一性のパーセント = 3 9 % ; 類似性のパーセント = 5 9 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 0 7 2 1 5 2 . 1 である; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである: 副腎分泌性セリンプロテアーゼ前駆体 [R a t t u s n o r v e g i c u s] 。

【 0 5 4 3 】

S G P r 4 5 2 1, 配列番号 5 6, 配列番号 1 1 5 は, 2 8 8 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 7 3 からアミノ酸 2 8 0 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 2 5 9 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (NR A A) のスミス・ウォーターマン検索 (P A M 1 0 0, ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである: P スコア = 1. 4 0 E - 8 1 ; 同一のアミノ酸の数 = 1 4

10

20

30

40

50

2 ; 同一性のパーセント = 57 % ; 類似性のパーセント = 72 % ; N R A A 中の最も類似するエントリーの受託番号は A A K 1 5 2 6 4 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである : (A F 3 0 5 4 2 5) 移植セリンプロテイナーゼ 2 [M u s m u s c u l u s] 。

【0544】

S G P r 5 0 4 , 配列番号 5 7 , 配列番号 1 1 6 は , 4 4 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 1 からアミノ酸 4 5 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 5 2 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0 , ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである : P スコア = 2 . 4 0 E - 1 3 ; 同一のアミノ酸の数 = 2 6 ; 同一性のパーセント = 6 1 % ; 類似性のパーセント = 8 8 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 0 0 2 0 9 5 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである : グランザイム K 前駆体 ; グランザイム 3 ; グランザイム K (セリンプロテアーゼ , グランザイム 3) ; トリプターゼ I I [H o m o s a p i e n s] 。

10

【0545】

S G P r 4 6 9 , 配列番号 5 8 , 配列番号 1 1 7 は , 4 5 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 1 からアミノ酸 4 6 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 2 1 0 からプロファイル位置 2 5 9 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0 , ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである : P スコア = 2 . 2 0 E - 1 7 ; 同一のアミノ酸の数 = 3 2 ; 同一性のパーセント = 6 9 % ; 類似性のパーセント = 8 4 % ; N R A A 中の最も類似するエントリーの受託番号は B A B 3 0 2 7 7 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである : (A K 0 1 6 5 0 9) 予測 [M u s m u s c u l u s] 。

20

30

【0546】

S G P r 4 0 0 , 配列番号 5 9 , 配列番号 1 1 8 は , 3 0 9 アミノ酸の長さの蛋白質をコードする。これはトリプシンファミリーのセリンプロテアーゼとして分類される。この蛋白質中のプロテアーゼドメインはアミノ酸 1 3 3 からアミノ酸 2 8 1 までトリプシン (P F 0 0 0 8 9) についての隠れマルコフプロファイルとマッチする。蛋白質配列とマッチする H M M R プロファイル中の位置はプロファイル位置 1 からプロファイル位置 1 9 8 である。この蛋白質配列を用いたアミノ酸配列の公共データベース (N R A A) のスミス・ウォーターマン検索 (P A M 1 0 0 , ギャップオープン 1 2 および延長ペナルティー 2) の結果は以下のとおりである : P スコア = 2 . 3 0 E - 1 6 ; 同一のアミノ酸の数 = 7 2 ; 同一性のパーセント = 3 8 % ; 類似性のパーセント = 4 8 % ; N R A A 中の最も類似するエントリーの受託番号は N P _ 0 3 6 1 6 4 . 1 である ; N R A A 中の最も類似する蛋白質の名前または説明および種は次のとおりである : 貫膜トリプターゼ [M u s m u s c u l u s] 。

40

【0547】

実施例 2 : 哺乳動物プロテアーゼの発現分析

材料および方法

定量的 P C R 分析

種々の正常ヒト組織および細胞株から R N A を単離する。一本鎖 c D N A は , 1 0 μ g の各 R N A から S u p e r s c r i p t P r e a m p l i f i c a t i o n S y s t e m

50

(GibcoBRL)を用いて上述のようにして合成する。次に、これらの一本鎖テンプレートを、Light Cycler (Roche Molecular Biochemical) 中で実時間PCR反応で、特異的プライマー対を用いて直線的に増幅する。グラフの読みは、全RNA調製物中の標的とする遺伝子の相対的量の定量的分析を提供する。

【0548】

DNAアレイに基づく発現分析

種々の正常ヒト組織、凍結切片および細胞株からDNAを含まないRNAを単離する。SMARTPCR cDNA合成手法 (Clontech) を改変して用いて、10 μg のRNAまたは1 μg のmRNAから一本鎖cDNAを合成する。ユニークなオリゴヌクレオチド配列を有する各転写産物の5'末端と3'末端を非対称的に標識するために、この方法を改変することができる。次に、Advantage ロングレンジPCR (Clontech) を用いてLight Cycler PCR装置で得られたsscDNAを直線的に増幅する。グラフの実時間ディスプレイが生成物がプラトーに達し始めたことを示したときに反応を停止させる。Millipore DNA精製マトリクスを用いて二本鎖cDNA生成物を精製し、乾燥し、再懸濁し、定量し、アガロースゲルで分析する。得られる要素を"組織cDNA"と称する。

10

【0549】

Genetic Microsystems (GMS) アレイ作製器を500 ng/μl で用いて、GAPS被覆ガラススライド (Corning) 上に組織cDNAをスポットする。

20

【0550】

各新規エクソンに対して蛍光標識オリゴヌクレオチドを合成して、これらが最も近い既知のホモログとの内部ミスマッチを含むことを確実にする。典型的にはオリゴは45ヌクレオチドの長さであり、5'末端でCy5で標識する。

【0551】

エクソン特異的Cy5標識オリゴをガラススライド上にアレイされた組織cDNAにハイブリダイズさせ、標準的緩衝液および条件を用いて洗浄する。次にハイブリダイズしたシグナルを、GMS Scannerを用いて定量する。

【0552】

あるいは、384ピンレプリケータを用いて組織cDNAを手動でナイロン膜上にスポットし、³²P末端標識オリゴプローブにハイブリダイズさせる。

30

【0553】

目的とする疾病領域に関連性のある情報を提供するように、および各プロテアーゼの作用の生物学的メカニズムを反映するように選択された多数のRNAテンプレートから組織cDNAを生成する。これらのテンプレートには、以下のものが含まれる：癌関連遺伝子を同定するためにヒト腫瘍細胞株、原発性ヒト腫瘍の凍結切片および32の正常ヒト組織；CNS関連遺伝子については正常、アルツハイマー病、パーキンソン病、および精神分裂病領域の切片；代謝関連遺伝子については正常および糖尿病または肥満骨格筋、脂肪細胞、または肝臓；および精製された造血細胞、および免疫関連遺伝子についてはリンパ様組織。作用の遺伝子メカニズムを特性決定するためには、新脈管形成 (VEGFリガンド、抗脈管形成剤、または低酸素で処理した培養内皮細胞)、運動性 (HGFリガンドで刺激したA549細胞、正常位転移、原発性腫瘍およびこれに対応する転移性腫瘍)、細胞サイクル (薬剤ブロックにより同調させ、細胞サイクルの種々の時間に回収したHela、H1299、および他の細胞株)、チェックポイントの完全性およびDNA修復 (γ線照射、UV、シスプラチナまたは酸化的ストレスで処理したp53正常および欠損細胞)、および細胞生存 (分化するよう誘導した細胞またはアポトーシスの種々の段階) を反映するように組織cDNAを生成する。

40

【0554】

実施例3：哺乳動物プロテアーゼをコードするcDNAの単離

50

材料および方法新規クローンの同定

総RNAは、原発性ヒト腫瘍、正常および腫瘍細胞株、正常ヒト組織、およびソートしたヒト造血細胞細胞からChomczynskiおよびSacchi (P. Chomczynski and N. Sacchi, Anal. Biochem. 162, 156 (1987)) のグアニジン塩／フェノール抽出プロトコルを用いて単離する。これらのRNAを用いて、Superscript Pre Amplification System (GIBCO BRL, Gaithersburg, MD; Gerard, G. F. et al. (1989), FOCUS 11, 66) を用いて、製造元により推奨される条件で一本鎖cDNAを生成する。典型的な反応においては、60 μ l の反応容量中で10 μ g の総RNAおよび1.5 μ g のオリゴ(dT)₁₂₋₁₈を用いる。生成物をRNase Hで処理し、H₂Oで100 μ l に希釈する。続くPCR増幅のためには、1-4 μ l のこのsscDNAを各反応において用いる。

【0555】

縮重オリゴヌクレオチドは、Applied Biosystems 3948 DNA合成機で確立されたホスホルアミダイト化学を用いて合成し、エタノールで沈澱させ、精製せずにPCR用に用いる。これらのプライマーは、いくつかの蛋白質プロテアーゼの触媒ドメイン中の保存モチーフのセンスおよびアンチセンス鎖に由来する。縮重ヌクレオチド残基の表示は以下のとおりである：N=A, C, G, またはT；R=AまたはG；Y=CまたはT；H=A, CまたはT, Gではない；D=A, GまたはT, Cではない；S=CまたはG；およびW=AまたはT。

【0556】

PCR反応は、多数の一本鎖cDNAに適用される縮重プライマーを用いて行う。プライマーをそれぞれ最終濃度5 μ Mで、10mM TrisHCl, pH 8.3, 50mM KCl, 1.5mM MgCl₂, 各200 μ Mデオキシヌクレオシド三リン酸, 0.001%ゼラチン, 1.5U AmpliTaq DNAポリメラーゼ (Perkin-Elmer/Cetus), および1-4 μ g のcDNAを含む混合物に加える。95℃で3分間変性させた後、サイクルの条件は、94℃で30秒間, 50℃で1分間, および72℃で1分45秒間を35サイクルである。300-350bpの間に移動したPCRフラグメントをGeneCleanキット (Bio101) を用いて2%アガロースゲルから単離し、製造元のプロトコルにしたがってpCRIIベクター (Invitrogen Corp. U. S. A.) 中にT-Aクローニングする。

【0557】

Qiagenカラムを用いるミニプラスミドDNA調製用にコロニーを選択し、サイクルシーケンシング染料ターミネータキットおよびAmpliTaq DNAポリメラーゼ, FS (ABI, Foster City, CA) を用いてプラスミドDNAを配列決定する。配列決定反応生成物はABI Prism 377 DNAシーケンサーにかけ、BLASTアラインメントアルゴリズム (Altschul, S. F. et al., J. Mol. Biol. 215:403-10) を用いて分析する。

【0558】

追加のPCR戦略を用いて、正確なまたはほぼ正確なオリゴヌクレオチドプライマーを用いて種々のPCRフラグメントまたはESTを接続する。PCR条件は上述したとおりであるが、ただし、アニーリング温度は、式： $T_m = 4(G + C) + 2(A + T)$ を用いて各オリゴ対について計算する。

【0559】cDNAクローンの単離：

ヒトcDNAライブラリをプロテアーゼ関連遺伝子に対応するPCRまたはESTフラグメントで探索する。プローブをランダムプライミングにより³²Pで標識し、ライブラリスクリーニングの標準的手法にしたがって、2 x 10⁶ cpm/mLで用いる。プレハイブリダイゼーション (3時間) およびハイブリダイゼーション (一夜) は、5 X S S C,

5 X デンハルト溶液, 2. 5 % 硫酸デキストラン, 50 mM $\text{Na}_2\text{PO}_4 / \text{NaHPO}_4$, pH 7. 0, 50 % ホルムアミド, 100 mg / mL 変性サケ精子 DNA 中で 42 °C で行う。ストリンジェントな洗浄は, 65 °C で, 0. 1 X S S C および 0. 1 % S D S 中で行う。DNA 配列決定は, サイクルシーケンシング染料ターミネータキットを用い, A m p l i T a q DNA ポリメラーゼ F S (A B I, F o s t e r C i t y, C A) を用いて両方の鎖について行う。配列決定反応生成物を A B I P r i s m 3 7 7 DNA シーケンサーにかける。

【0560】

実施例 4 : 哺乳動物プロテアーゼの発現分析

材料および方法

ノザンプロット分析

ノザンプロットは, ヒト成人組織 (例えば, 胸腺, 肺, 十二指腸, 結腸, 精巣, 脳, 小脳, 皮質, 唾液腺, 肝臓, 膵臓, 腎臓, 脾臓, 胃, 子宮, 前立腺, 骨格筋, 胎盤, 乳腺, 膀胱, リンパ節, 脂肪組織), および 2 つのヒト胎児正常組織 (胎児肝臓, 胎児脳) からの 60 種類のヒト腫瘍細胞株 (例えば, H O P - 9 2, E K V X, N C I - H 2 3, N C I - H 2 2 6, N C I - H 3 2 2 M, N C I - H 4 6 0, N C I - H 5 2 2, A 5 4 9, H O P - 6 2, O V C A R - 3, O V C A R - 4, O V C A R - 5, O V C A R - 8, I G R O V 1, S K - O V - 3, S N B - 1 9, S N B - 7 5, U 2 5 1, S F - 2 6 8, S F - 2 9 5, S F - 5 3 9, C C R F - C E M, K - 5 6 2, M O L T - 4, H L - 6 0, R P M I 8 2 2 6, S R, D U - 1 4 5, P C - 3, H T - 2 9, H C C - 2 9 9 8, H C T - 1 1 6, S W 6 2 0, C o l o 2 0 5, H T C 1 5, K M - 1 2, U O - 3 1, S N 1 2 C, A 4 9 8, C a K i l, R X F - 3 9 3, A C H N, 7 8 6 - 0, T K - 1 0, L O X I M V I, M a l m e - 3 M, S K - M E L - 2, S K - M E L - 5, S K - M E L - 2 8, U A C C - 6 2, U A C C - 2 5 7, M 1 4, M C F - 7, M C F - 7 / A D R R E S, H s 5 7 8 T, M D A - M B - 2 3 1, M D A - M B - 4 3 5, M D A - N, B T - 5 4 9, T 4 7 D) から単離した 10 μ g の総 RNA を変性ホルムアルデヒド 1. 2 % アガロースゲルに流し, ナイロン膜に移すことにより調製する。

【0561】

フィルターを, プロテアーゼ遺伝子のいくつかの挿入物から合成したランダムプライミング [$\alpha^{32}\text{P}$] d C T P 標識プローブでハイブリダイズさせる。ハイブリダイゼーションは, 6 X S S C, 0. 1 % S D S, 1 X デンハルト溶液, 100 μ g / mL の変性ニシン精子 DNA 中で, 1 - 2 x 10⁶ c p m / mL の ^{32}P 標識 DNA プローブで 42 °C で一夜行う。フィルターを 0. 1 X S S C / 0. 1 % S D S 中で 65 °C で洗浄し, M o l e c u l a r D y n a m i c s ホスファージメジャーで露光する。

【0562】

定量的 P C R 分析

種々の正常ヒト組織および細胞株から RNA を単離する。一本鎖 c D N A は, 10 μ g の上述した各 RNA から S u p e r s c r i p t P r e a m p l i f i c a t i o n S y s t e m (G i b c o B R L) を用いて合成する。次に, これらの一本鎖テンプレートを各クローンに特異的なプライマーとともに 25 サイクルの P C R 反応において用いる。反応生成物を 2 % アガロースゲルで電気泳動し, エチジウムブロマイドにより染色し, U V 光ボックスで写真を撮る。各試料について S T K 特異的バンドの相対的強度を評価する。

【0563】

DNA アレイに基づく発現分析

プラスミド DNA アレイプロットは, 各プロテアーゼについて 0. 5 μ g の変性プラスミドをナイロン膜に付加することにより調製する。 [$\gamma^{32}\text{P}$] d C T P 標識一本鎖 DNA プローブは, いくつかのヒト免疫組織起源または腫瘍細胞 (例えば, 胸腺, 樹状細胞, 肥満細胞, 単球, B 細胞 (初代, J u r k a t, R P M I 8 2 2 6, S R), T 細胞 (C D 8 / C D 4 +, T H 1, T H 2, C E M, M O L T 4), K 5 6 2 (巨核球) から単離さ

10

20

30

40

50

れた総RNAから合成する。ハイブリダイゼーションは、42℃で16時間、6XSSC、0.1%SDS、1Xデンハルト溶液、100μg/mL変性ニシン精子DNA中で、 10^6 cpm/mLの $[\gamma^{32}\text{P}]$ dCTP標識一本鎖プローブを用いて行う。フィルターを0.1XSSC/0.1%SDSで65℃で洗浄し、Molecular Dynamicsホスファイメージャーで定量的分析を行う。

【0564】

実施例5：プロテアーゼ遺伝子発現

ベクターの構築

材料および方法

発現ベクターの構築

いくつかのヒトcDNAから発現構築物を作成する：a) pCDNA発現ベクター中の全長クローン；b) GST発現カセットのC末端に融合させた新規プロテアーゼの触媒ドメインを含むGST融合構築物；およびc) pCDNAベクター中に挿入されたプロテアーゼドメイン中の予測ポリペプチド切断部位に変異を含む全長クローン。

【0565】

プロテアーゼのこれらの変異体は優性阻害構築物として機能するかもしれず、これらの新規プロテアーゼの機能を解明するために用いられる。

【0566】

実施例6：プロテアーゼに対する特異的免疫試薬の作成

材料および方法

単離されたプロテアーゼポリペプチドに対応するKLHまたはMAP-コンジュゲート化合成ペプチドに対する特異的免疫試薬をウサギで生成させる。C末端ペプチドをグルタルアルデヒドでKLHとコンジュゲートさせ、遊離C末端を残した。内部ペプチドは、ブロックされたN末端を用いてMAP-コンジュゲートさせた。追加の免疫試薬はまた、細菌で発現させた各新規PTKまたはSTKの細胞質ドメインを含むGST融合蛋白質でウサギを免疫することにより生成することができる。内因性起源について試験する前に、最初に、種々の免疫血清を、組換え蛋白質に対する反応性および選択性について試験する。

【0567】

ウエスタンブロット

SDS PAGEの蛋白質をイモビロン膜に移す。洗浄緩衝液はPBST（標準的リン酸緩衝化食塩水、pH7.4+0.1%TritonX-100）である。ブロッキングおよび抗体インキュベーション緩衝液はPBST+5%ミルクである。抗体希釈は1:1000から1:2000の範囲である。

【0568】

実施例7：プロテアーゼの組換え発現および生物学的アッセイ

材料および方法

哺乳動物細胞におけるプロテアーゼの過渡的発現

プロテアーゼ構築物を含むpcDNA発現プラスミド（10μgDNA/100mmプレート）をリポフェクタミン（Gibco BRL）とともに293細胞中に導入する。72時間後、細胞を0.5mLの可溶化緩衝液（20mMHEPES、pH7.35、150mMNaCl、10%グリセロール、1%TritonX-100、1.5mMMgCl₂、1mMEGTA、2mMフッ化フェニルメチルスルホニル、1μg/mLアプロチニン）中に回収する。試料アリコートに6%アクリルアミド/0.5%ビスアクリルアミドゲルのSDSポリアクリルアミドゲル電気泳動（PAGE）で分離し、電気泳動的にニトロセルロースに移す。非特異的結合は、ブロットをBlotto（5%w/v無脂乾燥ミルクおよび0.2%v/vNonidet P-40（Sigma）を含有するリン酸緩衝化食塩水）中でプレインキュベートすることによりブロックし、種々の抗ペプチドまたは抗GST融合蛋白質特異的抗血清を用いて組換え蛋白質を検出する。

【0569】

インビトロプロテアーゼアッセイ

蛍光発生ペプチドを用いるインビトロプロテアーゼアッセイ

アッセイは蛍光分光計、例えばPerkin-Elmer 204 Sを用いて行う。標準的
反応混合物（100 μ l）は、200 mM Tris-HCl, pH 8.5, および200
 μ M 蛍光発生ペプチド基質を含む。酵素を加えた後、反応混合物を37 $^{\circ}$ Cで30分間イン
キュベートし、1.9 mlの125 mM ZnSO₄を加えることにより停止する（Bre
nner, C., and Fuller, R. S., 1992, Proc. Natl. A
cad. Sci. U. S. A. 89: 922-926）。マイクロ遠心分離器で1分間遠心
分離することにより沈殿物を除去し（15,000 $\times$ g）、上清に放出される生成物（7
-アミノ-4-メチルクマリン）の速度を蛍光計で判定する〔（励起）= 385 nm,
（放出）= 465 nm〕。文献で用いられている基質の例には次のものがある：Boc-
Gly-Arg-Arg-4-メチルクマリル-7-アミド（MCA）、Boc-Gln
-Arg-Arg-MCA, Z-Arg-Arg-MCA, およびpGlu-Arg-Thr-Lys-Arg-MCA。保存溶液（100 mM）は、ペプチドをジメチルスルホ
キシドに溶解することにより調製し、次にこれを使用前に水で1 mMの作業溶液に希釈す
る（このアッセイの詳細は、R. Yosuf, et al. R Biol. Chem.,
Vol. 275, Issue 14, 9963-9969, April 7, 2000（
図面および表を含め本明細書の一部としてここに引用する））に見いだされる）。

10

【0570】

蛍光発生ペプチドを用いる無傷の細胞におけるプロテアーゼアッセイ

カルパイン活性は、細胞内チオールコンジュゲート化Boc-Leu-Met-CMAC
からの蛍光生成物であるAMCの生成速度により測定する（Rosser, B. G., P
owers, S. P., and Gores, G. J. (1993) J. Biol. Ch
em. 268, 23593-23600）。細胞を分散させ、ガラスカバースリップ上で
成長させ、生理食塩水溶液を37 $^{\circ}$ Cで連続的に注ぎ、順番に定量的蛍光イメージングシス
テムで画像化する。t = 0において、Boc-Leu-Met-CMAC（10 μ M, M
olecular Probes）を注入溶液に導入し、個々の細胞の平均蛍光強度（励
起350 nm, 放出470 nm）を60秒間隔で測定する。10分において、TNF-
（30 ng/ml）を10 μ MのBoc-Leu-Met-CMACを含む注入溶液に加える。
時間に対する蛍光の変化の傾きは細胞内カルパイン活性を表す（Rosser, et
al., 1993, J. Biol. Chem. 268: 23593-23600）。全
細胞集団におけるカルパインアッセイのためには、細胞の懸濁培養物に10 μ MのBoc
-Leu-Met-CMACを負荷し、TNF α を添加する前および後で細胞内蛍光の変
化をFACS Vantageシステムを用いて37 $^{\circ}$ Cで測定する。AMCの細胞蛍光は
360-nm励起フィルターおよび405-nm長経路放出フィルターを用いて測定する
（このアッセイの詳細は、Han, et al., 1999, J Biol Chem,
274: 787-794（図面および表を含めその全体を本明細書の一部としてここに引
用する））に見いだすことができる）。

20

30

【0571】

発色基質を用いるプロテアーゼアッセイ

酵素の蛋白質分解活性は、市販のアッセイシステム（Athena Environme
ntal Sciences, Inc.）を用いて測定する。アッセイには、マトリクス
に架橋させた染料-蛋白質コンジュゲートのユニバーサル基質を用いる。プロテアーゼ活
性は、マトリクスから上清に放出される染料の吸光度を測定することにより分光高度計で
測定する。酵素および基質を含む反応バイアルを37 $^{\circ}$ Cで3時間インキュベートする。異
なるインキュベーション時間で活性を測定し、500 μ lの0.2 N NaOHを各バイアル
に加えることにより反応を停止させる。各反応バイアルの上清の吸光度を450 nmで
測定する。蛋白質分解活性は、10 μ l（約10 μ g）の精製蛋白質を用いて、50 mM
Tris-HCl（pH 7.5）中の5 μ gのカゼイン（Sigma）とともに、37 $^{\circ}$ C
で30分間、1時間または2時間インキュベートしてモニターする。反応生成物をSDS
-ポリアクリルアミドゲル電気泳動により分離し、クマシーブルーで染色することにより

40

50

蛋白質を可視化する（このアッセイの詳細は，F a c c i o , e t a l . , 2 0 0 0 , J B i o l C h e m , 2 7 5 : 2 5 8 1 - 2 5 8 8 （図面および表を含めその全体を本明細書の一部としてここに引用する）に見いだすことができる）。

【0572】

膜に結合した放射性標識基質を用いるプロテアーゼアッセイ

標識していないプロテアーゼを緩衝液（100mM HEPES, 100mM NaCl, 125μM酢酸マグネシウム, 125μM酢酸亜鉛, pH7.5）中で放射性標識基質を含有する膜と混合し, 30℃でインキュベートする。典型的には, 各反応は最終容量80-100μlである。各反応は, 同じ最終濃度の溶解緩衝液成分（25mM Tris, 0.1Mソルビトール, 0.5mM EDTA, 0.01%NaN₃, pH7.5）に 10 対して標準化する。これは, 各反応に加える膜の量が様々であるためである。金属イオン特異性を調べるためには, 基質なしで反応を組立て, 1.125mMの1,10-オルトフェナントロリンで氷上で20分間予備処理する。次に, 金属イオンおよび基質含有膜を加え, 30℃でインキュベートすることにより反応を開始させる。これらを加えることにより, 1,10-オルトフェナントロリンは最終濃度1mMに希釈される。金属イオンを酢酸塩の形で, まず2mM濃HClで酸性にし, 次に1mMHEPES, pH7.5で中和した25-100mMの保存溶液から加える（Zn²⁺, Mg²⁺, Cu²⁺, Co²⁺, またはCa²⁺）。この工程は酢酸亜鉛を完全に溶解するために必要である。免疫沈 20 澱による分析のためには, 試料を0.1%SDSを含む免疫沈澱緩衝液で10-20倍に希釈し（Berkower, C., and Michaelis, S. (1991) EM BO J. 10:3777-3785）, 不溶性物質を除去し（13,000xg, 5-10分間, 4℃）, 基質特異的抗体で免疫沈澱させる。あるいは, 試料をSDS（最終濃度0.5%）により可溶化し, 3分間煮沸し, 免疫沈澱緩衝液で希釈した後, 直接免疫沈澱させる。免疫沈澱物を記載されるようにSDSポリアクリルアミドゲル電気泳動に供し, 20%トリクロロ酢酸で7分間固定し, 乾燥し, ホスファーイメージャースクリーンに露光して検出および定量を行う（Molecular Dynamics, Sunnyvale, CA）。上述の試薬はすべてSigmaから購入することができる。（このアッセイの詳細は, Schmidt, et al., 2000, J Biol Chem, 275:6227-6233（図面および表を含めその全体を本明細書の一部としてここに引用する）に見いだすことができる）。このアッセイを変形して膜に結合させていない基 30 質に簡単に適用することができる。

【0573】

種々のプロテアーゼアッセイの包括的議論は, The Handbook of Proteolytic Enzymes, Alan J. Barrett (Editor), Neil D. Rawlings (Editor), J. Fred Woessner (Editor) (February 1998) Academic Press, San Diego; ISBN: 0-12-079370-9（図面および表を含めその全体を本明細書の一部としてここに引用する）に見いだすことができる。

【0574】

細菌で発現させたプロテアーゼのGST融合構築物について類似のアッセイを行う。 40

【0575】

実施例8a:プロテアーゼの染色体位置

材料および方法

いくつかの情報源を用いて, 本明細書に記載される遺伝子のそれぞれの染色体上の位置に関する情報を得た。最初に, C e l e r a B r o w s e rを用いて遺伝子をマッピングした。あるいは, ゲノムコンティグ（NRNAに対するBLASTにより同定）の受託番号を用いて, E n t r e z G e n o m e B r o w s e r (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/MapViewHelp.html)をクエリーし, NCBIデータから細胞遺伝学的位置を読んだ。マップされた部位とヒト癌において見いだされる染色体増幅との関連についての参考文献 50

は, Knuutila, et al., Am J Pathol, 1998, 152: 1107-1123に見いだすことができる。マップされた位置についての情報はまた, 公表された文献 (NCBI, <http://www.ncbi.nlm.nih.gov/enke/querf.cgi>) を, マップされた位置とヒト疾病との関連性について検索することにより入手した。

【0576】

結果

マップされた遺伝子の染色体領域は表2に示される。

【0577】

以下の節は, 本出願に含まれるプロテアーゼについて確立された染色体位置にマップされる種々の疾病を記載する。本発明のプロテアーゼポリヌクレオチドを用いて, 関連する疾病を有するかまたはそれを発達させるリスクを有する個人を同定することができる。本明細書において説明されるように, 本発明のポリペプチドおよびポリヌクレオチドは, プロテアーゼ活性を調節し, 次に種々の疾病を改善する化合物を同定するのに有用である。

10

【0578】

SGPr 397, 配列番号1はヒト染色体位置8q12にマップされている。この領域中の染色体異常は乳癌と関連づけられている (Rummukainen J, et al., Cancer Genet Cytogenet. 2001 Apr; 126 (1): 1-7)。

【0579】

SGPr 413, 配列番号2はヒト染色体位置2q35にマップされている。この領域は変形性関節症における関与が強く示唆されている (Loughlin J, et al., Linkage analysis of chromosome 2q in osteoarthritis. Rheumatology. 2000 Apr; 39 (4): 377-81)。

20

【0580】

SGPr 404, 配列番号3はヒト染色体位置10q26にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuutila): 軟組織の悪性線維性組織球腫。

【0581】

SGPr 536 __ 1, 配列番号4はヒト染色体位置1p35にマップされている。

30

【0582】

SGPr 414, 配列番号5はヒト染色体位置2p14にマップされている。

【0583】

SGPr 430, 配列番号6はヒト染色体位置2q37にマップされている。この領域は変形性関節症における関与が強く示唆されている (Loughlin J, et al., Linkage analysis of chromosome 2q in osteoarthritis. Rheumatology. 2000 Apr; 39 (4): 377-81)。

【0584】

SGPr 496 __ 1, 配列番号7はヒト染色体位置Xp11.4にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuutila): 小細胞肺癌および前立腺癌。

40

【0585】

SGPr 495, 配列番号8はヒト染色体位置6q16にマップされている。

【0586】

SGPr 407, 配列番号9はヒト染色体位置2q37にマップされている。この領域は変形性関節症における関与が強く示唆されている (Loughlin J, et al., Linkage analysis of chromosome 2q in osteoarthritis. Rheumatology. 2000 Apr; 39 (4)

50

: 377-81)。

【0587】

SGPr453, 配列番号10はヒト染色体位置12q23にマップされている。

【0588】

SGPr445, 配列番号11はヒト染色体位置6q16にマップされている。

【0589】

SGPr401__1, 配列番号12はヒト染色体位置4q11にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 濾胞状癌。

【0590】

SGPr408, 配列番号13はヒト染色体位置11p15にマップされている。

10

【0591】

SGPr480, 配列番号14はヒト染色体位置17q24にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 非小細胞肺癌および精巣癌。

【0592】

SGPr431, 配列番号15はヒト染色体位置4q31.3にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 骨肉腫

【0593】

SGPr429, 配列番号16はヒト染色体位置1p36.2にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 肺癌。

20

【0594】

SGPr503, 配列番号17はヒト染色体位置12q24.3にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 非小細胞肺癌。

【0595】

SGPr427, 配列番号18はヒト染色体位置17p13にマップされている。

【0596】

SGPr092, 配列番号19はヒト染色体位置11p15にマップされている。

【0597】

SGPr359, 配列番号20はヒト染色体位置11q22にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 子宮頸癌。

30

【0598】

SGPr104__1, 配列番号21はヒト染色体位置3q27にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 頭頸部扁平上皮癌; 軟組織の悪性線維性組織球腫。

【0599】

SGPr303, 配列番号22はヒト染色体位置17q11.1にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 乳癌腫および肝細胞癌腫。

【0600】

SGPr402__1, 配列番号23はヒト染色体位置19q11にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 平滑筋肉腫。

40

【0601】

SGPr434, 配列番号24はヒト染色体位置3p21にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 膀胱癌腫。

【0602】

SGPr446__1, 配列番号25はヒト染色体位置3p21にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている(Knuutila): 膀胱癌腫。

【0603】

50

S G P r 4 4 7, 配列番号 2 6 はヒト染色体位置 1 6 p 1 3. 3 にマップされている。

【0 6 0 4】

S G P r 4 3 2 _ 1, 配列番号 2 7 は染色体位置が決定されていない。

【0 6 0 5】

S G P r 5 2 9, 配列番号 2 8 はヒト染色体位置 1 9 q 1 3. 4 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 乳癌腫。

【0 6 0 6】

S G P r 4 2 8 _ 1, 配列番号 2 9 はヒト染色体位置 8 p 2 3 にマップされている。

【0 6 0 7】

S G P r 4 2 5, 配列番号 3 0 はヒト染色体位置 6 q 1 4 にマップされている。

10

【0 6 0 8】

S G P r 5 4 8, 配列番号 3 1 はヒト染色体位置 1 9 q 1 3. 4 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 乳癌腫。

【0 6 0 9】

S G P r 3 9 6, 配列番号 3 2 はヒト染色体位置 4 q 3 2 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 非小細胞肺癌。

【0 6 1 0】

S G P r 4 2 6, 配列番号 3 3 はヒト染色体位置 4 q 1 3 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 非小細胞肺癌。

【0 6 1 1】

20

S G P r 5 5 2, 配列番号 3 4 はヒト染色体位置 4 q 1 3 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 非小細胞肺癌。

【0 6 1 2】

S G P r 4 0 5, 配列番号 3 5 はヒト染色体位置 1 6 p 1 3. 3 にマップされている。

【0 6 1 3】

S G P r 4 8 5 _ 1, 配列番号 3 6 はヒト染色体位置 8 p 2 3 にマップされている。

【0 6 1 4】

S G P r 5 3 4, 配列番号 3 7 はヒト染色体位置 1 6 q 2 3 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 胃の散在性大細胞リンパ種。

30

【0 6 1 5】

S G P r 3 9 0, 配列番号 3 8 はヒト染色体位置 1 9 q 1 1 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 平滑筋肉腫。

【0 6 1 6】

S G P r 5 2 1, 配列番号 3 9 はヒト染色体位置 1 9 q 1 3. 4 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 乳癌腫。

【0 6 1 7】

S G P r 5 3 0 _ 1, 配列番号 4 0 はヒト染色体位置 9 q 2 2 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 非小細胞肺癌。

40

【0 6 1 8】

S G P r 5 2 0, 配列番号 4 1 はヒト染色体位置 2 q 3 7 にマップされている。この領域は変形性関節症における関与が強く示唆されている (L o u g h l i n J, e t a l., L i n k a g e a n a l y s i s o f c h r o m o s o m e 2 q i n o s t e o a r t h r i t i s. R h e u m a t o l o g y. 2 0 0 0 A p r ; 3 9 (4) : 3 7 7 - 8 1)。

【0 6 1 9】

S G P r 4 5 5, 配列番号 4 2 はヒト染色体位置 1 2 p 1 1. 2 にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (K n u u t i l a) : 卵巣生殖細胞腫瘍, 精巣癌および非小細胞肺癌。

50

【0620】

SGPr507__2, 配列番号43はヒト染色体位置7q36にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuttilla) : 卵巣癌。

【0621】

SGPr559, 配列番号44はヒト染色体位置21q22にマップされている。

【0622】

SGPr567__1, 配列番号45はヒト染色体位置11q23にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuttilla) : 胸膜中皮腫。

【0623】

SGPr479__1, 配列番号46はヒト染色体位置1q42にマップされている。

【0624】

SGPr489__1, 配列番号47はヒト染色体位置11p15にマップされている。

【0625】

SGPr465__1, 配列番号48は染色体位置が決定されていない。

【0626】

SGPr524__1, 配列番号49は染色体位置が決定されていない。

【0627】

SGPr422, 配列番号50はヒト染色体位置4q13にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuttilla) : 非小細胞肺癌。

【0628】

SGPr538, 配列番号51はヒト染色体位置11q23にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuttilla) : 胸膜中皮腫。

【0629】

SGPr527__1, 配列番号52は染色体位置が決定されていない。

【0630】

SGPr542, 配列番号53はヒト染色体位置19q13.1にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuttilla) : 小細胞肺癌 (強く関連, 試験した35人の患者のうち10人が増幅を示した)。

【0631】

SGPr551, 配列番号54はヒト染色体位置22q13にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuttilla) : 骨肉腫。

【0632】

SGPr451, 配列番号55はヒト染色体位置12q23にマップされている。

【0633】

SGPr452__1, 配列番号56はヒト染色体位置16p13.3にマップされている。

【0634】

SGPr504, 配列番号57は染色体位置が決定されていない。

【0635】

SGPr469, 配列番号58は染色体位置が決定されていない。

【0636】

SGPr400, 配列番号59はヒト染色体位置4q32にマップされている。この領域のゲノム増幅は以下の癌と関連づけられている (Knuttilla) : 非小細胞肺癌。

【0637】

実施例8b : 一塩基多型 (SNPs) 候補

材料および方法

ヒトDNAにおける最も一般的な変種は一塩基多型 (SNPs) であり, これは, 約100-300塩基に1回生ずる。SNPsは大規模の関連遺伝学研究を容易にすると予測されているため, 最近, SNPの発見および検出に多大な興味をもたれている。本明細書に

10

20

30

40

50

記載される遺伝子の候補SNPsは、核酸配列を公共のデータベースに登録されているSNPsを含む配列に対してblastn検索することにより同定した(dbSNP, NCBI, <http://www.ncbi.nlm.nih.gov/SNP/snpblastpretty.html>)。SNP含有配列についてdbSNP受託番号が与えられている。SNPsはまた、発現遺伝子(dbEST, NRNA)およびゲノム配列(すなわちNRNA)のいくつかのデータベースを一塩基対ミスマッチについて比較することにより同定した。結果は、表1の”SNPs”のカラムに示される。これらは候補SNPsである。ヒト集団におけるこれらの実際の頻度は決定していない。以下のコードはそれぞれのDNA配列を表す：

10

G = グアノシン

A = アデノシン

T = チミジン

C = シチジン

R = GまたはA, (プリン)

Y = CまたはT, (ピリミジン)

K = GまたはT, (ケト)

W = AまたはT, (弱) (2つの水素結合)

S = CまたはG, (強) (3つの水素結合)

M = AまたはC, (アミノ)

20

B = C, GまたはT (すなわち, A以外)

D = A, GまたはT (すなわち, C以外)

H = A, CまたはT (すなわち, G以外)

V = A, CまたはG (すなわち, T以外)

N = A, C, GまたはT, (任意)

X = A, C, GまたはT

相補的DNA鎖

G A T C R Y W S K M B V D H N X

+++++

C T A G Y R S W M K V B H D N X

例えば、2つのバージョンの遺伝子が存在し、一方は所定の位置に”C”を有し、他方は同じ位置に”T”を有する場合、その位置はYと表され、これはCまたはTを意味する。 30

SNPsは、遺伝子に付随する遺伝性の特徴を同定するために重要であろう。

【0638】

結果

SNPの同定の結果は上述の表2および実施例1の新規プロテアーゼポリヌクレオチドの説明と題する節に含まれる。上述したように、本発明のプロテアーゼポリヌクレオチドにおいて種々のSNPsが同定された。

【0639】

実施例9：サザンブロッティングによる遺伝子増幅の証明材料および方法

40

ナイロン膜はBoehringer Mannheimから購入する。変性溶液は、0.4M NaOHおよび0.6M NaClを含む。中和溶液は、0.5M Tris-HCl, pH7.5および1.5M NaClを含む。ハイブリダイゼーション溶液は、50%ホルムアミド, 6X SSPE, 2.5X デンハルト溶液, 0.2mg/mL変性サケDNA, 0.1mg/mL酵母tRNA, および0.2%ドデシル硫酸ナトリウムを含む。制限酵素はBoehringer Mannheimから購入する。放射性標識プローブは、StratageneのPrime-it IIキットを用いて調製する。プローブテンプレートに用いるベータアクチンDNAフラグメントはClontechから購入する。

【0640】

種々の腫瘍細胞株(例えば, MCF-7, MDA-MB231, Calu-6, A549 50

, HCT-15, HT-29, Colo205, LS-180, DLD-1, HCT-116, PC3, CAPAN-2, MIA-PaCa-2, PANC-1, AspC-1, BxPC-3, OVCAR-3, SKOV3, SW626 および PA-1), および2つの正常細胞株から, ゲノムDNAを単離する。

【0641】

各ゲノムDNA試料の10 μ gのアリコートをEcoRI制限酵素で消化し, 別の10 μ gの試料をHindIII制限酵素で消化する。制限酵素消化したDNA試料を0.7%アガロースゲルに負荷し, 電気泳動分離した後, 標準的方法によりDNAをナイロン膜にキャピラリー移動させる (Sambrook, J. et al (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory)。

【0642】

実施例10: ファージディスプレイによる蛋白質-蛋白質相互作用の検出

材料および方法

ファージディスプレイは, 所望のおとりに対する親和性に基づいて分子相互作用を単離する方法を提供する。ファージ被覆蛋白質への融合としてクローニングされたcDNAフラグメントは, ファージの表面にディスプレイされる。おとりと相互作用するファージをアフィニティー精製により濃縮し, 個々のクローンからの挿入DNAを分析する。

【0643】

T7ファージディスプレイライブラリ

すべてのライブラリは, T7 Select 1-lbベクター (Novagen) 中で, 製造元の指針にしたがって構築した。

【0644】

おとりの提示

おとりとして用いるべき蛋白質ドメインは, GSTへのC末端融合体として作成し, E. coliで発現させる。ペプチドは化学的に合成し, 長鎖スペーサービオチン試薬を用いてN末端でビオチン化する。

【0645】

選択

PanMixおよびE. coli阻害剤カクテル (Sigma P-8465) を補充した, 新たに調製したライブラリ (10^{10} - 10^{12} pfu) のアリコートを, 固定化したおとりとともに室温で1-2時間インキュベートする。未結合ファージを洗浄緩衝液でよく洗浄する (少なくとも4回)。3-4ラウンドの選択の後, 結合したファージを100 μ lの1% SDS中に溶出し, アガロースプレートに播種して, 単一プラークを得る。

【0646】

挿入DNAの同定

個別のプラークを25 μ lの10mMEDTA中に取り出し, 70 $^{\circ}$ Cで10分間加熱することによりファージを破壊する。2 μ lの破壊ファージを50 μ lのPCR反応混合物に加える。35ラウンドの熱サイクル (94 $^{\circ}$ C, 50秒間; 50 $^{\circ}$ C, 1分間; 72 $^{\circ}$ C, 1分間) により挿入DNAを増幅する。

【0647】

緩衝液の組成

10x PanMix
5% Triton X-100
10% 無脂乾燥乳 (Carnation)
10mMEGTA
250mMNaF
250 μ g/mLヘパリン (sigma)
250 μ g/mL, 剪断し, 沸騰させたサケ精子DNA (sigma)
0.05% アジ化ナトリウム

10

20

30

40

50

PBS中で調製する。

【0648】

洗浄緩衝液

PBS, 以下のもので補充:

0.5% NP-40

25 µg/mL ヘパリン

PCR 反応混合物

1.0 mL 10x PCR 緩衝液 (Perkin-Elmer, 15 mM Mg を含む)

各 0.2 mL の dNTPs (10 mM 保存液)

0.1 mL T7 UP プライマー (15 pmol/L) GGAGCTGTCGTATTC
CAGTC 10

0.1 mL T7 DN プライマー (15 pmol/L) AACCCCTCAAGACCC
GTTTAG

0.2 mL 25 mM MgCl₂ または MgSO₄, EDTA 補償用

蒸留水で 1.0 mL とする

反応液 50 µl あたり 1 ユニットの Taq ポリメラーゼを加える

ライブラリ: T7 Select 1-H441

【0649】

実施例 11: Incyte および公共 EST に基づく遺伝子発現

材料および方法

プロテアーゼの核酸配列を発現配列の Incyte および公共 dbEST データベースの BLASTN 検索におけるクエリーとして用いた。プロテアーゼが存在したライブラリの組織起源を、特定の組織に生ずる遺伝子の頻度とともに以下に示す。この頻度は、所定の組織起源中で遺伝子の存在を示すクローンの数により決定した。Incyte 遺伝子 ID 番号または公共 NCBI 受託番号が与えられており、次に組織起源が示される。次に、各遺伝子について組織特異性の簡単なまとめが記載されている。

【0650】

結果

SGPr397, 配列番号 1,

Incyte366783.1 クローン: 2 前立腺, 3 結腸, 網膜および小腸 3667
83.3 1 前立腺クローン 30

前立腺 (3/8 クローン) および結腸 (3/8 クローン) において選択的に発現

【0651】

SGPr413, 配列番号 2,

Incyte475365.6 5 クローン, 3 小腸, および前立腺, 乳腫瘍

475365.5 10 クローン: 8 小腸, および脳 (2)

小腸 (11/15 クローン) において非常に選択的に発現

【0652】

SGPr404, 配列番号 3,

Incyte1129157.1 213 クローン, 脳 (18), 雄/雌生殖器 (21/
23), 乳 (14), および消化系 (25) において高い 40

1129157.2 1: 混合物

広い発現, いくつかは脳 (18/214 クローン), 消化組織 (25/118) および雄/雌生殖器 (21, 23 クローン) および乳 (14 クローン) において上昇

【0653】

SGPr536_1, 配列番号 4,

Incyte233762.17 149 クローン: 21 以上のヒットの組織なし

233762.15 15 クローン, 混合物

164 クローンに広い発現が見られる

【0654】

20

50

S G P r 4 1 4, 配列番号 5,
I n c y t e 3 9 9 7 7 3. 5 6 6 9 クローン, 4 0 4 ライブラリ, 広く分布 広くかつ
強く発現 (6 6 9 クローン)

【0 6 5 5】

S G P r 4 3 0, 配列番号 6,
I n c y t e 4 0 7 8 2 3. 1 2 1 クローン (4 精巣, 3 脳, 3 前立腺)
1 1 3 6 4 8 3. 1 1 前立腺
4 1 1 2 4 6. 1 肺腫瘍, 小腸, 胎児肝臓に各 2, および 1 心臓
3 2 2 7 0 0. 1 T 細胞
4 0 7 8 2 3. 2 胎児肝臓/脾臓

10

混合発現, 精巣 (4 / 3 1 クローン), 脳 (3 / 3 1) および前立腺 (4 / 3 1) および
胎児肝臓/脾臓 (4 / 3 1) において最も高い

【0 6 5 6】

S G P r 4 9 6 _ 1, 配列番号 7,
I n c y t e 9 8 6 0 3 1. 1 1 2 クローン: 肺, 脳, 副腎腫瘍に各 2
肺 (2 / 1 2 クローン), 副腎腫瘍 (2 / 1 2) および脳 (2 / 1 2) において選択的に
発現

【0 6 5 7】

S G P r 4 9 5, 配列番号 8,
I n c y t e 3 5 0 9 2 1. 2 1 6 クローン: 2 胸腺, 3 結腸, 3 脳
3 5 0 9 2 1. 7 副腎腫瘍
3 5 0 9 2 1. 1 0 1 4 クローン: 2 結腸, 1 副腎腫瘍
3 5 0 9 2 1. 6 1 0 クローン: 2 副腎 (1 腫瘍), 2 脳
3 5 0 9 2 1. 1 副腎
3 5 0 9 2 1. 9 結腸 (2), 小腸腸, 肺腫瘍
副腎 (5 / 4 6 クローン) および結腸 (7 / 4 6) において選択的に発現

20

【0 6 5 8】

S G P r 4 0 7, 配列番号 9,
E S T s なし

【0 6 5 9】

30

S G P r 4 5 3, 配列番号 1 0,
I n c y t e 4 2 8 4 2 8. 1 1 7 クローン: 3 肺 (2 腫瘍), 2 前立腺, 4 精巣, 2
奇形腫 (h N T 2)
4 2 8 4 2 8. 5 脳, 肺, 奇形腫
4 2 8 4 2 8. 6 奇形腫 (2), 肺, 腎臓
h N T 2 奇形腫細胞株 (5 / 2 4 クローン) において高度に発現, 肺 (5 クローン) およ
び精巣 (4 クローン) に選択的

【0 6 6 0】

S G P r 4 4 5, 配列番号 1 1,
I n c y t e 3 5 0 9 2 1. 7 1 副腎腫瘍
3 5 0 9 2 1. 1 0 1 4, 例えば 1 副腎腫瘍
3 5 0 9 2 1. 6 1 0 クローン: 正常および腫瘍副腎 (各 1) 結腸腫瘍 (2) 3 5 0 9
2 1. 1 副腎
3 5 0 9 2 1. 8 2 前立腺腫瘍, 1 網膜
副腎 (5 / 2 8 クローン) において最も高く, 副腎ホルモンプロセッシングにおける関与の
可能性を示す

40

【0 6 6 1】

S G P r 4 0 1 _ 1, 配列番号 1 2,
I n c y t e 2 3 2 4 1 4. 1 1 6 9 クローン: 6 9 N S, 1 8 雄生殖器, (1 0 前立
腺), 1 1 雌生殖器, 1 1 呼吸系, 5 腎臓, 1 つの神経膠芽細胞腫ライブラリに 9

50

2 3 2 4 1 4. 2 精巣

神経系 (6 9 / 1 7 0 クローン), 特に神経膠芽細胞腫に選択的

【0 6 6 2】

S G P r 4 0 8, 配列番号 1 3,

I n c y t e 2 3 3 6 6 0. 2 3 5 7 クローン, 2 4 8 ライブラリ: 5 4 脳, 2 6 造血
/ 免疫, 2 4 / 2 1 f / m 生殖器, 2 1 消化系, 2 0 心臓血管

2 3 3 6 6 0. 1 1 1 3 クローン, 広い発現

2 3 3 6 6 0. 1 0 7 クローン, 広い発現

広くかつ強く発現 (3 7 7 クローン)

【0 6 6 3】

S G P r 4 8 0, 配列番号 1 4,

I n c y t e 1 3 2 6 2 5 6. 3 2 7 4 クローン, 広いが, N S (5 9), 造血 (3 5)
) 生殖器 (2 4 / 1 0, m / f) で最も高い。1 つの唾液腺ライブラリに 6 クローン

1 3 2 6 2 5 6. 8 4 混合クローン

1 3 2 6 2 5 6. 1 2 6 クローン, 1 3 雄生殖器

1 3 2 6 2 5 6. 1 0 1 0 クローン, 混合

広く強い発現 (3 0 0 クローン以上)

【0 6 6 4】

S G P r 4 3 1, 配列番号 1 5,

I n c y t e 2 3 6 3 6 8. 1 1 5 1 クローン, 1 1 0 ライブラリ, 消化系, 神経, 造 20
血 (1 8, 1 7, 1 8) において最も高い。5, 2 つの胎児肝臓 / 悲脾臓ライブラリのそ
れぞれにおいて 4 ヒット。

2 3 6 3 6 8. 2 1 胎児心臓

2 3 6 3 6 8. 1 4 7 : 混合

広く中程度に強い発現 (合計 1 5 9 クローン)

【0 6 6 5】

S G P r 4 2 9, 配列番号 1 6,

I n c y t e 8 9 0 5 4 0. 9 4 1 クローン, 広い

8 9 0 5 4 0. 1 1 2 5 クローン, 広い

8 9 0 5 4 0. 8 1 5 クローン, 広い

広く中程度に強い発現 (1 8 1 クローン)

【0 6 6 6】

S G P r 5 0 3, 配列番号 1 7,

I n c y t e 1 4 4 7 3 5 7. 3 1 0 7 クローン, N S (1 5), 雄生殖器 (1 1) お
よび消化組織 (1 5) において最も高い。

1 4 4 7 3 5 7. 1 樹状突起細胞

2 4 5 0 4 5. 1 1 6 混合

広い発現 (1 2 4 クローン), 神経系 (1 6), 雄生殖器 (1 1) および消化組織 (1 5)
) において最も高い

【0 6 6 7】

S G P r 4 2 7, 配列番号 1 8,

I n c y t e 9 0 3 0 9 2. 3 1 4 1 クローン, 3 5 ライブラリ; 9 クローン, 前立腺
, それ以外は非常に広い

9 0 3 0 9 2. 2 3 1 脳

前立腺 (9 / 4 2 クローン) において発現が上昇

【0 6 6 8】

S G P r 0 9 2, 配列番号 1 9,

I n c y t e 3 3 9 2 5 1. 1 4 / 5 子宮, 1 混合組織

3 3 9 2 5 1. 2 1 / 1 子宮; 子宮 (5 / 6 クローン) において非常に選択的に発現

【0 6 6 9】

10

20

30

40

50

S G P r 3 5 9 , 配列番号 2 0 ,
 I n c y t e 3 9 1 1 3 3 . 1 混合組織 (胎児肺, 精巣, B-細胞)
 g i | 7 2 8 0 3 9 9 = I n c y t e と同じ
 混合組織, 1つの E S T のみ

【0670】

S G P r 1 0 4 _ 1 , 配列番号 2 1 ,
 I n c y t e 1 2 / 2 3 クローン, 脳
 2 3 2 0 1 5 . 5 6 / 7 クローン, 脳
 2 3 2 0 1 5 . 2 2 / 4 クローン, 脳
 2 3 2 0 1 5 . 1 1 / 1 脳
 2 3 2 0 1 5 . 6 1 / 1 脳

脳特異的

【0671】

S G P r 3 0 3 , 配列番号 2 2 ,
 I n c y t e 3 2 3 8 4 6 . 1 5 3 8 サンプル, 脳 (8) 乳 (4), 子宮および卵巣 (7) において最も高い。
 3 2 3 8 4 6 . 1 3 0 4 クローン, 神経系 (72) および生殖器 (28 f, 24 m) において最も高い。他の組織
 4 1 4 0 4 8 . 3 4 4 5 クローン, N S において最も高い
 3 2 3 8 4 6 . 1 1 8 クローン

広い発現

【0672】

S G P r 4 0 2 _ 1 , 配列番号 2 3 ,
 I n c y t e 2 4 4 4 0 7 . 4 2 5 クローン, 精巣 (8) 脳 (4), 子宮 (2; 1 腫瘍) において最も高い。
 2 4 4 4 0 7 . 2 子宮
 2 4 4 4 0 7 . 1 精巣
 2 4 4 4 0 7 . 6 子宮腫瘍
 2 4 4 4 0 7 . 5 ファローピウス管腫瘍
 2 4 4 4 0 7 . 9 混合組織, 例えば腫瘍, 鼻腫瘍

生殖器試料で濃縮

【0673】

S G P r 4 3 4 , 配列番号 2 4 ,
 I n c y t e 1 1 0 1 5 4 . 4 クローン起源なし,
 1 1 0 1 5 4 . 6 2 前立腺
 1 1 0 1 5 4 . 1 2 1 前立腺, 1 下垂体
 1 1 0 1 5 4 . 1 1 1 下垂体
 1 1 0 1 5 4 . 8 ファローピウス管腫瘍 (2), 混合 (1)
 1 1 0 1 5 4 . 7 下垂体
 1 1 0 1 5 4 . 5 大腿筋肉 (2) - 組織特異的スプライシング
 1 1 0 1 5 4 . 1 0 3 心臓, 2 脳, 下垂体 (61 / 62 マッチ)
 前立腺 (3 / 17 クローン), 4 唾液腺 (4 / 17 クローン) およびファローピウス管腫瘍 (2 / 17 クローン) において選択的発現。下垂体におけるホルモンプロセッシングおよび前立腺ホルモンプロセッシングにおける役割を示すかもしれない。

【0674】

S G P r 4 4 6 _ 1 , 配列番号 2 5 ,
 I n c y t e 1 0 4 0 6 4 1 . 1 心臓, 筋肉
 1 3 8 8 3 7 1 . 1 2 心臓
 筋肉 (3 / 3 クローン), 特に心臓筋肉 (2 / 3 クローン) に特異的

【0675】

10

20

30

40

50

S G P r 4 4 7, 配列番号 2 6,
I n c y t e 1 3 5 2 9 3 2. 1 膵臓腫瘍
膵臓腫瘍から 1 個のクローン

【0 6 7 6】

S G P r 4 3 2 _ 1, 配列番号 2 7,
I n c y t e 4 7 4 6 7 4. 1 5 2 9 クローン, 混合
4 7 4 6 7 4. 3 0 9 0 クローン, 混合
4 7 4 6 7 4. 1 8 2 クローン, 混合
広く強い発現 (合計 2 0 1 クローン)

【0 6 7 7】

S G P r 5 2 9, 配列番号 2 8,
I n c y t e 9 8 8 0 1 9. 3 7 1 クローン, 2 3, 雌生殖器。9, 1 つの卵巢腫瘍ラ
イブラリ, 2 その他, 2 その他, およびさらにその他から 1 (正常卵巢なし)。5, 1 つ
の膵臓腫瘍細胞株, 4, 膵臓腫瘍ライブラリ
9 8 8 0 1 9. 1 乳皮膚
膵臓 (4 / 7 2 クローン, 1 つの膵臓腫瘍ライブラリから, および膵臓腫瘍細胞株から 5
) および卵巢 (卵巢腫瘍から 1 4, 正常卵巢から 0) において選択的に発現

【0 6 7 8】

S G P r 4 2 8 _ 1, 配列番号 2 9,
I n c y t e 8 9 1 1 4 6. 1 4 : 脳, 下垂体, 血液, 胸腺
広く低レベルの発現 (4 クローン, すべて異なる組織より)

【0 6 7 9】

S G P r 4 2 5, 配列番号 3 0,
I n c y t e 4 0 0 8 3 3. 1 2 5 クローン, 混合 (任意の組織から 4 未満, ただし胎
児性から 5)
広いが強くない発現 (合計 2 5 クローン)

【0 6 8 0】

S G P r 5 4 8, 配列番号 3 1,
I n c y t e 9 7 1 2 3 6. 1 2 クローン, 混合精巣, 胎児肺, B 細胞の混合物から
転写産物はほとんど存在せず, 精巣, 胎児肺および B 細胞の混合物ライブラリから 2 つの
クローンのみ

【0 6 8 1】

S G P r 3 9 6, 配列番号 3 2,
I n c y t e 2 0 9 0 5 1. 1 肺 (1)
8 8 9 1 2 6. 1 脳 (1)
2 つの E S T s のみー肺および脳

【0 6 8 2】

S G P r 4 2 6, 配列番号 3 3,
I n c y t e E S T s なし

【0 6 8 3】

S G P r 5 5 2, 配列番号 3 4,
I n c y t e 1 5 1 0 5 1 2. 1 扁桃, 脊髄
1 5 1 1 2 2 2. 1 扁桃
4 0 6 2 2 1. 1 8 3 クローン: N S に 1 6, 造血 / 免疫に 1 0, 雄生殖器に 9, およ
びいくつかの他の組織。1 扁桃,
9 8 1 3 5 5. 3 8 クローン, 2 卵巢腫瘍, 1 扁桃, 種々
9 4 クローンのうち扁桃 (3 / 9 4, しかし, 扁桃は通常は発現起源としてはあまり見ら
れない), および神経系 (1 7 クローン) においてある程度の選択性が見られる

【0 6 8 4】

S G P r 4 0 5, 配列番号 3 5,

10

20

30

40

50

I n c y t e 1 3 4 3 6 0 . 1 1 腎臓

1つのクローン, 腎臓から

【0685】

S G P r 4 8 5 _ 1 , 配列番号36,

I n c y t e 1 8 0 5 7 6 . 2 5 / 5 精巣のクローン

精巣特異的 (5 / 5 クローン)

【0686】

S G P r 5 3 4 , 配列番号37,

I n c y t e 1 3 8 3 3 9 1 . 2 0 1 1 2 / 1 1 4 , 開始 (1 0 3 - 1 6 5 = 完全マッ
チ) においてはよくマッチするが, テンプレートアルチファクトかもしれない 10

1 4 5 0 8 1 2 . 1 1 / 1 脾臓, いくつかのミスマッチがN側に存在

1 3 8 3 3 9 1 . 1 3 5 / 5 脾臓

1 0 4 5 8 3 4 . 1 1 / 1 脾臓

ほぼ完全に脾臓特異的 (1 1 8 / 1 2 0 , 脾臓からのクローン)

【0687】

S G P r 3 9 0 , 配列番号38,

I n c y t e 1 9 9 4 2 8 . 9 骨腫瘍, 小腸

1 9 9 4 2 8 . 3 3 8 2 クローン: 4 1 脳, 3 4 / 2 3 生殖器 (m / f) , 2 2 造血 /
免疫, 2 7 消化系

広い組織分布, 脳 (4 1 / 3 8 2 クローン) , 雄および雌生殖器 (それぞれ 3 4 および 2
3 クローン) および消化系 (2 7 クローン) において最も高い 20

【0688】

S G P r 5 2 1 , 配列番号39,

I n c y t e 4 2 7 8 2 6 . 1 2 8 クローン, 小腸腫瘍のほとんど (5 , 1 ライブラリ
) , 新生児ケラチノサイト (2 ライブラリから各 3) , 8 , 卵巣腫瘍, 5 , 乳皮膚

卵巣腫瘍 (8 / 2 8 クローン) , 新生児ケラチノサイト (6 / 2 8) , 乳ケラチノサイト
(5) および小腸腫瘍ライブラリ (5 クローン, 1 つのライブラリより) において選択的
発現

【0689】

S G P r 5 3 0 _ 1 , 配列番号40,

E S T s なし

【0690】

S G P r 5 2 0 , 配列番号41,

I n c y t e 4 0 5 9 4 7 . 1 4 / 4 クローン副腎腫瘍 (褐色細胞腫) (1 つのライブ
ラリより 3 , 別のライブラリより 1)

1 3 3 8 6 5 2 . 1 1 / 1 , 副腎腫瘍 (褐色細胞腫) からのクローン

1 4 7 7 1 8 9 . 1 1 / 1 , 副腎 (混合正常および褐色細胞腫) からのクローン

褐色細胞腫 (副腎腫瘍) に特異的: 褐色細胞腫より 4 / 5 クローン, 混合正常副腎および
褐色細胞腫より 1 / 5 クローン

【0691】

S G P r 4 5 5 , 配列番号42,

I n c y t e 1 1 1 5 8 3 3 . 1 , 混合胎児肺 / 精巣 / B 細胞

9 8 7 2 7 9 . 1 脳 (1) , 腫瘍を含む混合組織 (1)

3 つのクローン, 1 つ (脳) のみが特異的起源

【0692】

S G P r 5 0 7 _ 2 , 配列番号43,

I n c y t e 4 0 3 8 9 1 . 1 1 0 クローン: 精巣に 6 , 混合物 (精巣, 肺, B 細胞)
に 6

4 0 3 8 9 1 . 2 1 脳

精巣選択的: 精巣より 6 / 1 1 クローン, 精巣試料を含む混合物ライブラリより 5 / 1 1 50

クローン

【0693】

S G P r 5 5 9, 配列番号 4 4,
I n c y t e 4 7 5 1 0 0. 1 3 5 クローン, 1 1 雌生殖器, 8 消化器, 7 子宮腫瘍 (正常にはなし), 4 乳, 2 卵巣腫瘍, 1 H e L a 子宮頸部腫瘍
4 7 5 1 0 0. 6 T h l 細胞, H e L a 細胞
子宮 (7 / 3 7 クローン), 卵巣 (2 / 3 7) 子宮頸部 (2 / 3 7, H e L a 子宮頸部腫瘍細胞株から), ならびに乳 (4) の腫瘍において選択的発現

【0694】

S G P r 5 6 7 _ 1, 配列番号 4 5, 10
I n c y t e 9 8 1 3 5 5. 3 混合 (卵巣腫瘍ライブラリより 2 / 8 クローン, 扁桃, 脳, 肺腫瘍, 心臓, 胎盤, 脊髄神経節より各 1)
広い発現はほとんどない (7 種の異なる組織より 8 クローン)

【0695】

S G P r 4 7 9 _ 1, 配列番号 4 6,
I n c y t e 2 1 9 2 1 4. 1 1 精巣
1 つの E S T, 精巣において発現

【0696】

S G P r 4 8 9 _ 1, 配列番号 4 7, 20
I n c y t e 3 3 8 9 5 6. 1 2 腎臓, 1 胎盤
1 0 4 2 3 0 6. 1 1 口腫瘍, 2 ファローピウス管腫瘍
1 3 8 4 8 2 4. 1 小腸, 腎臓
まれであるが広い発現, 腎臓 (3 / 8 クローン) およびファローピウス管腫瘍 (2 / 8, 1 つのライブラリから) に選択的

【0697】

S G P r 4 6 5 _ 1, 配列番号 4 8,
E S T s なし

【0698】

S G P r 5 2 4 _ 1, 配列番号 4 9, 30
I n c y t e 9 5 2 1 8 2. 3 1 精巣
9 5 2 1 8 2. 2 1 精巣
9 5 2 1 8 2. 4 1 前立腺
雄生殖器に特異的 (2 / 3 クローン, 精巣, 1 / 3, 前立腺)

【0699】

S G P r 4 2 2, 配列番号 5 0,
I n c y t e 1 5 1 1 2 8 4. 1 1 扁桃
1 3 5 1 2 5 9. 1 1 脳
扁桃 (1 / 2 クローン) および脳 (1 / 2 クローン) においてのみまれな転写産物が見られる

【0700】

S G P r 5 3 8, 配列番号 5 1, 40
I n c y t e 9 0 3 0 9 2. 2 9 4 クローン, 3 脳, 1 乳
9 0 3 0 9 2. 1 9 1 脳
9 0 3 0 9 2. 2 2 2 脳
9 0 3 0 9 2. 2 8 3 8 クローン, 脳に 2 4
9 0 3 0 9 2. 2 4 2 脳, 1 小腸
神経系 (3 2 / 4 8 クローン) において選択的に発現

【0701】

S G P r 5 2 7 _ 1, 配列番号 5 2, 50
I n c y t e 6 5 4 5 0. 1 1 前立腺腫瘍

1 0 3 5 5 4 . 2 混合組織，腫瘍を含む
 2 2 8 4 5 6 . 2 1 1 クローン：混合（2 脳，2 血液）
 1 0 3 5 5 4 . 1 混合（3）
 広い低レベル発現（1 6 クローン）

【0702】

S G P r 5 4 2 , 配列番号 5 3 ,
 I n c y t e 2 4 4 0 8 5 . 1 発現は造血細胞に選択的である：1 1 クローンはすべて造血組織からである：胎児肝臓／脾臓より 6 , このうち 4 は肥満細胞，臍帯血より 2 , C D 3 4 + 骨髄より 1 , およびリンパ種より 2 クローン：A M L 芽球より 1 および C M L より 1 。

10

【0703】

S G P r 5 5 1 , 配列番号 5 4 ,
 I n c y t e 3 1 9 5 2 9 . 1 2 2 クローン：7 肝臓，1 胎児肝臓／脾臓，3 肺
 3 1 9 5 2 9 . 2 1 肝臓
 3 1 9 5 2 9 . 3 2 混合組織，例えば精巣，1 精巣（組織特異的スプライシング）
 肝臓（9 / 2 6 クローン）において選択的に発現。精巣特異的スプライシング型を有するかもしれない（1 つのテンプレートの 3 / 3 クローン）

【0704】

S G P r 4 5 1 , 配列番号 5 5 ,
 I n c y t e 1 4 7 1 5 4 1 . 1 1 混合

20

【0705】

S G P r 4 5 2 _ 1 , 配列番号 5 6 ,
 I n c y t e 4 4 6 3 7 4 . 1 混合（メラノサイト，子宮，胎児心臓）
 発現データなし（混合組織より 1 つの E S T ）

【0706】

S G P r 5 0 4 , 配列番号 5 7 ,
 I n c y t e 2 4 4 0 8 5 . 1 発現は造血細胞に特異的：1 1 クローンはすべて造血組織より：胎児肝臓／脾臓より 6 , このうち 4 は肥満細胞，臍帯血より 2 , C D 3 4 + 骨髄より 1 , およびリンパ種より 2 クローン：A M L 芽球より 1 および C M L より 1

【0707】

S G P r 4 6 9 , 配列番号 5 8 ,
 I n c y t e 1 1 0 1 5 4 . 1 0 7 クローン：3 心臓，2 脳，1 下垂体
 1 1 0 1 5 4 . 3 心臓，筋肉，精巣心臓（4 / 1 0 クローン）において選択的発現

30

【0708】

S G P r 4 0 0 , 配列番号 5 9 ,
 I n c y t e 8 8 9 1 2 6 . 1 脳
 1 つの E S T のみ，脳

【0709】

結論

当業者は，本発明は，その目的を実施し，記載される結果および利点，ならびに本明細書に固有のものを得るのによく適合していることを容易に理解するであろう。本明細書に記載される分子複合体および方法，手順，処理，分子，特定の化合物は，現在のところ好ましい態様の代表的なものであり，例示的なものであって，本発明の範囲を限定することを意図するものではない。当業者は，本発明の範囲および精神から逸脱することなく，本明細書に開示される本発明に対して置換基の変更および改変をなすことができることを容易に理解するであろう。

40

【0710】

本明細書において言及されるすべての特許および刊行物は，本発明の属する技術分野の技術者のレベルを示す。すべての特許および刊行物は，それぞれの刊行物が特定の個々に本明細書の一部としてここに引用されることと同じ程度に，本明細書の一部として引用さ

50

れる。

【0711】

本明細書に例示的に記載されている発明は、本明細書に特定の開示されていない任意の要素または限定なしでも適切に実施することができる。すなわち、例えば、本明細書における各例において、“・・・を含む”、“・・・から本質的になる”および“・・・からなる”との用語は、他の2つのいずれかと置き換えることができる。本明細書において用いた用語および表現は、説明の用語として用いるものであり、限定ではない。そのような用語および表現の使用においては、示されかつ記載されている特徴またはその一部の等価物を除外することを意図するものではなく、特許請求の範囲に記載される本発明の範囲中で種々の変更が可能であることが理解される。すなわち、好ましい態様および任意の特徴により本発明を特定の開示してきたが、当業者には本明細書に記載される概念の変更および変種が可能であり、そのような変更および変種も特許請求の範囲に定義される本発明の範囲内であると考えられることが理解されるべきである。

10

【0712】

さらに、発明の特徴および局面がマーカッシュグループの用語で記載されている場合、当業者は、本発明が、マーカッシュグループのメンバーの個々のメンバーまたはサブグループに関してもまた記載されていることを認識するであろう。例えば、Xが、臭素、塩素およびヨウ素からなる群より選択されるとして記載されている場合、Xが臭素である特許請求の範囲およびXが臭素および塩素である特許請求の範囲も完全に記載されている。

【0713】

遺伝コードの縮重の観点から、核酸の他の組み合わせもまた本明細書のペプチドおよび蛋白質をコードする。例えば、GCT, GCC, GCA, GCGの4つの核酸配列はすべてアミノ酸アラニンをコードする。したがって、あるアミノ酸について平均で3つのコドンが存在し、100アミノ酸の長さのポリペプチドは平均で 3^{100} 、もしくは 5×10^4 種類の核酸配列によりコードされるであろう。すなわち、日常的な方法を用いて過度の実験なしに、核酸配列を改変して第1の核酸配列によりコードされるものと同じポリペプチドをコードする第2の核酸配列を形成することができる。すなわち、特許請求の範囲に記載されるペプチドおよび蛋白質をコードするすべての可能な核酸もまた、コドン使用、特にヒトにおいて好ましいものを完全に考慮してこれらがすべて書き出されているように、本明細書に完全に記載されている。さらに、ポリペプチドの有意な活性が変更しない配列の領域内において、ポリペプチドのアミノ酸配列、またはそのようなポリペプチドをコードする対応する核酸配列の変更が生ずるよう設計しまたは選択することができる。例えば、ポリペプチドの活性部位と離れたβターン中で、アミノ酸変化を生じさせることができる。また、欠失（例えば活性部位に影響を与えないポリペプチドのセグメントまたはそのようなポリペプチドをコードする対応する核酸配列を除去する）および付加（例えば、活性部位の機能に影響を与えずに、ポリペプチド配列により多くのアミノ酸を付加する、例えばGST融合蛋白質を形成する、または活性部位の機能に影響を与えずにそのようなポリペプチドをコードする対応する核酸配列に付加する）もまた本発明の範囲内である。ポリペプチドに対するそのような変更は、当業者が日常的な方法を用いて過度の実験なしに行うことができる。すなわち、本発明のペプチドまたは蛋白質の有意な活性に影響を与えないと容易に決定することができるすべての可能な核酸および／またはアミノ酸配列もまた本明細書に完全に記載されている。

20

30

40

【0714】

本明細書においては、本発明を広くかつ一般的に記載している。一般的開示に含まれるより狭い種および垂属のそれぞれのグループもまた本発明の一部を形成する。これには、除かれたものが具体的に記載されているか否かにかかわらず、属から任意の主題を除く「ただし・・・」またはネガティブ限定を含む発明の一般的記載が含まれる。

【0715】

他の態様は特許請求の範囲の範囲内である。

【図面の簡単な説明】

50

【図 1】図 1 は、ヒト蛋白質プロテアーゼのヌクレオチド配列（配列番号 1－59）を 5' から 3' の方向で示す。

【図 2】図 2 は、配列番号 1－59 によりコードされるヒトプロテアーゼのアミノ酸配列（配列番号 60－118）を翻訳の方向で示す。

WO 02/00860 A2

WO 02/00860 A2

**Published:**

— without international search report and to be republished
upon receipt of that report

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

WO 02/00860

PCT/US01/20171

5

10

NOVEL PROTEASES

The present invention claims priority to provisional application serial no. 60/214,047 filed June 26, 2000, which is hereby incorporated by reference in its entirety.

15

FIELD OF THE INVENTION

The present invention relates to protease polypeptides, nucleotide sequences encoding the protease polypeptides, as well as various products and methods useful for the diagnosis and treatment of various protease-related diseases and conditions.

20

BACKGROUND OF THE INVENTION**Proteases and Human Disease**

"Protease," "proteinase," and "peptidase" are synonymous terms applying to all enzymes that hydrolyse peptide bonds, *i.e.* proteolytic enzymes. Proteases are an exceptionally important group of enzymes in medical research and biotechnology. They are necessary for the survival of all living creatures, and are encoded by 1-2% of all mammalian genes. Rawlings and Barrett (MEROPS: the peptidase database. *Nucleic Acids Res.*, 1999, 27:325-331) (<http://www.babraham.co.uk/Merops/Merops.htm>) (Which is incorporated herein by reference in its entirety including any figures, tables, or drawings.) have classified peptidases into 157 families based on

25

30

structural similarity at the catalytic core sequence. These families are further classed into 26 clans, based on indications of common evolutionary relationship. Peptidases play key roles in both the normal physiology and disease-related pathways in mammalian cells. Examples include the modulation of apoptosis (caspases), control of blood pressure (renin, angiotensin-converting enzymes), tissue remodeling and tumor invasion (collagenase), the development of Alzheimer's Disease (β -secretase), protein turnover and cell-cycle regulation (proteasome), and inflammation (TNF- α convertase). (Barrett *et al.*, Handbook of Proteolytic Enzymes, 1998, Academic Press, San Diego which is incorporated herein by reference in its entirety including any figures, tables, or drawings.)

Peptidases are classed as either exopeptidases or endopeptidases. The exopeptidases act only near the ends of polypeptide chains: aminopeptidases act at the free N-terminus and carboxypeptidases at the free C-terminus. The endopeptidases are divided, on the basis of their mechanism of action, into six subclasses: aspartyl endopeptidases (3.4.23), cysteine endopeptidases (3.4.22), metalloendopeptidases (3.4.24), serine endopeptidases (3.4.21), threonine endopeptidases (3.4.25), and a final group that could not be assigned to any of the above classes (3.4.99). (Enzyme nomenclature and numbering are based on "Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) 1992, (<http://www.chem.qmw.ac.uk/iubmb/enzyme/EC34/intro.html>).)

In serine-, threonine- and cysteine-type peptidases, the catalytic nucleophile is the reactive group of an amino acid side chain, either a hydroxyl group (serine- and threonine-type peptidases) or a sulfhydryl group (cysteine-type peptidases). In aspartic-type and metallopeptidases, the nucleophile is commonly an activated water molecule. In aspartic-type peptidases, the water molecule is directly bound by the side chains of aspartate residues. In metallopeptidases, one or two metal ions hold the water molecule in place, and charged amino acid side chains are ligands for the metal ions. The metal may be zinc, cobalt or manganese. One metal ion is usually

WO 02/00860

PCT/US01/20171

attached to three amino acid ligands. Families of peptidases are referred to by use of the numbering system of Rawlings & Barrett (Rawlings, N. D. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Research* 27 (1999) 325-331, which is incorporated herein by reference in its entirety including any figures, tables, or drawings).). Enzyme nomenclature and numbering are based on "Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) 1992, (<http://www.chem.qmw.ac.uk/iubmb/enzyme/EC34/intro.html>).

10

Protease Families

1. Aspartyl proteases (Prosite number PS00141)

Aspartyl proteases, also known as acid proteases, are a widely distributed family of proteolytic enzymes in vertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue. The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Enzymes in this class include cathepsin E, renin, presenilin (PS1), and the APP secretases.

2. Cysteine proteases (Prosite PDOC00126)

Eukaryotic cysteine proteases are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. Peptidases in this family with important roles in disease include the caspases, calpain, hedgehog, ubiquitin hydrolases, and papain.

3. Metalloproteases (Prosite PDOC00129)

WO 02/00860

PCT/US01/20171

The metalloproteases are a class which includes matrix metalloproteases (MMPs), collagenase, stromelysin, gelatinase, neprylisin, carboxypeptidase, dipeptidase, and membrane-associated metalloproteases, such as those of the ADAM family. They require a metal co-factor for activity; frequently the required metal ion is zinc but some metalloproteases utilize cobalt and manganese.

Proteins of the extracellular matrix interact directly with cell surface receptors thereby initiating signal transduction pathways and modulating those triggered by growth factors, some of which may require binding to the extracellular matrix for optimal activity. Therefore the extracellular matrix has a profound effect on the cells encased by it and adjacent to it. Remodeling of the extracellular matrix requires protease of several families, including metalloproteases (MMPs).

4. Serine proteases (S1) (Prosite PS00134 trypsin-his; PS00135 trypsin-ser)
The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases. A partial list of proteases known to belong to this large and important family include: blood coagulation factors VII, IX, X, XI and XII; thrombin; plasminogen; complement components C1r, C1s, C2; complement factors B, D and I; complement-activating component of RA-reactive factor; elastases 1, 2, 3A, 3B (protease E); hepatocyte growth factor activator; glandular (tissue) kallikreins including EGF-binding protein types A, B, and C; NGF- γ hain, γ -renin, and prostate specific antigen (PSA); plasma kallikrein; mast cell proteases; myeloblastin (proteinase 3) (Wegener's autoantigen); plasminogen activators (urokinase-type, and tissue-type); and the trypsins I, II, III, and IV. These peptidases play key roles in coagulation, tumorigenesis, control of blood pressure, release of growth factors, and other roles.

WO 02/00860

PCT/US01/20171

5. Threonine peptidases (T1) -- (Prosite PDOC00326/PDOC00668)

Threonine proteases are characterized by their use of a hydroxyl group of a threonine residue in the catalytic site of these enzymes. Only a few of these enzymes have been characterized thus far, such as the 20S proteasome from the archaeobacterium *Thermoplasma acidophilum* (Seemuller *et al.*, 1995, *Science*, 268:579-82, and chapter 167 of Barrett *et al.*, *Handbook of Proteolytic Enzymes*, 1998, Academic Press, San Diego).

SUMMARY OF THE INVENTION

10 This invention concerns the isolation and characterization of novel sequences of human proteases. These sequences are obtained via bioinformatics searching strategies on the predicted amino acid translations of new human genetic sequences. These sequences, now identified as proteases, are translated into polypeptides which are further characterized. Additionally, the nucleic acid sequences of these proteases
15 are used to obtain full-length cDNA clones of the proteases. The partial or complete sequences of these proteases are presented here, together with their classification, predicted or deduced protein structure.

Modulation of the activities of these proteases will prove useful therapeutically. Additionally, the presence or absence of these proteases or the
20 DNA sequence encoding them will prove useful in diagnosis or prognosis of a variety of diseases. In this regard, Example 8 describes the chromosomal localization of proteases of the present invention, and describes diseases mapping to the chromosomal locations of the proteases of the invention.

A first aspect of the invention features an identified, isolated, enriched, or
25 purified nucleic acid molecule having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID

WO 02/00860

PCT/US01/20171

NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID
 NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID
 NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID
 NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID
 5 NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID
 NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ
 ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106,
 SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID
 NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115,
 10 SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 and biological domains
 thereof.

The term "identified" in reference to a nucleic acid is meant that a sequence
 was selected from a genomic, EST, or cDNA sequence database based on being
 predicted to encode a portion of a previously unknown or novel protease.

15 By "isolated" in reference to nucleic acid is meant a polymer of 10
 (preferably 21, more preferably 39, most preferably 75) or more nucleotides
 conjugated to each other, including DNA and RNA that is isolated from a natural
 source or that is synthesized as the sense or complementary antisense strand. In
 certain embodiments of the invention, longer nucleic acids are preferred, for
 20 example those of 300, 600, 900, 1200, 1500, or more nucleotides and/or those
 having at least 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,
 97%, 98% or 99% identity to a sequence selected from the group consisting of those
 set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID
 NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10,
 25 SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15,
 SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20,
 SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25,
 SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30,
 SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35,

WO 02/00860

PCT/US01/20171

SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40,
SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45,
SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50,
SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55,
5 SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ ID NO:59.

It is understood that by nucleic acid it is meant, without limitation, DNA,
RNA or cDNA, and where the nucleic acid is RNA, the thymine (T) will be uracil
(U).

The isolated nucleic acid of the present invention is unique in the sense that
10 it is not found in a pure or separated state in nature. Use of the term "isolated"
indicates that a naturally occurring sequence has been removed from its normal
cellular (*i.e.*, chromosomal) environment. Thus, the sequence may be in a cell-free
solution or placed in a different cellular environment. The term does not imply that
the sequence is the only nucleotide chain present, but that it is essentially free
15 (preferably about 90% pure, more preferably at least about 95% pure) of non-
nucleotide material naturally associated with it, and thus is distinguished from
isolated chromosomes.

By the use of the term "enriched" in reference to nucleic acid is meant that
the specific DNA or RNA sequence constitutes a significantly higher fraction (2- to
20 5-fold) of the total DNA or RNA present in the cells or solution of interest than in
normal or diseased cells or in the cells from which the sequence was taken. This
could be caused by a person by preferential reduction in the amount of other DNA or
RNA present, or by a preferential increase in the amount of the specific DNA or
RNA sequence, or by a combination of the two. However, it should be noted that
25 enriched does not imply that there are no other DNA or RNA sequences present, just
that the relative amount of the sequence of interest has been significantly increased.
The term "significant" is used to indicate that the level of increase is useful to the
person making such an increase, and generally means an increase relative to other
nucleic acids of about at least 2-fold, more preferably at least 5-fold, more

WO 02/00860

PCT/US01/20171

preferably at least 10-fold or even more. The term also does not imply that there is no DNA or RNA from other sources. The DNA from other sources may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes from naturally occurring events, such as viral
5 infection, or tumor-type growths, in which the level of one mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in
10 purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation). Instead, it represents an indication that the sequence is relatively more pure than in the natural environment (compared to the natural level this level should be at least 2- to 5-fold greater, e.g., in terms of mg/mL). Individual clones isolated from a cDNA library may be
15 purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones could be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic
20 substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^6 -fold purification of the native message. Thus, purification of at least one order of magnitude, preferably
25 two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

By a "protease polypeptide" is meant 32 (preferably 40, more preferably 45, most preferably 55) or more contiguous amino acids in a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID

WO 02/00860

PCT/US01/20171

NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID
 NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID
 NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID
 NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID
 5 NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID
 NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID
 NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID
 NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID
 NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104,
 10 SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID
 NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113,
 SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID
 NO:118 and biological domains thereof. In certain aspects, polypeptides of 100,
 200, 300, 400, 450, 500, 550, 600, 700, 800, 900 or more amino acids are preferred.

15 The protease polypeptide can be encoded by a full-length nucleic acid sequence or
 any portion of the full-length nucleic acid sequence, so long as a functional activity
 of the polypeptide is retained. It is well known in the art that due to the degeneracy
 of the genetic code numerous different nucleic acid sequences can code for the same
 amino acid sequence. Equally, it is also well known in the art that conservative
 20 changes in amino acid can be made to arrive at a protein or polypeptide which
 retains the functionality of the original. Such substitutions may include the
 replacement of an amino acid by a residue having similar physicochemical
 properties, such as substituting one aliphatic residue (Ile, Val, Leu or Ala) for
 another, or substitution between basic residues Lys and Arg, acidic residues Glu and
 25 Asp, amide residues Gln and Asn, hydroxyl residues Ser and Tyr, or aromatic
 residues Phe and Tyr. Further information regarding making amino acid exchanges
 which have only slight, if any, effects on the overall protein can be found in Bowie
et al., *Science*, 1990, 247:1306-1310, which is incorporated herein by reference in its

WO 02/00860

PCT/US01/20171

entirety including any figures, tables, or drawings. In all cases, all permutations are intended to be covered by this disclosure.

The amino acid sequence of the protease peptide of the invention will be substantially similar to a sequence having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, or the corresponding full-length amino acid sequence, or fragments thereof.

A sequence that is substantially similar to a sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106,

WO 02/00860

PCT/US01/20171

SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 will preferably have at least 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or

5 99% identity to a sequence selected from the group consisting of SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80,

10 SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID

15 NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118. Preferably the protease polypeptide will have at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to one of the aforementioned sequences.

20 By "identity" is meant a property of sequences that measures their similarity or relationship. Identity is measured by dividing the number of identical residues by the total number of residues and gaps and multiplying the product by 100. "Gaps" are spaces in an alignment that are the result of additions or deletions of amino acids. Thus, two copies of exactly the same sequence have 100% identity, but

25 sequences that are less highly conserved, and have deletions, additions, or replacements, may have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity using standard parameters, for example Gapped BLAST or PSI-BLAST (Altschul, *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402), BLAST (Altschul, *et al.*

WO 02/00860

PCT/US01/20171

(1990) *J. Mol. Biol.* 215:403-410), and Smith-Waterman (Smith, *et al.* (1981) *J. Mol. Biol.* 147:195-197). Preferably, the default settings of these programs will be employed, but those skilled in the art recognize whether these settings need to be changed and know how to make the changes.

- 5 “Similarity” is measured by dividing the number of identical residues plus the number of conservatively substituted residues (see Bowie, *et al. Science*, 1999), 247:1306-1310, which is incorporated herein by reference in its entirety, including any drawings, figures, or tables) by the total number of residues and gaps and multiplying the product by 100.
- 10 In preferred embodiments, the invention features isolated, enriched, or purified nucleic acid molecules encoding a protease polypeptide comprising a nucleotide sequence that: (a) encodes a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 and biological domains thereof, (b) is the complement of the nucleotide sequence of (a); or (c) hybridizes under highly stringent conditions to the nucleotide molecule of (a) and encodes a naturally occurring protease polypeptide.
- 25

WO 02/00860

PCT/US01/20171

In preferred embodiments, the invention features isolated, enriched or purified nucleic acid molecules comprising a nucleotide sequence substantially identical to a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, 5 SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, 10 SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, 15 SEQ ID NO:58, and SEQ ID NO:59. Preferably the sequence has at least 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to the above listed sequences.

The term "complement" refers to two nucleotides that can form multiple favorable interactions with one another. For example, adenine is complementary to 20 thymine as they can form two hydrogen bonds. Similarly, guanine and cytosine are complementary since they can form three hydrogen bonds. A nucleotide sequence is the complement of another nucleotide sequence if all of the nucleotides of the first sequence are complementary to all of the nucleotides of the second sequence.

Various low or high stringency hybridization conditions may be used 25 depending upon the specificity and selectivity desired. These conditions are well known to those skilled in the art. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having more than 1 or 2 mismatches out of 20 contiguous nucleotides, more preferably, such conditions

WO 02/00860

PCT/US01/20171

prevent hybridization of nucleic acids having more than 1 or 2 mismatches out of 50 contiguous nucleotides, most preferably, such conditions prevent hybridization of nucleic acids having more than 1 or 2 mismatches out of 100 contiguous nucleotides. In some instances, the conditions may prevent hybridization of nucleic acids having more than 5 mismatches in the full-length sequence.

By stringent hybridization assay conditions is meant hybridization assay conditions at least as stringent as the following: hybridization in 50% formamide, 5X SSC, 50 mM NaH₂PO₄, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5X Denhardt's solution at 42 °C overnight; washing with 2X SSC, 0.1% SDS at 45 °C; and washing with 0.2X SSC, 0.1% SDS at 45 °C. Under some of the most stringent hybridization assay conditions, the second wash can be done with 0.1X SSC at a temperature up to 70 °C (Berger *et al.* (1987) Guide to Molecular Cloning Techniques pg 421, hereby incorporated by reference herein in its entirety including any figures, tables, or drawings.). However, other applications may require the use of conditions falling between these sets of conditions. Methods of determining the conditions required to achieve desired hybridizations are well known to those with ordinary skill in the art, and are based on several factors, including but not limited to, the sequences to be hybridized and the samples to be tested. Washing conditions of lower stringency frequently utilize a lower temperature during the washing steps, such as 65 °C, 60 °C, 55 °C, 50 °C, or 42 °C.

The term "activity" means that the polypeptide hydrolyzes peptide bonds.

The term "catalytic activity", as used herein, defines the rate at which a protease catalytic domain cleaves a substrate. Catalytic activity can be measured, for example, by determining the amount of a substrate cleaved as a function of time. Catalytic activity can be measured by methods of the invention by holding time constant and determining the concentration of a cleaved substrate after a fixed period of time. Cleavage of a substrate occurs at the active site of the protease. The active site is normally a cavity in which the substrate binds to the protease and is cleaved.

WO 02/00860

PCT/US01/20171

The term "biological domain" means a domain or region of the protease polypeptide which has catalytic activity or which binds to the substrate of the protease.

5 The term "substrate" as used herein refers to a polypeptide or protein which is cleaved by a protease of the invention. The term "cleaved" refers to the severing of a covalent bond between amino acid residues of the backbone of the polypeptide or protein.

10 The term "insert" as used herein refers to a portion of a protease that is absent from a close homolog. Inserts may or may not be the product alternative splicing of exons. Inserts can be identified by using a Smith-Waterman sequence alignment of the protein sequence against the non-redundant protein database, or by means of a multiple sequence alignment of homologous sequences using the DNASTar program Megalign (Preferably, the default settings of this program will be used, but those skilled in the art will recognize whether these settings need to be changed and know how to make the changes.). Inserts may play a functional role by
15 presenting a new interface for protein-protein interactions, or by interfering with such interactions.

In other preferred embodiments, the invention features isolated, enriched, or purified nucleic acid molecules encoding protease polypeptides, further comprising
20 a vector or promoter operably linked to the nucleotide sequence. The invention also features recombinant nucleic acid, preferably in a cell or an organism. The recombinant nucleic acid may contain a sequence selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ
25 ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ

WO 02/00860

PCT/US01/20171

ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ ID NO:59, or
5 a functional derivative thereof and a vector or a promoter operably linked to the nucleotide sequence. The recombinant nucleic acid can alternatively contain a transcriptional initiation region functional in a cell, a sequence complementary to an RNA sequence encoding a protease polypeptide and a transcriptional termination
10 region functional in a cell. Specific vectors and host cell combinations are discussed herein.

The term "vector" relates to a single or double-stranded circular nucleic acid molecule that can be transfected into cells and replicated within or independently of a cell genome. A circular double-stranded nucleic acid molecule can be cut and
15 thereby linearized upon treatment with restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a protease can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together.

An operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression may vary from organism to organism, but shall in general
20 include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which,
25 when transcribed into RNA, will signal synthesis initiation.

The term "transfecting" defines a number of methods to insert a nucleic acid vector or other nucleic acid molecules into a cellular organism. These methods involve a variety of techniques, such as treating the cells with high concentrations of

WO 02/00860

PCT/US01/20171

salt, an electric field, detergent, or DMSO to render the outer membrane or wall of the cells permeable to nucleic acid molecules of interest or use of various viral transduction strategies.

The term "promoter" as used herein, refers to nucleic acid sequence needed for gene sequence expression. Promoter regions vary from organism to organism, but are well known to persons skilled in the art for different organisms. For example, in prokaryotes, the promoter region contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

In preferred embodiments, the isolated nucleic acid comprises, consists essentially of, or consists of a nucleic acid sequence selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ ID NO:59 which encodes an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID

WO 02/00860

PCT/US01/20171

NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, a functional derivative thereof, or at least 35, 40, 45, 50, 60, 75, 100, 200, or 300 contiguous amino acids selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118. The nucleic acid may be isolated from a natural source by cDNA cloning or by subtractive hybridization. The natural source may be mammalian, preferably human, blood, semen, or tissue, and the nucleic acid may be synthesized by the triester method or by using an automated DNA synthesizer.

WO 02/00860

PCT/US01/20171

The term "mammal" refers preferably to such organisms as mice, rats, rabbits, guinea pigs, sheep, and goats, more preferably to cats, dogs, monkeys, and apes, and most preferably to humans.

In yet other preferred embodiments, the nucleic acid is a conserved or unique
 5 region, for example those useful for: the design of hybridization probes to facilitate identification and cloning of additional polypeptides, the design of PCR probes to facilitate cloning of additional polypeptides, obtaining antibodies to polypeptide regions, and designing antisense oligonucleotides.

By "conserved nucleic acid regions", are meant regions present on two or
 10 more nucleic acids encoding a protease polypeptide, to which a particular nucleic acid sequence can hybridize under lower stringency conditions. Examples of lower stringency conditions suitable for screening for nucleic acid encoding protease polypeptides are provided in Wahl *et al. Meth. Enzym.* 152:399-407 (1987) and in Wahl *et al. Meth. Enzym.* 152:415-423 (1987), which are hereby incorporated by
 15 reference herein in its entirety, including any drawings, figures, or tables. Preferably, conserved regions differ by no more than 5 out of 20 nucleotides, even more preferably 2 out of 20 nucleotides or most preferably 1 out of 20 nucleotides.

By "unique nucleic acid region" is meant a sequence present in a nucleic acid coding for a protease polypeptide that is not present in a sequence coding for any
 20 other naturally occurring polypeptide. Such regions preferably encode 32 (preferably 40, more preferably 45, most preferably 55) or more contiguous amino acids set forth in a full-length amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67,
 25 SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92,

WO 02/00860

PCT/US01/20171

SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97,
 SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID
 NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106,
 SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID
 5 NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115,
 SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 in a sample. The nucleic
 acid probe contains a nucleotide base sequence that will hybridize to the sequence
 selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID
 NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7,
 10 SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12,
 SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17,
 SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22,
 SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27,
 SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32,
 15 SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37,
 SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42,
 SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47,
 SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52,
 SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57,
 20 SEQ ID NO:58, and SEQ ID NO:59, or a functional derivative thereof.

In preferred embodiments, the nucleic acid probe hybridizes to nucleic acid
 encoding at least 12, 32, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous
 amino acids of a full-length sequence selected from the group consisting of those set
 forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID
 25 NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID
 NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID
 NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID
 NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID
 NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID

WO 02/00860

PCT/US01/20171

NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, 5 SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, or a functional derivative thereof.

Methods for using the probes include detecting the presence or amount of protease RNA in a sample by contacting the sample with a nucleic acid probe under 10 conditions such that hybridization occurs and detecting the presence or amount of the probe bound to protease RNA. The nucleic acid duplex formed between the probe and a nucleic acid sequence coding for a protease polypeptide may be used in the identification of the sequence of the nucleic acid detected (Nelson *et al.*, in Nonisotopic DNA Probe Techniques, Academic Press, San Diego, Kricka, ed., p. 15 275, 1992, hereby incorporated by reference herein in its entirety, including any drawings, figures, or tables). Kits for performing such methods may be constructed to include a container means having disposed therein a nucleic acid probe.

Methods for using the probes also include using these probes to find the full-length clone of each of the predicted proteases by techniques known to one skilled in 20 the art. These clones will be useful for screening for small molecule compounds that inhibit the catalytic activity of the encoded protease with potential utility in treating cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically disorders including cancers of tissues, blood, or hematopoietic origin, particularly those 25 involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental

WO 02/00860

PCT/US01/20171

retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, multiple sclerosis, and amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

In another aspect, the invention describes a recombinant cell or tissue comprising a nucleic acid molecule encoding a protease polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118. In such cells, the nucleic acid may be under the control of the genomic regulatory elements, or may be under the control of exogenous regulatory elements including an exogenous promoter. By "exogenous" it is meant a promoter that is not

WO 02/00860

PCT/US01/20171

normally coupled *in vivo* transcriptionally to the coding sequence for the protease polypeptides.

The polypeptide is preferably a fragment of the protein encoded by a full-length amino acid sequence selected from the group consisting of those set forth in

5 SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84,

10 SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108,

15 SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118. By "fragment," is meant an amino acid sequence present in a protease polypeptide. Preferably, such a sequence comprises at least 32, 45, 50, 60, 100, 200, or 300 contiguous amino acids of a full-length sequence selected from the

20 group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID

25 NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106,

WO 02/00860

PCT/US01/20171

SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.

In another aspect, the invention features an isolated, enriched, or purified
 5 protease polypeptide having a sequence substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75,
 10 SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100,
 15 SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 and biological domains thereof. Preferable the polypeptide sequence has at least
 20 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to the above listed sequences.

By "isolated" in reference to a polypeptide is meant a polymer of 6 (preferably 12, more preferably 18, most preferably 25, 32, 40, or 50) or more amino acids conjugated to each other, including polypeptides that are isolated from a
 25 natural source or that are synthesized. In certain aspects longer polypeptides are preferred, such as those with 100, 200, 300, 400, 450, 500, 550, 600, 700, 800, 900 or more contiguous amino acids of a full-length sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67,

WO 02/00860

PCT/US01/20171

SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72,
 SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77,
 SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82,
 SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87,
 5 SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92,
 SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97,
 SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID
 NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106,
 SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID
 10 NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115,
 SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, and/or those polypeptides
 having at least 50%, 60%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,
 97%, 98% or 99% identity to a sequence selected from the group consisting of SEQ
 ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ
 15 ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ
 ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ
 ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ
 ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ
 ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ
 20 ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ
 ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ
 ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104,
 SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID
 NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113,
 25 SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID
 NO:118.

The isolated polypeptides of the present invention are unique in the sense
 that they are not found in a pure or separated state in nature. Use of the term
 "isolated" indicates that a naturally occurring sequence has been removed from its

normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only amino acid chain present, but that it is essentially free (at least about 90% pure, more preferably at least about 95% pure or more) of non-amino acid-based material naturally associated with it.

By the use of the term "enriched" in reference to a polypeptide is meant that the specific amino acid sequence constitutes a significantly higher fraction (2- to 5-fold) of the total amino acid sequences present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other amino acid sequences present, or by a preferential increase in the amount of the specific amino acid sequence of interest, or by a combination of the two. However, it should be noted that enriched does not imply that there are no other amino acid sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term significant here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other amino acid sequences of about at least 2-fold, more preferably at least 5- to 10-fold or even more. The term also does not imply that there is no amino acid sequence from other sources. The other source of amino acid sequences may, for example, comprise amino acid sequence encoded by a yeast or bacterial genome, or a cloning vector such as pUC19. The term is meant to cover only those situations in which man has intervened to increase the proportion of the desired amino acid sequence.

It is also advantageous for some purposes that an amino acid sequence be in purified form. The term "purified" in reference to a polypeptide does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment. Compared to the natural level this level should be at least 2-to 5-fold greater (e.g., in terms of mg/mL). Purification of at least one order of magnitude, preferably two or

WO 02/00860

PCT/US01/20171

three orders, and more preferably four or five orders of magnitude is expressly contemplated. The substance is preferably free of contamination at a functionally significant level, for example 90%, 95%, or 99% pure.

In preferred embodiments, the protease polypeptide is a fragment of the protein encoded by a full-length amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118. Preferably, the protease polypeptide contains at least 32, 45, 50, 60, 100, 200, or 300 contiguous amino acids of a full-length sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID

WO 02/00860

PCT/US01/20171

NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, or a functional derivative thereof.

5 In preferred embodiments, the protease polypeptide comprises an amino acid sequence having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.

20 The polypeptide can be isolated from a natural source by methods well-known in the art. The natural source may be mammalian, preferably human, blood, semen, or tissue, and the polypeptide may be synthesized using an automated polypeptide synthesizer.

In some embodiments the invention includes a recombinant protease polypeptide having (a) an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78,

WO 02/00860

PCT/US01/20171

SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118. By "recombinant protease polypeptide" is meant a polypeptide produced by recombinant DNA techniques such that it is distinct from a naturally occurring polypeptide either in its location (*e.g.*, present in a different cell or tissue than found in nature), purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

The polypeptides to be expressed in host cells may also be fusion proteins which include regions from heterologous proteins. Such regions may be included to allow, *e.g.*, secretion, improved stability, or facilitated purification of the polypeptide. For example, a sequence encoding an appropriate signal peptide can be incorporated into expression vectors. A DNA sequence for a signal peptide (secretory leader) may be fused in-frame to the polynucleotide sequence so that the polypeptide is translated as a fusion protein comprising the signal peptide. A signal peptide that is functional in the intended host cell promotes extracellular secretion of the polypeptide. Preferably, the signal sequence will be cleaved from the polypeptide upon secretion of the polypeptide from the cell. Thus, preferred fusion proteins can be produced in which the N-terminus of a protease polypeptide is fused to a carrier peptide.

In one embodiment, the polypeptide comprises a fusion protein which includes a heterologous region used to facilitate purification of the polypeptide. Many of the available peptides used for such a function allow selective binding of

WO 02/00860

PCT/US01/20171

the fusion protein to a binding partner. A preferred binding partner includes one or more of the IgG binding domains of protein A are easily purified to homogeneity by affinity chromatography on, for example, IgG-coupled Sepharose. Alternatively, many vectors have the advantage of carrying a stretch of histidine residues that can
 5 be expressed at the N-terminal or C-terminal end of the target protein, and thus the protein of interest can be recovered by metal chelation chromatography. A nucleotide sequence encoding a recognition site for a proteolytic enzyme such as enterokinase, factor X procollagenase or thrombin may immediately precede the sequence for a protease polypeptide to permit cleavage of the fusion protein to
 10 obtain the mature protease polypeptide. Additional examples of fusion-protein binding partners include, but are not limited to, the yeast I-factor, the honeybee melatin leader in sf9 insect cells, 6-His tag, thioredoxin tag, hemagglutinin tag, GST tag, and OmpA signal sequence tag. As will be understood by one of skill in the art, the binding partner which recognizes and binds to the peptide may be any ion,
 15 molecule or compound including metal ions (*e.g.*, metal affinity columns), antibodies, or fragments thereof, and any protein or peptide which binds the peptide, such as the FLAG tag.

Antibodies

20 In another aspect, the invention features an antibody (*e.g.*, a monoclonal or polyclonal antibody) having specific binding affinity to a protease polypeptide or a protease polypeptide domain or fragment where the polypeptide is selected from the group having a sequence at least about 90% identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID
 25 NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID

WO 02/00860

PCT/US01/20171

- NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105,
- 5 SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.
- Preferably the polypeptide is has at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 99% or 100% identity with the sequences listed above. By
- 10 "specific binding affinity" is meant that the antibody binds to the target protease polypeptide with greater affinity than it binds to other polypeptides under specified conditions. Antibodies or antibody fragments are polypeptides that contain regions that can bind other polypeptides. The term "specific binding affinity" describes an antibody that binds to a protease polypeptide with greater affinity than it binds to
- 15 other polypeptides under specified conditions. Antibodies can be used to identify an endogenous source of protease polypeptides, to monitor cell cycle regulation, and for immuno-localization of protease polypeptides within the cell.
- The term "polyclonal" refers to antibodies that are heterogenous populations of antibody molecules derived from the sera of animals immunized with an antigen
- 20 or an antigenic functional derivative thereof. For the production of polyclonal antibodies, various host animals may be immunized by injection with the antigen. Various adjuvants may be used to increase the immunological response, depending on the host species.
- "Monoclonal antibodies" are substantially homogenous populations of
- 25 antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. Monoclonal antibodies may be obtained by methods known to those skilled in the art (Kohler *et al.*, *Nature*, 1975, 256:495-497, and U.S. Patent No. 4,376,110,

both of which are hereby incorporated by reference herein in their entirety including any figures, tables, or drawings).

An antibody of the present invention includes "humanized" monoclonal and polyclonal antibodies. Humanized antibodies are recombinant proteins in which
5 non-human (typically murine) complementarity determining regions of an antibody have been transferred from heavy and light variable chains of the non-human (*e.g.* murine) immunoglobulin into a human variable domain, followed by the replacement of some human residues in the framework regions of their murine counterparts. Humanized antibodies in accordance with this invention are suitable
10 for use in therapeutic methods. General techniques for cloning murine immunoglobulin variable domains are described, for example, by the publication of Orlandi *et al.*, *Proc. Nat'l Acad. Sci. USA* 86: 3833 (1989). Techniques for producing humanized monoclonal antibodies are described, for example, by Jones *et al.*, *Nature* 321:522 (1986), Riechmann *et al.*, *Nature* 332:323 (1988), Verhoeven *et al.*, *Science* 239:1534 (1988), Carter *et al.*, *Proc. Nat'l Acad. Sci. USA* 89:4285
15 (1992), Sandhu, *Crit. Rev. Biotech.* 12:437 (1992), and Singer *et al.*, *J. Immun.* 150:2844 (1993).

The term "antibody fragment" refers to a portion of an antibody, often the hypervariable region and portions of the surrounding heavy and light chains, that
20 displays specific binding affinity for a particular molecule. A hypervariable region is a portion of an antibody that physically binds to the polypeptide target.

An antibody fragment of the present invention includes a "single-chain antibody," a phrase used in this description to denote a linear polypeptide that binds
25 antigen with specificity and that comprises variable or hypervariable regions from the heavy and light chain chains of an antibody. Such single chain antibodies can be produced by conventional methodology. The Vh and Vl regions of the Fv fragment can be covalently joined and stabilized by the insertion of a disulfide bond. See Glockshuber, *et al.*, *Biochemistry* 1362 (1990). Alternatively, the Vh and Vl regions can be joined by the insertion of a peptide linker. A gene encoding the Vh, Vl and

WO 02/00860

PCT/US01/20171

peptide linker sequences can be constructed and expressed using a recombinant expression vector. See Colcher, *et al.*, *J. Nat'l Cancer Inst.* 82: 1191 (1990). Amino acid sequences comprising hypervariable regions from the Vh and Vl antibody chains can also be constructed using disulfide bonds or peptide linkers.

5 Antibodies or antibody fragments having specific binding affinity to a protease polypeptide of the invention may be used in methods for detecting the presence and/or amount of protease polypeptide in a sample by probing the sample with the antibody under conditions suitable for protease-antibody immunocomplex formation and detecting the presence and/or amount of the antibody conjugated to
10 the protease polypeptide. Diagnostic kits for performing such methods may be constructed to include antibodies or antibody fragments specific for the protease as well as a conjugate of a binding partner of the antibodies or the antibodies themselves.

15 An antibody or antibody fragment with specific binding affinity to a protease polypeptide of the invention can be isolated, enriched, or purified from a prokaryotic or eukaryotic organism. Routine methods known to those skilled in the art enable production of antibodies or antibody fragments, in both prokaryotic and eukaryotic organisms. Purification, enrichment, and isolation of antibodies, which are polypeptide molecules, are described above.

20 Antibodies having specific binding affinity to a protease polypeptide of the invention may be used in methods for detecting the presence and/or amount of protease polypeptide in a sample by contacting the sample with the antibody under conditions such that an immunocomplex forms and detecting the presence and/or amount of the antibody conjugated to the protease polypeptide. Diagnostic kits for
25 performing such methods may be constructed to include a first container containing the antibody and a second container having a conjugate of a binding partner of the antibody and a label, such as, for example, a radioisotope. The diagnostic kit may also include notification of an FDA approved use and instructions therefor.

WO 02/00860

PCT/US01/20171

In another aspect, the invention features a hybridoma which produces an antibody having specific binding affinity to a protease polypeptide or a protease polypeptide domain, where the polypeptide is selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, 5 SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, 10 SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, 15 SEQ ID NO:117 and SEQ ID NO:118. By "hybridoma" is meant an immortalized cell line that is capable of secreting an antibody, for example an antibody to a protease of the invention. In preferred embodiments, the antibody to the protease comprises a sequence of amino acids that is able to specifically bind a protease 20 polypeptide of the invention.

In another aspect, the present invention is also directed to kits comprising antibodies that bind to a polypeptide encoded by any of the nucleic acid molecules described above, and a negative control antibody.

The term "negative control antibody" refers to an antibody derived from 25 similar source as the antibody having specific binding affinity, but where it displays no binding affinity to a polypeptide of the invention.

In another aspect, the invention features a protease polypeptide binding agent able to bind to a protease polypeptide selected from the group having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60,

WO 02/00860

PCT/US01/20171

SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65,
SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70,
SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75,
SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80,
5 SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85,
SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90,
SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95,
SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100,
SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID
10 NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109,
SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID
NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.
The binding agent is preferably a purified antibody that recognizes an epitope
present on a protease polypeptide of the invention. Other binding agents include
15 molecules that bind to protease polypeptides and analogous molecules that bind to a
protease polypeptide. Such binding agents may be identified by using assays that
measure protease binding partner activity, or they may be identified using assays
that measure protease activity, such as the release of a fluorogenic or radioactive
marker attached to a substrate molecule.

20

Screening Methods to Detect Protease Polypeptides

The invention also features a method for screening for human cells
containing a protease polypeptide of the invention or an equivalent sequence. The
method involves identifying the novel polypeptide in human cells using techniques
25 that are routine and standard in the art, such as those described herein for identifying
the proteases of the invention (*e.g.*, cloning, Southern or Northern blot analysis, *in
situ* hybridization, PCR amplification, etc.).

Screening Methods to Identify Substances that Modulate Protease**Activity**

In another aspect, the invention features methods for identifying a substance that modulates protease activity comprising the steps of: (a) contacting a protease

5 polypeptide comprising an amino acid sequence substantially identical to a sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID

10 NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID

15 NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 with a test substance; (b) measuring the activity of said polypeptide; and (c) determining

20 whether said substance modulates the activity of said polypeptide. More preferably the sequence is at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to the listed sequences.

The term "modulates" refers to the ability of a compound to alter the function of a protease of the invention. A modulator preferably activates or inhibits

25 the activity of a protease of the invention depending on the concentration of the compound exposed to the protease.

The term "modulates" also refers to altering the function of proteases of the invention by increasing or decreasing the probability that a complex forms between the protease and a natural binding partner. A modulator preferably increases the

WO 02/00860

PCT/US01/20171

probability that such a complex forms between the protease and the natural binding partner, more preferably increases or decreases the probability that a complex forms between the protease and the natural binding partner depending on the concentration of the compound exposed to the protease, and most preferably decreases the

5 probability that a complex forms between the protease and the natural binding partner.

The term "activates" refers to increasing the cellular activity of the protease.

The term "inhibits" refers to decreasing the cellular activity of the protease.

The term "complex" refers to an assembly of at least two molecules bound to
10 one another. Signal transduction complexes often contain at least two protein molecules bound to one another. For instance, a protein tyrosine receptor protein kinase, GRB2, SOS, RAF, and RAS assemble to form a signal transduction complex in response to a mitogenic ligand. Similarly, the proteases involved in blood
15 coagulation and their cofactors are known to form macromolecular complexes on cellular membranes. Additionally, proteases involved in modification of the extracellular matrix are known to form complexes with their inhibitors and also with components of the extracellular matrix.

The term "natural binding partner" refers to polypeptides, lipids, small molecules, or nucleic acids that bind to proteases in cells. A change in the
20 interaction between a protease and a natural binding partner can manifest itself as an increased or decreased probability that the interaction forms, or an increased or decreased concentration of protease/natural binding partner complex.

The term "contacting" as used herein refers to mixing a solution comprising the test compound with a liquid medium bathing the cells of the methods. The
25 solution comprising the compound may also comprise another component, such as dimethyl sulfoxide (DMSO), which facilitates the uptake of the test compound or compounds into the cells of the methods. The solution comprising the test compound may be added to the medium bathing the cells by utilizing a delivery apparatus, such as a pipette-based device or syringe-based device.

WO 02/00860

PCT/US01/20171

In another aspect, the invention features methods for identifying a substance that modulates protease activity in a cell comprising the steps of: (a) expressing a protease polypeptide in a cell, wherein said polypeptide has a sequence substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118; (b) adding a test substance to said cell; and (c) monitoring a change in cell phenotype, cell proliferation, cell differentiation or the interaction between said polypeptide and a natural binding partner.

The term "expressing" as used herein refers to the production of proteases of the invention from a nucleic acid vector containing protease genes within a cell. The nucleic acid vector is transfected into cells using well known techniques in the art as described herein.

Another aspect of the instant invention is directed to methods of identifying compounds that bind to protease polypeptides of the present invention, comprising contacting the protease polypeptides with a compound, and determining whether the compound binds the protease polypeptides. Binding can be determined by binding assays which are well known to the skilled artisan, including, but not limited to, gel-shift assays, Western blots, radiolabeled competition assay, phage-based expression

WO 02/00860

PCT/US01/20171

cloning, co-fractionation by chromatography, co-precipitation, cross linking, interaction trap/two-hybrid analysis, southwestern analysis, ELISA, and the like, which are described in, for example, *Current Protocols in Molecular Biology*, 1999, John Wiley & Sons, NY, which is incorporated herein by reference in its entirety.

- 5 The compounds to be screened include, but are not limited to, compounds of extracellular, intracellular, biological or chemical origin.

The methods of the invention also embrace compounds that are attached to a label, such as a radiolabel (e.g., ^{125}I , ^{35}S , ^{32}P , ^3H), a fluorescence label, a chemiluminescent label, an enzymic label and an immunogenic label. The protease polypeptides employed in such a test may either be free in solution, attached to a solid support, borne on a cell surface, located intracellularly or associated with a portion of a cell. One skilled in the art can, for example, measure the formation of complexes between a protease polypeptide and the compound being tested. Alternatively, one skilled in the art can examine the diminution in complex formation between a protease polypeptide and its substrate caused by the compound being tested.

Other assays can be used to examine enzymatic activity including, but not limited to, photometric, radiometric, HPLC, electrochemical, and the like, which are described in, for example, *Enzyme Assays: A Practical Approach*, eds. R. Eisinger and M. J. Danson, 1992, Oxford University Press, which is incorporated herein by reference in its entirety.

Another aspect of the present invention is directed to methods of identifying compounds which modulate (i.e., increase or decrease) activity of a protease polypeptide comprising contacting the protease polypeptide with a compound, and determining whether the compound modifies activity of the protease polypeptide. These compounds are also referred to as "modulators of proteases." The activity in the presence of the test compound is measured to the activity in the absence of the test compound. Where the activity of a sample containing the test compound is higher than the activity in a sample lacking the test compound, the compound will

WO 02/00860

PCT/US01/20171

have increased the activity. Similarly, where the activity of a sample containing the test compound is lower than the activity in the sample lacking the test compound, the compound will have inhibited the activity.

The present invention is particularly useful for screening compounds by using a protease polypeptide in any of a variety of drug screening techniques. The compounds to be screened include, but are not limited to, extracellular, intracellular, biological or chemical origin. The protease polypeptide employed in such a test may be in any form, preferably, free in solution, attached to a solid support, borne on a cell surface or located intracellularly. One skilled in the art can measure the change in rate that a protease of the invention cleaves a substrate polypeptide. One skilled in the art can also, for example, measure the formation of complexes between a protease polypeptide and the compound being tested. Alternatively, one skilled in the art can examine the diminution in complex formation between a protease polypeptide and its substrate caused by the compound being tested.

The activity of protease polypeptides of the invention can be determined by, for example, examining the ability to bind or be activated by chemically synthesised peptide ligands. Alternatively, the activity of the protease polypeptides can be assayed by examining their ability to bind metal ions such as calcium, hormones, chemokines, neuropeptides, neurotransmitters, nucleotides, lipids, odorants, and photons. Thus, modulators of the protease polypeptide's activity may alter a protease function, such as a binding property of a protease or an activity such as cleaving protein substrates or polypeptide substrates, or membrane localization.

In various embodiments of the method, the assay may take the form of a yeast growth assay, an Aequorin assay, a Luciferase assay, a mitogenesis assay, a MAP Kinase activity assay, as well as other binding or function-based assays of protease activity that are generally known in the art. In several of these embodiments, the invention includes any of the serine proteases, cysteine proteases, aspartyl proteases, metalloproteases, threonine proteases, and other proteases. Biological activities of proteases according to the invention include, but are not

limited to, the binding of a natural or a synthetic ligand, as well as any one of the functional activities of proteases known in the art. Non-limiting examples of protease activities include cleavage of polypeptide chains, processing the pro-form of a polypeptide chain to the active product, transmembrane signaling of various

5 forms, and/or the modification of the extracellular matrix.

The modulators of the invention exhibit a variety of chemical structures, which can be generally grouped into mimetics of natural protease ligands, and peptide and non-peptide allosteric effectors of proteases. The invention does not restrict the sources for suitable modulators, which may be obtained from natural

10 sources such as plant, animal or mineral extracts, or non-natural sources such as small molecule libraries, including the products of combinatorial chemical approaches to library construction, and peptide libraries.

The use of cDNAs encoding proteins in drug discovery programs is well-known; assays capable of testing thousands of unknown compounds per day in high-throughput screens (HTSs) are thoroughly documented. The literature is replete with examples of the use of radiolabelled ligands in HTS binding assays for drug discovery (*see, Williams, Medicinal Research Reviews*, 1991, 11:147-184.; Sweetnam, *et al., J. Natural Products*, 1993, 56:441-455 for review). Recombinant proteins are preferred for binding assay HTS because they allow for better

20 specificity (higher relative purity), provide the ability to generate large amounts of receptor material, and can be used in a broad variety of formats (see Hodgson, *Bio/Technology*, 1992, 10:973-980 which is incorporated herein by reference in its entirety). A variety of heterologous systems is available for functional expression of recombinant proteins that are well known to those skilled in the art. Such systems

25 include bacteria (Strosberg, *et al., Trends in Pharmacological Sciences*, 1992, 13:95-98), yeast (Pausch, *Trends in Biotechnology*, 1997, 15:487-494), several kinds of insect cells (Vanden Broeck, *Int. Rev. Cytology*, 1996, 164:189-268), amphibian cells (Jayawickreme *et al., Current Opinion in Biotechnology*, 1997, 8:629-634) and several mammalian cell lines (CHO, HEK293, COS, etc.; *see, Gerhardt, et al., Eur.*

J. Pharmacology, 1997, 334:1-23). These examples do not preclude the use of other possible cell expression systems, including cell lines obtained from nematodes (PCT application WO 98/37177).

An expressed protease can be used for HTS binding assays in conjunction
5 with its defined ligand, in this case the corresponding peptide that activates it. The identified peptide is labeled with a suitable radioisotope, including, but not limited to, ^{125}I , ^3H , ^{35}S or ^{32}P , by methods that are well known to those skilled in the art. Alternatively, the peptides may be labeled by well-known methods with a suitable fluorescent derivative (Baindur, *et al.*, *Drug Dev. Res.*, 1994, 33:373-398; Rogers,
10 *Drug Discovery Today*, 1997, 2:156-160). Radioactive ligand specifically bound to the receptor in membrane preparations made from the cell line expressing the recombinant protein can be detected in HTS assays in one of several standard ways, including filtration of the receptor-ligand complex to separate bound ligand from unbound ligand (Williams, *Med. Res. Rev.*, 1991, 11:147-184.; Sweetnam, *et al.*, *J.*
15 *Natural Products*, 1993, 56:441-455). Alternative methods include a scintillation proximity assay (SPA) or a FlashPlate format in which such separation is unnecessary (Nakayama, *Cur. Opinion Drug Disc. Dev.*, 1998, 1:85-91 Bossé, *et al.*, *J. Biomolecular Screening*, 1998, 3:285-292.). Binding of fluorescent ligands can be detected in various ways, including fluorescence energy transfer (FRET), direct
20 spectrophotofluorometric analysis of bound ligand, or fluorescence polarization (Rogers, *Drug Discovery Today*, 1997, 2:156-160; Hill, *Cur. Opinion Drug Disc. Dev.*, 1998, 1:92-97).

The proteases and natural binding partners required for functional expression of heterologous protease polypeptides can be native constituents of the host cell or
25 can be introduced through well-known recombinant technology. The protease polypeptides can be intact or chimeric. The protease activation may result in the stimulation or inhibition of other native proteins, events that can be linked to a measurable response.

WO 02/00860

PCT/US01/20171

Examples of such biological responses include, but are not limited to, the following: the ability to survive in the absence of a limiting nutrient in specifically engineered yeast cells (Pausch, *Trends in Biotechnology*, 1997, 15:487-494); changes in intracellular Ca^{2+} concentration as measured by fluorescent dyes
5 (Murphy, *et al.*, *Cur. Opinion Drug Disc. Dev.*, 1998, 1:192-199). Fluorescence changes can also be used to monitor ligand-induced changes in membrane potential or intracellular pH; an automated system suitable for HTS has been described for these purposes (Schroeder, *et al.*, *J. Biomolecular Screening*, 1996, 1:75-80). Assays are also available for the measurement of common second but these are not
10 generally preferred for HTS.

The invention contemplates a multitude of assays to screen and identify inhibitors of ligand binding to protease polypeptides or of substrate cleavage by protease polypeptides. In one example, the protease polypeptide is immobilized and interaction with a binding partner or substrate is assessed in the presence and
15 absence of a candidate modulator such as an inhibitor compound. In another example, interaction between the protease polypeptide and its binding partner or a substrate is assessed in a solution assay, both in the presence and absence of a candidate inhibitor compound. In either assay, an inhibitor is identified as a compound that decreases binding between the protease polypeptide and its natural
20 binding partner or the activity of a protease polypeptide in cleaving a substrate molecule. Another contemplated assay involves a variation of the di-hybrid assay wherein an inhibitor of protein/protein interactions is identified by detection of a positive signal in a transformed or transfected host cell, as described in PCT publication number WO 95/20652, published August 3, 1995 and is included by
25 reference herein including any figures, tables, or drawings.

Candidate modulators contemplated by the invention include compounds selected from libraries of either potential activators or potential inhibitors. There are a number of different libraries used for the identification of small molecule modulators, including: (1) chemical libraries, (2) natural product libraries, and (3)

WO 02/00860

PCT/US01/20171

combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules. Chemical libraries consist of random chemical structures, some of which are analogs of known compounds or analogs of compounds that have been identified as "hits" or "leads" in other drug discovery screens, while others are derived from natural products, and still others arise from non-directed synthetic organic chemistry. Natural product libraries are collections of microorganisms, animals, plants, or marine organisms which are used to create mixtures for screening by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of plants or marine organisms. Natural product libraries include polyketides, non-ribosomal peptides, and variants (non-naturally occurring) thereof. For a review, *see, Science* 282:63-68 (1998). Combinatorial libraries are composed of large numbers of peptides, oligonucleotides, or organic compounds as a mixture. These libraries are relatively easy to prepare by traditional automated synthesis methods, PCR, cloning, or proprietary synthetic methods. Of particular interest are non-peptide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, *see, Myers, Curr. Opin. Biotechnol.* 8:701-707 (1997). Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to modulate activity.

Still other candidate inhibitors contemplated by the invention can be designed and include soluble forms of binding partners, as well as such binding partners as chimeric, or fusion, proteins. A "binding partner" as used herein broadly encompasses both natural binding partners as described above as well as chimeric polypeptides, peptide modulators other than natural ligands, antibodies, antibody fragments, and modified compounds comprising antibody domains that are immunospecific for the expression product of the identified protease gene.

Other assays may be used to identify specific peptide ligands of a protease polypeptide, including assays that identify ligands of the target protein through measuring direct binding of test ligands to the target protein, as well as assays that identify ligands of target proteins through affinity ultrafiltration with ion spray mass spectroscopy/HPLC methods or other physical and analytical methods.

Alternatively, such binding interactions are evaluated indirectly using the yeast two-hybrid system described in Fields *et al.*, *Nature*, 340:245-246 (1989), and Fields *et al.*, *Trends in Genetics*, 10:286-292 (1994), both of which are incorporated herein by reference. The two-hybrid system is a genetic assay for detecting interactions between two proteins or polypeptides. It can be used to identify proteins that bind to a known protein of interest, or to delineate domains or residues critical for an interaction. Variations on this methodology have been developed to clone genes that encode DNA binding proteins, to identify peptides that bind to a protein, and to screen for drugs. The two-hybrid system exploits the ability of a pair of interacting proteins to bring a transcription activation domain into close proximity with a DNA binding domain that binds to an upstream activation sequence (UAS) of a reporter gene, and is generally performed in yeast. The assay requires the construction of two hybrid genes encoding (1) a DNA-binding domain that is fused to a first protein and (2) an activation domain fused to a second protein. The DNA-binding domain targets the first hybrid protein to the UAS of the reporter gene; however, because most proteins lack an activation domain, this DNA-binding hybrid protein does not activate transcription of the reporter gene. The second hybrid protein, which contains the activation domain, cannot by itself activate expression of the reporter gene because it does not bind the UAS. However, when both hybrid proteins are present, the noncovalent interaction of the first and second proteins tethers the activation domain to the UAS, activating transcription of the reporter gene. For example, when the first protein is a protease gene product, or fragment thereof, that is known to interact with another protein or nucleic acid, this assay can be used to detect agents that interfere with the binding interaction. Expression of the reporter

WO 02/00860

PCT/US01/20171

gene is monitored as different test agents are added to the system. The presence of an inhibitory agent results in lack of a reporter signal.

When the function of the protease polypeptide gene product is unknown and no ligands are known to bind the gene product, the yeast two-hybrid assay can also be used to identify proteins that bind to the gene product. In an assay to identify proteins that bind to a protease polypeptide, or fragment thereof, a fusion polynucleotide encoding both a protease polypeptide (or fragment) and a UAS binding domain (*i.e.*, a first protein) may be used. In addition, a large number of hybrid genes each encoding a different second protein fused to an activation domain are produced and screened in the assay. Typically, the second protein is encoded by one or more members of a total cDNA or genomic DNA fusion library, with each second protein coding region being fused to the activation domain. This system is applicable to a wide variety of proteins, and it is not even necessary to know the identity or function of the second binding protein. The system is highly sensitive and can detect interactions not revealed by other methods; even transient interactions may trigger transcription to produce a stable mRNA that can be repeatedly translated to yield the reporter protein.

Other assays may be used to search for agents that bind to the target protein. One such screening method to identify direct binding of test ligands to a target protein is described in U.S. Patent No. 5,585,277, incorporated herein by reference. This method relies on the principle that proteins generally exist as a mixture of folded and unfolded states, and continually alternate between the two states. When a test ligand binds to the folded form of a target protein (*i.e.*, when the test ligand is a ligand of the target protein), the target protein molecule bound by the ligand remains in its folded state. Thus, the folded target protein is present to a greater extent in the presence of a test ligand which binds the target protein, than in the absence of a ligand. Binding of the ligand to the target protein can be determined by any method which distinguishes between the folded and unfolded states of the target protein. The function of the target protein need not be known in order for this assay to be

WO 02/00860

PCT/US01/20171

performed. Virtually any agent can be assessed by this method as a test ligand, including, but not limited to, metals, polypeptides, proteins, lipids, polysaccharides, polynucleotides and small organic molecules.

Another method for identifying ligands of a target protein is described in
5 Wieboldt *et al.*, *Anal. Chem.*, 69:1683-1691 (1997), incorporated herein by reference. This technique screens combinatorial libraries of 20-30 agents at a time in solution phase for binding to the target protein. Agents that bind to the target protein are separated from other library components by simple membrane washing. The specifically selected molecules that are retained on the filter are subsequently
10 liberated from the target protein and analyzed by HPLC and pneumatically assisted electrospray (ion spray) ionization mass spectroscopy. This procedure selects library components with the greatest affinity for the target protein, and is particularly useful for small molecule libraries.

In preferred embodiments of the invention, methods of screening for
15 compounds which modulate protease activity comprise contacting test compounds with protease polypeptides and assaying for the presence of a complex between the compound and the protease polypeptide. In such assays, the ligand is typically labelled. After suitable incubation, free ligand is separated from that present in bound form, and the amount of free or uncomplexed label is a measure of the ability
20 of the particular compound to bind to the protease polypeptide.

In another embodiment of the invention, high throughput screening for compounds having suitable binding affinity to protease polypeptides is employed. Briefly, large numbers of different small peptide test compounds are synthesised on a solid substrate. The peptide test compounds are contacted with the protease
25 polypeptide and washed. Bound protease polypeptide is then detected by methods well known in the art. Purified polypeptides of the invention can also be coated directly onto plates for use in the aforementioned drug screening techniques. In addition, non-neutralizing antibodies can be used to capture the protein and immobilize it on the solid support.

WO 02/00860

PCT/US01/20171

Other embodiments of the invention comprise using competitive screening assays in which neutralizing antibodies capable of binding a polypeptide of the invention specifically compete with a test compound for binding to the polypeptide. In this manner, the antibodies can be used to detect the presence of any peptide that

5 shares one or more antigenic determinants with a protease polypeptide.

Radiolabeled competitive binding studies are described in A.H. Lin *et al.*

Antimicrobial Agents and Chemotherapy, 1997, vol. 41, no. 10, pp. 2127-2131, the disclosure of which is incorporated herein by reference in its entirety.

Therapeutic Methods

10 The invention includes methods for treating a disease or disorder by administering to a patient in need of such treatment a protease polypeptide substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, and any other protease polypeptide of the present invention. As discussed in the section "Gene Therapy," a protease polypeptide of the invention may also be administered indirectly by via administration of suitable polynucleotide means for *in vivo* expression of the protease polypeptide. Preferably the protease polypeptide will have at least about

25

WO 02/00860

PCT/US01/20171

90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to one of the aforementioned sequences.

In another aspect, the invention provides methods for treating a disease or disorder by administering to a patient in need of such treatment a substance that modulates the activity of a protease substantially identical to a sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.

Preferably the disease is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis,

WO 02/00860

PCT/US01/20171

- and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.
- 10 In preferred embodiments, the invention provides methods for treating or preventing a disease or disorder by administering to a patient in need of such treatment a substance that modulates the activity of a protease polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.

Preferably the disease is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases

include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

Preferably the disease is selected from the group consisting of immune-related diseases and disorders, cardiovascular disease, and cancer. Most preferably, the immune-related diseases and disorders are selected from the group consisting of rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplantation.

Substances useful for treatment of protease-related disorders or diseases preferably show positive results in one or more *in vitro* assays for an activity corresponding to treatment of the disease or disorder in question (Examples of such assays are provided herein, including Example 7). Examples of substances that can be screened for favorable activity are provided and referenced throughout the specification, including this section (**Screening Methods to Identify Substances**

that Modulate Protease Activity). The substances that modulate the activity of the proteases preferably include, but are not limited to, antisense oligonucleotides, ribozymes, and other inhibitors of proteases, as determined by methods and screens referenced this section and in Example 7, below, and any other suitable methods.

- 5 The use of antisense oligonucleotides and ribozymes are discussed more fully in the Section "Gene Therapy," below.

The term "preventing" refers to decreasing the probability that an organism contracts or develops an abnormal condition.

- 10 The term "treating" refers to having a therapeutic effect and at least partially alleviating or abrogating an abnormal condition in the organism.

- The term "therapeutic effect" refers to the inhibition or activation factors causing or contributing to the abnormal condition. A therapeutic effect relieves to some extent one or more of the symptoms of the abnormal condition. In reference to the treatment of abnormal conditions, a therapeutic effect can refer to one or more of
- 15 the following: (a) an increase or decrease in the proliferation, growth, and/or differentiation of cells; (b) activation or inhibition (*i.e.*, slowing or stopping) of cell death; (c) inhibition of degeneration; (d) relieving to some extent one or more of the symptoms associated with the abnormal condition; and (e) enhancing the function of the affected population of cells. Compounds demonstrating efficacy against
- 20 abnormal conditions can be identified as described herein.

The term "abnormal condition" refers to a function in the cells or tissues of an organism that deviates from their normal functions in that organism. An abnormal condition can relate to cell proliferation, cell differentiation, or cell survival.

- 25 Abnormal cell proliferative conditions include cancers such as fibrotic and mesangial disorders, abnormal angiogenesis and vasculogenesis, wound healing, psoriasis, diabetes mellitus, and inflammation.

Abnormal differentiation conditions include, but are not limited to neurodegenerative disorders, slow wound healing rates, and slow tissue grafting healing rates.

5 Abnormal cell survival conditions relate to conditions in which programmed cell death (apoptosis) pathways are activated or abrogated. A number of proteases are associated with the apoptosis pathways. Aberrations in the function of any one of the proteases could lead to cell immortality or premature cell death.

10 The term "aberration", in conjunction with the function of a protease in a signal transduction process, refers to a protease that is over- or under-expressed in an organism, mutated such that its catalytic activity is lower or higher than wild-type protease activity, mutated such that it can no longer interact with a natural binding partner, is no longer modified by another protein, or no longer interacts with a natural binding partner.

15 The term "administering" relates to a method of incorporating a compound into cells or tissues of an organism. The abnormal condition can be prevented or treated when the cells or tissues of the organism exist within the organism or outside of the organism. Cells existing outside the organism can be maintained or grown in cell culture dishes. For cells harbored within the organism, many techniques exist in the art to administer compounds, including (but not limited to) oral, parenteral, 20 dermal, injection, and aerosol applications. For cells outside of the organism, multiple techniques exist in the art to administer the compounds, including (but not limited to) cell microinjection techniques, transformation techniques, and carrier techniques.

25 The abnormal condition can also be prevented or treated by administering a compound to a group of cells having an aberration in a signal transduction pathway to an organism. The effect of administering a compound on organism function can then be monitored. The organism is preferably a mouse, rat, rabbit, guinea pig, or goat, more preferably a monkey or ape, and most preferably a human.

In another aspect, the invention features methods for detection of a protease polypeptide in a sample as a diagnostic tool for diseases or disorders, wherein the method comprises the steps of: (a) contacting the sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid target region of a protease polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, said probe comprising the nucleic acid sequence encoding the polypeptide, fragments thereof, and the complements of the sequences and fragments; and (b) detecting the presence or amount of the probe:target region hybrid as an indication of the disease.

In preferred embodiments of the invention, the disease or disorder is selected from the group consisting of rheumatoid arthritis, arteriosclerosis, autoimmune disorders, organ transplantation, myocardial infarction, cardiomyopathies, stroke, renal failure, oxidative stress-related neurodegenerative disorders, and cancer. Preferably the disease is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders. More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon,

WO 02/00860

PCT/US01/20171

lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

The protease "target region" is the nucleotide base sequence selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58,

WO 02/00860

PCT/US01/20171

and SEQ ID NO:59, or the corresponding full-length sequences, a functional derivative thereof, or a fragment thereof or a domain thereof to which the nucleic acid probe will specifically hybridize. Specific hybridization indicates that in the presence of other nucleic acids the probe only hybridizes detectably with the nucleic acid target region of the protease of the invention. Putative target regions can be identified by methods well known in the art consisting of alignment and comparison of the most closely related sequences in the database.

In preferred embodiments the nucleic acid probe hybridizes to a protease target region encoding at least 6, 12, 75, 90, 105, 120, 150, 200, 250, 300 or 350 contiguous amino acids of a sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, or the corresponding full-length amino acid sequence, or a functional derivative thereof. Hybridization conditions should be such that hybridization occurs only with the protease genes in the presence of other nucleic acid molecules. Under stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having more than 1 or 2 mismatches out of 20 contiguous nucleotides. Such conditions are defined in Berger *et al.* (1987) (Guide

WO 02/00860

PCT/US01/20171

to Molecular Cloning Techniques pg 421, hereby incorporated by reference herein in its entirety including any figures, tables, or drawings.).

5 The diseases for which detection of protease genes in a sample could be diagnostic include diseases in which protease nucleic acid (DNA and/or RNA) is amplified in comparison to normal cells. By "amplification" is meant increased numbers of protease DNA or RNA in a cell compared with normal cells. In normal cells, proteases may be found as single copy genes. In selected diseases, the chromosomal location of the protease genes may be amplified, resulting in multiple copies of the gene, or amplification. Gene amplification can lead to amplification of
10 protease RNA, or protease RNA can be amplified in the absence of protease DNA amplification.

"Amplification" as it refers to RNA can be the detectable presence of protease RNA in cells, since in some normal cells there is no basal expression of protease RNA. In other normal cells, a basal level of expression of protease exists,
15 therefore in these cases amplification is the detection of at least 1-2-fold, and preferably more, protease RNA, compared to the basal level.

The diseases that could be diagnosed by detection of protease nucleic acid in a sample preferably include cancers. The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid
20 extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample that is compatible with the method utilized.

25 In a final aspect, the invention features a method for detection of a protease polypeptide in a sample as a diagnostic tool for a disease or disorder, wherein the method comprises: (a) comparing a nucleic acid target region encoding the protease polypeptide in a sample, where the protease polypeptide has an amino acid sequence selected from the group consisting those set forth in SEQ ID NO:60, SEQ ID

WO 02/00860

PCT/US01/20171

- NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, or one or more fragments thereof, with a control nucleic acid target region encoding the protease polypeptide, or one or more fragments thereof; and (b) detecting differences in sequence or amount between the target region and the control target region, as an indication of the disease or disorder. Preferably the disease is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, and metabolic disorders.
- More specifically these diseases include cancer of tissues, blood, or hematopoietic origin, particularly those involving breast, colon, lung, prostate, cervical, brain, ovarian, bladder, or kidney; central or peripheral nervous system diseases and conditions including migraine, pain, sexual dysfunction, mood disorders, attention disorders, cognition disorders, hypotension, and hypertension; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Tourette's Syndrome; neurodegenerative diseases including Alzheimer's, Parkinson's, Multiple sclerosis, and Amyotrophic lateral sclerosis; viral or non-viral infections caused by HIV-1, HIV-2 or other viral- or prion-agents

WO 02/00860

PCT/US01/20171

or fungal- or bacterial- organisms; metabolic disorders including Diabetes and obesity and their related syndromes, among others; cardiovascular disorders including reperfusion restenosis, coronary thrombosis, clotting disorders, unregulated cell growth disorders, atherosclerosis; ocular disease including

5 glaucoma, retinopathy, and macular degeneration; inflammatory disorders including rheumatoid arthritis, chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, atherosclerosis, rhinitis, autoimmunity, and organ transplant rejection.

The term "comparing" as used herein refers to identifying discrepancies

10 between the nucleic acid target region isolated from a sample, and the control nucleic acid target region. The discrepancies can be in the nucleotide sequences, e.g. insertions, deletions, or point mutations, or in the amount of a given nucleotide sequence. Methods to determine these discrepancies in sequences are well-known to one of ordinary skill in the art. The "control" nucleic acid target region refers to the

15 sequence or amount of the sequence found in normal cells, e.g. cells that are not diseased as discussed previously.

The term "domain" refers to a region of a polypeptide which serves a particular function. For instance, N-terminal or C-terminal domains of signal transduction proteins can serve functions including, but not limited to, binding

20 molecules that localize the signal transduction molecule to different regions of the cell or binding other signaling molecules directly responsible for propagating a particular cellular signal. Some domains can be expressed separately from the rest of the protein and function by themselves, while others must remain part of the intact protein to retain function. The latter are termed functional regions of proteins

25 and also relate to domains.

The expression of proteases can be modulated by signal transduction pathways such as the Ras/MAP kinase signaling pathways. Additionally, the activity of proteases can modulate the activity of the MAP kinase signal transduction

WO 02/00860

PCT/US01/20171

pathway. Furthermore, proteases can be shown to be instrumental in the communication between disparate signal transduction pathways.

The term "signal transduction pathway" refers to the molecules that propagate an extracellular signal through the cell membrane to become an intracellular signal. This signal can then stimulate a cellular response. The polypeptide molecules involved in signal transduction processes are typically receptor and non-receptor protein tyrosine kinases, receptor and non-receptor protein phosphatases, polypeptides containing SRC homology 2 and 3 domains, phosphotyrosine binding proteins (SRC homology 2 (SH2) and phosphotyrosine binding (PTB and PH) domain containing proteins), proline-rich binding proteins (SH3 domain containing proteins), GTPases, phosphodiesterases, phospholipases, prolyl isomerases, proteases, Ca^{2+} binding proteins, cAMP binding proteins, guanyl cyclases, adenylyl cyclases, NO generating proteins, nucleotide exchange factors, and transcription factors.

The summary of the invention described above is not limiting and other features and advantages of the invention will be apparent from the following detailed description of the invention, and from the claims.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A-WW shows the nucleotide sequences for human proteases oriented in a 5' to 3' direction (SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39,

WO 02/00860

PCT/US01/20171

SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44,
 SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49,
 SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54,
 SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ ID
 5 NO:59). In the sequences, N means any nucleotide.

Figure 2A-S shows the amino acid sequences for the human proteases
 encoded by SEQ ID No. 1-59 in the direction of translation (SEQ ID NO:60, SEQ
 ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ
 ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70 SEQ
 10 ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ
 ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ
 ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ
 ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ
 ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ
 15 ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ
 ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105,
 SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID
 NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114,
 SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118). In the
 20 sequences, X means any amino acid.

DETAILED DESCRIPTION OF THE INVENTION

The following description of the background of the invention is provided to
 aid in understanding the invention, but is not admitted to be or to describe prior art
 25 to the invention.

Proteases are enzymes capable of severing the amino acid backbone of other
 proteins, and are involved in a large number of diverse processes within the body.
 Their normal functions include modulation of apoptosis (caspases) (Salvesen and

WO 02/00860

PCT/US01/20171

Dixon, *Cell*, 1997, 91:443-46), control of blood pressure (renin, angiotensin-converting enzymes) (van Hooft *et al.*, 1991, *N Engl J Med.* 324(19):1305-11, and chapters 254 and 359 in Barrett *et al.*, *Handbook of Proteolytic Enzymes*, 1998, Academic Press, San Diego), tissue remodeling and tumor invasion (collagenase) (Vu *et al.*, 1998, *Cell* 93:411-22, Werb, 1997, *Cell*, 91:439-442), development of Alzheimer's Disease (β -secretase) (De Strooper *et al.*, 1999, *Nature* 398:518-22), protein turnover and cell-cycle regulation (proteasome) (Bastians *et al.*, 1999, *Mol. Biol. Cell.* 10:3927-41, Gottesman, *et al.*, 1997, *Cell*, 91:435-38, Larsen *et al.*, 1997, *Cell*, 91:431-34), inflammation (TNF- α convertase) (Black *et al.*, *Nature*, 1997, 385:729-33), and protein turnover (Bochtler *et al.*, 1999, *Annu. Rev. Biophys Biomol Struct.* 28:295-317). Proteases may be classified into several major groups including serine proteases, cysteine proteases, aspartyl proteases, metalloproteases, threonine proteases, and other proteases.

15 **1. Aspartyl proteases (A1; Prosite number PS00141):**

Aspartyl proteases, also known as acid proteases, are a widely distributed family of proteolytic enzymes in vertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue. The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Enzymes in this class include cathepsin E, renin, presenilin (PS1), and the APP secretases.

Cathepsin E

25 Cathepsin E is an immunologically discrete aspartic protease found in the gastrointestinal tract (Azuma *et al.*, 1992 *J. Biol. Chem.*, 267:1609-1614). Cathepsin E is an intracellular proteinase that does not appear to be involved in the digestion of

WO 02/00860

PCT/US01/20171

dietary protein. It is found in highest concentration in the surface of epithelial mucus-producing cells of the stomach. It is the first aspartic proteinase expressed in the fetal stomach and is found in more than half of gastric cancers. It appears, therefore, to be an 'oncofetal' antigen. Its association with stomach cancers suggests
5 it may play a role in the development of this disease.

Renin

Released by the juxtaglomerular cells of the kidney, renin catalyzes the first step in the activation pathway of angiotensinogen--a cascade that can result in
10 aldosterone release, vasoconstriction, and increase in blood pressure. Renin cleaves angiotensinogen to form angiotensin I, which is converted to angiotensin II by angiotensin I converting enzyme, an important regulator of blood pressure and electrolyte balance. Renin occurs in other organs than the kidney, e.g., in the brain, where it is implicated in the regulation of numerous activities.
15

Presenilin proteins

Alzheimer's disease (AD) patients with an inherited form of the disease carry mutations in the presenilin proteins (PSEN1; PSEN2) or the amyloid precursor protein (APP). These disease-linked mutations result in increased production of the
20 longer form of amyloid-beta (main component of amyloid deposits found in AD brains) (Saftig et al., *Eur. Arch. Psychiatry Clin. Neurosci.*, 1999, 249:271-79). Presenilins are postulated to regulate APP processing through their effects on gamma-secretase, an enzyme that cleaves APP (Cruts et al., 1998, *Hum. Mutat.*, 11:183-190, Haass et al., *Science*, 1999, 286:916-19). Also, it is thought that the
25 presenilins are involved in the cleavage of the Notch receptor, such that that they either directly regulate gamma-secretase activity or themselves are protease enzymes

WO 02/00860

PCT/US01/20171

(De Strooper et al., *Nature*, 1999, 398:518-22). Two alternative transcripts of PSEN2 have been identified (Sato et al., 1999, *J. Neurochem.* 72(6):2498-505). Point mutations in the PS1 gene result in a selective increase in the production of the amyloidogenic peptide amyloid-beta (1-42) by proteolytic processing of the amyloid precursor protein (APP) (Lemere et al., 1996, *Nat Med* 2(10):1146-50). The possible role of PS1 in normal APP processing was studied by De Strooper et al. (*Nature* 391: 387-390, 1998) in neuronal cultures derived from PS1-deficient mouse embryos. They found that cleavage by alpha- and beta-secretase of the extracellular domain of APP was not affected by the absence of PS1, whereas cleavage by gamma-secretase of the transmembrane domain of APP was prevented, causing C-terminal fragments of APP to accumulate and a 5-fold drop in the production of amyloid peptide. Pulse-chase experiments indicated that PS1 deficiency specifically decreased the turnover of the membrane-associated fragments of APP. Thus, PS1 appears to facilitate a proteolytic activity that cleaves the integral membrane domain of APP. The results indicated to the authors that mutations in PS1 that manifest clinically cause a gain of function, and that inhibition of PS1 activity is a potential target for anti-amyloidogenic therapy in Alzheimer disease.

Beta-secretase

Beta-secretase, expressed specifically in the brain, is responsible for the proteolytic processing of the amyloid precursor protein (APP) associated with Alzheimer's disease (Potter et al., 2000, *Nat. Biotechnol* 18(2):125-26). It cleaves at the amino terminus of the beta peptide sequence, between residues 671 and 672 of APP, leading to the generation and extracellular release of beta-cleaved soluble APP, and a carboxyterminal fragment that is later released by gamma-secretase (Kinberly et al., 2000 *J. Biol. Chem.* 275(5):3173-78). Yan et al. (*Nature*, 1999, 402:533-37) identified a new membrane-bound aspartyl protease (Asp2) with beta-secretase activity. The Asp2 gene is expressed widely in brain and other tissues. Decreasing the expression of Asp2 in cells reduces amyloid beta-peptide production

and blocks the accumulation of the carboxy-terminal APP fragment that is created by beta-secretase cleavage. Asp2 is a new protein target for drugs that are designed to block the production of amyloid beta-peptide and the consequent formation of amyloid plaque in Alzheimer's disease.

5 Two aspartyl proteases involved in human placentation have recently been isolated: decidua aspartyl protease (DAP-1) and DAP-2 (Moses et al., *Mol. Hum. Reprod.* 1999, 5:983-89).

Another member of the aspartyl peptidase family is HIV-1 retropepsin, from the human immunodeficiency virus type 1. This enzyme is vital for processing of
10 the viral polyprotein and maturation of the mature virion.

2. Cysteine proteases

Another class of proteases which perform a wide variety of functions within the body are the cysteine proteases. Among their roles are the processing of precursor proteins, and intracellular degradation of proteins marked for disposal via
15 the ubiquitin pathway. Eukaryotic cysteine proteases are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. Peptidases in this family with
20 important roles in disease include the caspases, calpain, hedgehog, and Ubiquitin hydrolases.

Cysteine proteases are produced by a large number of cells including those of the immune system (macrophages, monocytes, etc.). These immune cells exercise their protective role in the body, in part, by migrating to sites of
25 inflammation and secreting molecules, among the secreted molecules are cysteine proteases.

Under some conditions, the inappropriate regulation of cysteine proteases of the immune system can lead to autoimmune diseases such as rheumatoid arthritis. For example, the over-secretion of the cysteine protease cathepsin C causes the

WO 02/00860

PCT/US01/20171

degradation of elastin, collagen, laminin, and other structural proteins found in bones. Bone subjected to this inappropriate digestion is more susceptible to metastasis.

Caspase (C14) - apoptosis

5 A cascade of protease reactions is believed to be responsible for the apoptotic changes observed in mammalian cells undergoing programmed cell death. This cascade involves many members of the aspartate-specific cysteine proteases of the caspase family, including caspases 2, 3, 6, 7, 8 and 10 (Salvesen and Dixit, *Cell* 1997, 91:443-446). Cancer cells that escape apoptotic signals, generated by
10 cytotoxic chemotherapeutics or loss of normal cellular survival signals (as in metastatic cells), can go on to develop palpable tumors.

Other caspases are also involved in the activation of pro-inflammatory cytokines. Caspase 1 specifically processes the precursors of IL-1 β , and IL-18 (interferon- γ -inducing factor)(Salvesen and Dixit *Cell* 1997).

15

Calpain (C2) - axonal death, dystrophies

Calcium-dependent cysteine proteases, collectively called calpain, are widely distributed in mammalian cells (Wang, 2000, *Trends Neurosci.* 23(1):20-26). The calpains are nonlysosomal intracellular cysteine proteases. The mammalian calpains
20 include 2 ubiquitous proteins, CAPN1 and CAPN2, as well as 2 stomach-specific proteins, and CAPN3, which is muscle-specific (Herasse *et al.*, 1999, *Mol. Cell. Biol.* 19(6):4047-55). The ubiquitous enzymes consist of heterodimers with distinct large subunits associated with a common small subunit, all of which are encoded by different genes. The large subunits of calpains can be subdivided into 4 domains;
25 domains I and III, whose functions remain unknown, show no homology with known proteins. The former, however, may be important for the regulation of the proteolytic activity. Domain II shows similarity with other cysteine proteases, which share histidine, cysteine, and asparagine residues at their active sites. Domain IV is calmodulin-like. CAPN5 and CAPN6 differ from previously identified

WO 02/00860

PCT/US01/20171

vertebrate calpains in that they lack a calmodulin-like domain IV (Ohno *et al.*, 1990, *Cytogenet. Cell Genet.* 53(4):225-29).

Mutations in the CAPN3 gene have been associated with limb-girdle muscular dystrophy, type 2A (LGMD2A) (Allamand *et al.*, 1995, *Hum. Molec. Genet.* 4:459-463). The slowly progressive muscle weakness associated with this disease is usually first evident in the pelvic girdle and then spreads to the upper limbs while sparing facial muscles. Calpain has also been implicated in the development of hyperactive Cdk5 leading to neuronal cell death associated with Alzheimer's disease (Patrick *et al.*, 1999, *Nature* 402:615-622).

10

Hedgehog (C46) – Cancer

The organization and morphology of the developing embryo are established through a series of inductive interactions. One family of vertebrate genes has been described related to the *Drosophila* gene 'hedgehog' (*hh*) that encodes inductive signals during embryogenesis (Johnson and Tabin, 1997, *Cell* 90:979-990). 'Hedgehog' encodes a secreted protein that is involved in establishing cell fates at several points during *Drosophila* development (Marigo *et al.*, 1995, *Genomics* 28:44-51). There are 3 known mammalian homologs of *hh*: Sonic hedgehog (*Shh*), Indian hedgehog (*Ihh*), and desert hedgehog (*Dhh*) (Johnson and Tabin, 1997, *Cell* 90:979-990). Like its *Drosophila* cognate, *Shh* encodes a signal that is instrumental in patterning the early embryo. It is expressed in Hensen's node, the floorplate of the neural tube, the early gut endoderm, the posterior of the limb buds, and throughout the notochord (Chiang *et al.*, 1996, *Nature* 383:407-413). It has been implicated as the key inductive signal in patterning of the ventral neural tube, the anterior-posterior limb axis, and the ventral somites. Oro *et al.* ("Basal cell carcinomas in mice overexpressing sonic hedgehog." *Science* 276: 817-821, 1997) showed that transgenic mice overexpressing SHH in the skin developed many features of the basal cell nevus syndrome, demonstrating that SHH is sufficient to induce basal cell carcinomas (BCCs) in mice. The data suggested that SHH may

15

20

25

WO 02/00860

PCT/US01/20171

have a role in human tumorigenesis. Activating mutations of SHH or another 'hedgehog' gene may be an alternative pathway for BCC formation in humans. The human mutation his133tyr (his134tyr in mouse) is a candidate. It is distinct from loss-of-function mutations reported for individuals with holoprosencephaly (Oro *et al.*, 1997, *Science* 276:817-821). His133 lies adjacent in the catalytic site to his134, one of the conserved residues thought to be necessary for catalysis. SHH may be a dominant oncogene in multiple human tumors, a mirror of the tumor suppressor activity of the opposing 'patched' (PTCH) gene (Aszterbaum *et al.*, 1998, *J. Invest. Derm.* 110:885-888). The rapid and frequent appearance of Shh-induced tumors in the mice suggested that disruption of the SHH-PTC pathway is sufficient to create BCCs.

Members of the vertebrate hedgehog family (Sonic, Indian, and Desert) have been shown to be essential for the development of various organ systems, including neural, somite, limb, skeletal, and for male gonad morphogenesis. Desert hedgehog is expressed in the developing retina, whereas Indian hedgehog (Ihh) is expressed in the developing and mature retinal pigmented epithelium beginning at embryonic day 13 (Levine *et al.*, *J. Neurosci.*, 1997, 17(16):6277-88). Dhh has also been implicated in having a role in the regulation of spermatogenesis. Sertoli cell precursors express Sry, sex determining gene, which leads to testis development in mammals. Dhh expression is initiated in Sertoli cell precursors shortly after the activation of Sry and persists in the testis into the adult. Bitgood *et al.* (*Curr. Biol.*, 1996, 6(3):298-304) disclose that female mice homozygous for a Dhh-null mutation show no obvious phenotype, whereas males are viable but infertile having a complete absence of mature sperm, demonstrating that Dhh signaling plays an essential role in the regulation of mammalian spermatogenesis. Dhh has also been found to have a role in the and maintenance of protective nerve sheaths endo-, peri- and epineurium. In Dhh knockout mice, the connective tissue sheaths in adult nerves appear highly abnormal by electron microscopy. Mirsky *et al.*, (*Ann. N.Y. Acad. Sci.*, 1999, 883:196-202) demonstrate that Dhh signaling from Schwann cells to the

WO 02/00860

PCT/US01/20171

mesenchyme is involved in the formation of a morphologically and functionally normal perineurium.

Recent advances in developmental and molecular biology during embryogenesis and organogenesis have provided new insights into the mechanism of bone formation. Iwasaki *et al.*, (*J. Bone Joint Surg. Br.*, 1999, 81(6):1076-82) demonstrate that Indian Hedgehog (Ihh) is expressed in cartilage cell precursors and later in mature and hypertrophic chondrocytes. Ihh plays a critical role in the morphogenesis of the vertebrate skeleton. Becker *et al.* (*Dev. Biol.*, 1997, 187(2):298-310) provide data which suggests that Ihh is also involved in mediating differentiation of extraembryonic endoderm during early mouse embryogenesis. Short limbed dwarfism, with decreased chondrocyte proliferation and extensive hypertrophy are the results of targeted deletion of Ihh (Karp *et al.*, 2000, Development 127(3):543-48). The expression of Ihh mRNA and protein is unregulated dramatically as F9 cells differentiate in response to retinoic acid, into either parietal endoderm or embryoid bodies, containing an outer visceral endoderm layer. RT-PCR analysis of blastocyst outgrowth cultures demonstrates that whereas little or no Ihh message is present in blastocysts, significant levels appear upon subsequent days of culture, coincident with the emergence of parietal endoderm cells.

20 Ubiquitin hydrolases (C12) - apoptosis, checkpoint integrity

14 genes in this patent belong to the ubiquitin hydrolase family, SEQID:5, SEQID:6, SEQID:7, SEQID:8, SEQID:9, SEQID:10, SEQID:11, SEQID:12, SEQID:13, SEQID:14, SEQID:15, SEQID:16, SEQID:17, SEQID:18. The polypeptides encoded by these genes may have one or more of the following activities.

Ubiquitin carboxyl-terminal hydrolases (3.1.2.15) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of

WO 02/00860

PCT/US01/20171

poly-ubiquitin precursors as well as that of ubiquitinated proteins. In eukaryotic cells, the covalent attachment of ubiquitin to proteins plays a role in a variety of cellular processes. In many cases, ubiquitination leads to protein degradation by the 26S proteasome. Protein ubiquitination is reversible, and the removal of ubiquitin is catalyzed by deubiquitinating enzymes, or DUBs. A defect in these enzymes, catalyzing the removal of ubiquitin from ubiquitinated proteins, may be characteristic of neurodegenerative diseases such as Alzheimer's, Parkinson's, progressive supranuclear palsy, and Pick's and Kuf's disease.

10 Papain (C1) — cathepsins K, S and B - bone resorption. Ag processing (Prosite PS00139).

One gene in this patent belongs to the Papain family, SEQID:4. The polypeptide encoded by this gene may have one or more of the following activities.

15 Cathepsin K, a member of the papain family of peptidases, is involved in osteoclastic resorption. It plays an important role in extracellular degradation and may have a role in disorders of bone remodeling, such as pycnodysostosis, an autosomal recessive osteochondrodysplasia characterized by osteosclerosis and short stature. Antigen presentation by major histocompatibility complex (MHC) class II molecules requires the participation of different proteases in the endocytic route to
20 degrade endocytosed antigens as well as the MHC class II-associated invariant chain. Only cathepsin S, a member of the papain family, appears to be essential for complete destruction of the invariant chain. Cathepsin B is overexpressed in tumors of the lung, prostate, colon, breast, and stomach. Hughes et al. (*Proc. Nat. Acad. Sci.* 95: 12410-12415, 1998) found an amplicon at 8p23-p22 that resulted in cathepsin B
25 overexpression in esophageal adenocarcinoma. Abundant extracellular expression of CTSB protein was found in 29 of 40 (72.5%) of esophageal adenocarcinoma specimens by use of immunohistochemical analysis. The findings were thought to support an important role for CTSB in esophageal adenocarcinoma and possibly in other tumors.

WO 02/00860

PCT/US01/20171

Cathepsin B, a lysosomal protease, is being studied as a prognostic marker in various cancers (breast, pulmonary adenocarcinomas).

Cysteine Protease AEP

- 5 The cysteine protease AEP plays another role in the immune functions. It has been implicated in the protease step required for antigen processing in B cells. (Manoury *et al. Nature* 396:695-699 (1998))

Hepatitis A viral protease (C3E)

- 10 The Hepatitis A genome encodes a cysteine protease required for enzymatic cleavages *in vivo* to yield mature proteins (Wang, 1999, *Prog. Drug Res.* 52:197-219). This enzyme and its homologs in other viruses (such as hepatitis E virus) are potential targets for chemotherapeutic intervention.

15 3. Metalloproteases

Collagenase (M10) – invasion

Two genes in this patent are members of the M10 family, SEQID:19 and SEQID:20. The polypeptides encoded by these genes may have one or more of the following activities.

- 20 Matrix degradation is an essential step in the spread of cancer. The 72- and 92-kD type IV collagenases are members of a group of secreted zinc metalloproteases which, in mammals, degrade the collagens of the extracellular matrix. Other members of this group include interstitial collagenase and stromelysin (Nagase *et al.*, 1992, *Matrix Suppl.* 1:421-424). By targeted disruption in embryonic stem cells, Vu *et al.* (*Cell*, 1998, 934:11-22) created homozygous mice with a null mutation in the MMP9/gelatinase B gene. These mice exhibited an abnormal pattern of skeletal growth plate vascularization and ossification. Growth plates from MMP9-null mice in culture showed a delayed release of an angiogenic activator, establishing a role for this proteinase in controlling angiogenesis.

WO 02/00860

PCT/US01/20171

MMP2 (gelatinase A) have been associated with the aggressiveness of human cancers (Chenard *et al.*, 1999, *Int. J. Cancer*, 82:208-12). In a study comparing basal cell carcinomas (BCC) with the more aggressive squamous cell carcinomas (SCC), both MMP2 and MMP9 were expressed at a higher level in SCC (Dumas *et al.*, 1999, *Anticancer Res.*, 19(4B):2929-38). Additionally, expression of MMP2 and MMP9 in T lymphocytes has recently been shown to be modulated by the Ras/MAP kinase signaling pathways (Esparza *et al.*, 1999, *Blood*, 94:2754-66) (see also, Li *et al.*, 1998, *Biochim. Biophys. Acta*, 1405:110-20).

10 ADAMs (M12) - TNF, inflammation, growth factor processing

The ADAM peptidases are a family of proteins containing a disintegrin and metalloproteinase (ADAM) domain (Werb and Yan, *Science*, 1998, 282:1279-1280). Members of this family are cell surface proteins with a unique structure possessing both potential adhesion and protease domains (Primakoff and Myles, *Trends in Genet.*, 2000, 16:83-87). Activity of these proteases can be linked to TNF, inflammation, and/or growth factor processing.

ADAM proteases have also been characterized as having a pro- and metalloproteinase domain, a disintegrin domain, a cysteine-rich region and an EGF repeat (Blobel, 1997, *Cell*, 90:589-592 which is hereby incorporated herein by reference in its entirety including any figures, tables, or drawings). They have been associated with the release from the plasma membrane of numerous proteins including Tumor Necrosis Factor- α (TNF- α), kit-ligand, TGF α , Fas-ligand, cytokine receptors such as the IL-6 receptor and the NGF receptor, as well as adhesion proteins such as L-selectin, and the b amyloid precursor proteins (Blobel, 1997, *Cell*, 90:589-592).

Tumor necrosis factor- α is synthesized as a proinflammatory cytokine from a 233-amino acid precursor. Conversion of the membrane-bound precursor to a secreted mature protein is mediated by a protease termed TNF- α convertase. TNF- α

WO 02/00860

PCT/US01/20171

is involved in a variety of diseases. ADAM17, which contains a disintegrin and metalloproteinase domains, is also called 'tumor necrosis factor- α converting enzyme' (TACE) (Black *et al.*, *Nature*, 1997, 385:729-33). The gene encodes an 824-amino acid polypeptide containing the features of the ADAM family: a secretory signal sequence, a disintegrin domain, and a metalloprotease domain. Expression studies showed that the encoded protein cleaves precursor tumor necrosis factor- α to its mature form. This enzyme may also play a role in the processing of Transforming Growth Factor- α (TGF- α), as mice which lack the gene are similar in phenotype to those that lack TGF- α (Peschon *et al.*, *Science*, 282:1281-1284).

Neprylisin (M13) - Endothelin-converting enzyme

One gene in this patent, SEQID:21, is a member of this family. The polypeptide encoded by this gene may have one or more of the following activities. Neprylisin, a metalloproteinase active in degradation of enkephalins and other bioactive peptides, is a drug target in hypertension and renal disease (Oefner, *et al.*, *J. Mol. Biol.*, 2000, 296:341-49).

Carboxypeptidase (M14) - Neurotransmitter processing

Three genes in this application are Zn carboxypeptidases, SEQID:1, SEQID:2, and SEQID:3. The polypeptides encoded by these genes may have one or more of the following activities.

Carboxypeptidases specifically remove COOH-terminal basic amino acids (arginine or lysine). They have important functions in many biologic processes, including activation, inactivation, or modulation of peptide hormone activity, neurotransmitter processing, and alteration of physical properties of proteins and enzymes.

WO 02/00860

PCT/US01/20171

Dipeptidase (M2) -- ACE

One protease in this patent is a member of the M2 family: SEQID:22. The polypeptide encoded by this gene may have one or more of the following activities.

- 5 Angiotensin I converting enzyme (EC 3.4.15.1), or kininase II, is
 adipeptidyl carboxypeptidase that plays an important role in blood pressure
 regulation and electrolyte balance by hydrolyzing angiotensin I into angiotensin II, a
 potent vasopressor, and aldosterone-stimulating peptide. The enzyme is also able to
 inactivate bradykinin, a potent vasodilator. Although angiotensin-converting
 10 enzyme has been studied primarily in the context of its role in blood pressure
 regulation, this widely distributed enzyme has many other physiologic functions.
 There are two forms of ACE: a testis-specific isozyme and a somatic isozyme which
 has two active centers.

15 Matrix metalloproteases (M10B) -- tissue remodeling and inflammation

- The matrix metalloproteases (MMPs) are a family of related matrix-
 degrading enzymes that are important in tissue remodeling and repair during
 development and inflammation. Abnormal expression is associated with various
 diseases such as tumor invasiveness, arthritis, and atherosclerosis. MMP activity
 20 may also be related to tobacco-induced pulmonary emphysema.

- The matrix metalloproteases (MMPs) are a family of related matrix-
 degrading enzymes that are important in tissue remodeling and repair during
 development and inflammation (Belotti *et al.*, 1999, *Int. J. Biol. Markers* 14(4):232-
 38). Abnormal expression is associated with various diseases such as tumor
 25 invasiveness (Johansson and Kahari, 2000, *Histol. Histopathol.* 15(1):225-37),
 arthritis (Malemud *et al.*, 1999, *Front. Biosci.* 4:D762-71), and atherosclerosis
 (Nagase, 1997, *Biol. Chem.* 378(3-4):151-60). MMP activity may also be related to

WO 02/00860

PCT/US01/20171

tobacco-induced pulmonary emphysema (Dhami *et al.*, *Am. J. Respir. Cell Mol. Biol.*, 2000, 22:244-52).

SREBP Protease (M50)

5 The sterol regulatory element-binding proteins protease functions in the intra-membrane proteolysis and release of sterol-regulatory binding proteins (SREBPs) (Duncan *et al.*, 1997, *J. Biol. Chem.* 272:12778-85). SREBPs activate genes of cholesterol and fatty acid metabolism, making the SREBP protease an attractive target for therapeutic modulation (Brown *et al.*, 1997, *Cell* 89:331-340).

10

Metalloprotease processing of growth factors

In addition to the processing of TGF- α described above, metalloproteases have been directly demonstrated to be active in the processing of the precursor of other growth factors such as heparin-binding EGF (proHB-EGF) (Izumi *et al.*, 15 *EMBO J.*, 1998, 17:7260-72), and amphiregulin (Brown *et al.*, 1998, *J. Biol. Chem.*, 27:17258-68).

Additionally, metalloproteases have recently been shown to be instrumental in the communication whereby stimulation of a GPCR pathway results in stimulation of the MAP kinase pathway (Prenzel *et al.*, 1999, *Nature*, 402:884-888). 20 The growth factor intermediate in the pathway, HB-EGF is released by the cell in a proteolytic step regulated by the GPCR pathway involving an uncharacterized metalloprotease. After release, the HB-EGF is bound by the extracellular matrix and then presented to the EGF receptors on the surface, resulting in the activation of the MAP kinase pathway (Prenzel *et al.*, 1999, *Nature*, 402:884-888).

25 A recent study by Gallea-Robache *et al.* (1997) has also implicated a metalloprotease family displaying different substrate specificities in the shedding of other growth factors including macrophage colony-stimulating factor (M-CSF) and stem cell factor (SCF) (Gallea-Robache *et al.*, 1997, *Cytokine* 9:340-46). The

WO 02/00860

PCT/US01/20171

shedding of M-CSF (also known as CSF-1) has been linked to activation of Protein Kinase C by phorbol esters (Stein *et al.*, 1991, *Oncogene*, 6:601-05).

4. Serine Proteases

5 The serine proteases are a class which includes trypsin, kallikrein, chymotrypsin, elastase, thrombin, tissue plasminogen activator (tPA), urokinase plasminogen activator (uPA), plasmin (Werb, *Cell*, 1997, 91:439-442), kallikrein (Clements, *Biol. Res.*, 1998, 31:151-59), and cathepsin G (Shamamian *et al.*, *Surgery*, 2000, 127:142-47). These proteases have in common a well-conserved
10 catalytic triad of amino acid residues in their active site consisting of histidine-57, aspartic acid-102, and serine-195 (using the chymotrypsin numbering system). Serine protease activity has been linked to coagulation and they may have use as tumor markers.

Serine proteases can be further subclassified by their specificity in substrates.
15 The elastases prefer to cleave substrates adjacent to small aliphatic residues such as valine, chymases prefer to cleave near large aromatic hydrophobic residues, and tryptases prefer positively charged residues. One additional class of serine protease has been described recently which prefers to cleave adjacent to a proline. This prolyl endopeptidase has been implicated in the progression of memory loss in
20 Alzheimer's patients (Toide *et al.*, 1998, *Rev. Neurosci.* 9(1):17-29).

A partial list of proteases known to belong to this large and important family include: blood coagulation factors VII, IX, X, XI and XII; thrombin; plasminogen; complement components C1r, C1s, C2; complement factors B, D and I; complement-activating component of RA-reactive factor; elastases 1, 2, 3A, 3B
25 (protease E); hepatocyte growth factor activator; glandular (tissue) kallikreins including EGF-binding protein types A, B, and C; NGF- γ chain, γ -renin, and prostate specific antigen (PSA); plasma kallikrein; mast cell proteases; myeloblastin (proteinase 3) (Wegener's autoantigen); plasminogen activators (urokinase-type, and tissue-type); and the trypsins I, II, III, and IV. These peptidases play key roles in

WO 02/00860

PCT/US01/20171

coagulation, tumorigenesis, control of blood pressure, release of growth factors, and other roles. (<http://www.babraham.co.uk/Merops/Merops.htm>).

- Proteases of the trypsin family in this patent include SGPr434, SEQID:24; SGPr446_1, SEQID:25; SGPr447, SEQID:26; SGPr432_1, SEQID:27; SGPr529, SEQID:28; SGPr428_1, SEQID:29; SGPr425, SEQID:30; SGPr548, SEQID:31; SGPr396, SEQID:32; SGPr426, SEQID:33; SGPr552, SEQID:34; SGPr405, SEQID:35; SGPr485_1, SEQID:36; SGPr534, SEQID:37; SGPr390, SEQID:38; SGPr521, SEQID:39; SGPr530_1, SEQID:40; SGPr520, SEQID:41; SGPr455, SEQID:42; SGPr507_2, SEQID:43; SGPr559, SEQID:44; SGPr567_1, SEQID:45; SGPr479_1, SEQID:46; SGPr489_1, SEQID:47; SGPr465_1, SEQID:48; SGPr524_1, SEQID:49; SGPr422, SEQID:50; SGPr538, SEQID:51; SGPr527_1, SEQID:52; SGPr542, SEQID:53; SGPr551, SEQID:54; SGPr451, SEQID:55; SGPr452_1, SEQID:56; SGPr504, SEQID:57; SGPr469, SEQID:58; SGPr400, SEQID:59. SEQID:23 is a serine protease of the subtilase sub-family. Limited proteolysis of most large protein precursors is carried out in vivo by the subtilisin-like pro-protein convertases. Many important biological processes such as peptide hormone synthesis, viral protein processing and receptor maturation involve proteolytic processing by these enzymes, making them potential targets for the development of novel therapeutic agents (Bergeron F, J Mol Endocrinol 2000 Feb;24(1):1-22)

5. Threonine peptidases (T1) – (Prosite PDOC00326/PDOC00668)

Proteasomal subunits (T1A)

- The proteasome is a multicatalytic threonine proteinase complex involved in ATP/ubiquitin dependent non-lysosomal proteolysis of cellular substrates. It is responsible for selective elimination of proteins with aberrant structures, as well as naturally occurring short-lived proteins related to metabolic regulation and cell-cycle progression (Momand *et al.*, 2000, *Gene* 242(1-2):15-29, Bochtler *et al.*, 1999, *Annu. Rev. Biophys Biomol Struct.*28:295-317). The proteasome inhibitor

WO 02/00860

PCT/US01/20171

lactacystin reversibly inhibits proliferation of human endothelial cells, suggesting a role for proteasomes in angiogenesis (Kumeda, *et al.*, *Anticancer Res.* 1999 Sep-Oct;19(5B):3961-8). Another important function of the proteasome in higher vertebrates is to generate the peptides presented on MHC-class 1 molecules to circulating lymphocytes (Castelli *et al.*, 1997, *Int. J. Clin. Lab. Res.* 27(2):103-10). The proteasome has a sedimentation coefficient of 26S and is composed of a 20S catalytic core and a 22S regulatory complex. Eukaryotic 20S proteasomes have a molecular mass of 700 to 800 kD and consist of a set of over 15 kinds of polypeptides of 21 to 32 kD. All eukaryotic 20S proteasome subunits can be classified grossly into 2 subfamilies, α and β , by their high similarity with either the α or β subunits of the archaeobacterium *Thermoplasma acidophilum* (Mayr *et al.*, 1999, *Biol. Chem.* 380(10):1183-92). Several of the components have been identified as threonine peptidases, suggesting that this class of peptidases plays a key role in regulating metabolic pathways and cell-cycle progression, among other functions (Yorgin *et al.*, 2000, *J. Immunol.* 164(6):2915-23).

6. Peptidases of Unknown Catalytic Mechanism

The prenyl-protein specific protease responsible for post-translational processing of the Ras proto-oncogene and other prenylated proteins falls into this class. This class also includes several viral peptidases that may play a role in mammalian infection, including cardiovirus endopeptidase 2A (encephalomyocarditis virus) (Molla *et al.*, 1993, *J. Virol.* 67(8):4688-95), NS2-3 protease (hepatitis C virus) (Blight *et al.*, 1998, *Antivir. Ther.* 3(Suppl 3):71-81), endopeptidase (infectious pancreatic necrosis virus) (Lejal *et al.*, *J. Gen. Virol.*, 2000, 81:983-992), and the Npro endopeptidase (hog cholera virus) (Tratschin *et al.*, 1998, *J. Virol.* 72(9):7681-84).

Nucleic Acid Probes, Methods, and Kits for Detection of Proteases

A nucleic acid probe of the present invention may be used to probe an appropriate chromosomal or cDNA library by usual hybridization methods to obtain other nucleic acid molecules of the present invention. A chromosomal DNA or cDNA library may be prepared from appropriate cells according to recognized methods in the art (*cf.* "Molecular Cloning: A Laboratory Manual", second edition, Cold Spring Harbor Laboratory, Sambrook, Fritsch, & Maniatis, eds., 1989).

In the alternative, chemical synthesis can be carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the amino acid sequence of the polypeptide of interest. The synthesized nucleic acid probes may be used as primers in a polymerase chain reaction (PCR) carried out in accordance with recognized PCR techniques, essentially according to PCR Protocols, "A Guide to Methods and Applications", Academic Press, Michael, *et al.*, eds., 1990, utilizing the appropriate chromosomal or cDNA library to obtain the fragment of the present invention.

One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer alignment and sequence analysis known in the art ("Molecular Cloning: A Laboratory Manual", 1989, *supra*). The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be visualized using known methods.

The nucleic acid probes of the present invention include RNA, as well as DNA probes, such probes being generated using techniques known in the art. The nucleic acid probe may be immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, and acrylic resins, such as polyacrylamide and latex beads. Techniques for coupling nucleic acid probes to such solid supports are well known in the art.

WO 02/00860

PCT/US01/20171

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The samples used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample which is compatible with the method utilized.

One method of detecting the presence of nucleic acids of the invention in a sample comprises (a) contacting said sample with the above-described nucleic acid probe under conditions such that hybridization occurs, and (b) detecting the presence of said probe bound to said nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art as described above. Samples to be tested include but should not be limited to RNA samples of human tissue.

A kit for detecting the presence of nucleic acids of the invention in a sample comprises at least one container means having disposed therein the above-described nucleic acid probe. The kit may further comprise other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents include, but are not limited to radiolabelled probes, enzymatic labeled probes (horseradish peroxidase, alkaline phosphatase), and affinity labeled probes (biotin, avidin, or streptavidin). Preferably, the kit further comprises instructions for use.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a

WO 02/00860

PCT/US01/20171

container which contains the probe or primers used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, and the like), and containers which contain the reagents used to detect the hybridized probe, bound antibody, amplified product, or the like. One skilled in the art will readily recognize that the nucleic acid probes described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

DNA Constructs Comprising a Protease Nucleic Acid Molecule and Cells Containing These Constructs.

The present invention also relates to a recombinant DNA molecule comprising, 5' to 3', a promoter effective to initiate transcription in a host cell and the above-described nucleic acid molecules. In addition, the present invention relates to a recombinant DNA molecule comprising a vector and an above-described nucleic acid molecule. The present invention also relates to a nucleic acid molecule comprising a transcriptional region functional in a cell, a sequence complementary to an RNA sequence encoding an amino acid sequence corresponding to the above-described polypeptide, and a transcriptional termination region functional in said cell. The above-described molecules may be isolated and/or purified DNA molecules.

The present invention also relates to a cell or organism that contains an above-described nucleic acid molecule and thereby is capable of expressing a polypeptide. The polypeptide may be purified from cells which have been altered to express the polypeptide. A cell is said to be "altered to express a desired polypeptide" when the cell, through genetic manipulation, is made to produce a protein which it normally does not produce or which the cell normally produces at lower levels. One skilled in the art can readily adapt procedures for introducing and expressing either genomic, cDNA, or synthetic sequences into either eukaryotic or prokaryotic cells.

WO 02/00860

PCT/US01/20171

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains nucleotide sequences which contain transcriptional and translational regulatory information and such sequences are "operably linked" to nucleotide sequences which encode the polypeptide. An operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression may vary from organism to organism, but shall in general include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

If desired, the non-coding region 3' to the sequence encoding a protease of the invention may be obtained by the above-described methods. This region may be retained for its transcriptional termination regulatory sequences, such as termination and polyadenylation. Thus, by retaining the 3'-region naturally contiguous to the DNA sequence encoding a protease of the invention, the transcriptional termination signals may be provided. Where the transcriptional termination signals are not satisfactorily functional in the expression host cell, then a 3' region functional in the host cell may be substituted.

Two DNA sequences (such as a promoter region sequence and a sequence encoding a protease of the invention) are said to be operably linked if the nature of the linkage between the two DNA sequences allows the protease sequence to be transcribed, i.e., where the linkage does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of a gene sequence encoding a protease of the invention, or (3) interfere with the ability of the gene sequence of a protease of the invention to be

WO 02/00860

PCT/US01/20171

transcribed by the promoter region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence. Thus, to express a gene encoding a protease of the invention, transcriptional and translational signals recognized by an appropriate host are necessary.

The present invention encompasses the expression of a gene encoding a protease of the invention (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Prokaryotic hosts are, generally, very efficient and convenient for the production of recombinant proteins and are, therefore, one type of preferred expression system for proteases of the invention. Prokaryotes most frequently are represented by various strains of *E. coli*. However, other microbial strains may also be used, including other bacterial strains.

In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Examples of suitable plasmid vectors may include pBR322, pUC118, pUC119 and the like; suitable phage or bacteriophage vectors may include λ gt10, λ gt11 and the like; and suitable virus vectors may include pMAM-neo, pKRC and the like. Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as *E. coli*, *Bacillus*, *Streptomyces*, *Pseudomonas*, *Salmonella*, *Serratia*, and the like. However, under such conditions, the polypeptide will not be glycosylated. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

To express a protease of the invention (or a functional derivative thereof) in a prokaryotic cell, it is necessary to operably link the sequence encoding the protease of the invention to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (*i.e.*, inducible or derepressible). Examples of constitutive promoters include the *int* promoter of bacteriophage λ , the

bla promoter of the β -lactamase gene sequence of pBR322, and the *cat* promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage λ (P_L and P_R), the *trp*, *recA*, *lacZ*, *lacI*, and *gal* promoters of *E. coli*, the α -amylase (Ullmanen *et al.*, *J. Bacteriol.* 162:176-182, 1985) and the ζ -28-specific promoters of *B. subtilis* (Gilman *et al.*, *Gene Sequence* 32:11-20, 1984), the promoters of the bacteriophages of *Bacillus* (Gryczan, in: The Molecular Biology of the Bacilli, Academic Press, Inc., NY, 1982), and *Streptomyces* promoters (Ward *et al.*, *Mol. Gen. Genet.* 203:468-478, 1986).

10 Prokaryotic promoters are reviewed by Glick (*Ind. Microbiol.* 1:277-282, 1987), Cenatiempo (*Biochimie* 68:505-516, 1986), and Gottesman (*Ann. Rev. Genet.* 18:415-442, 1984).

Proper expression in a prokaryotic cell may also require the presence of a ribosome-binding site upstream of the gene sequence-encoding sequence. Such

15 ribosome-binding sites are disclosed, for example, by Gold *et al.* (*Ann. Rev. Microbiol.* 35:365-404, 1981). The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on the type of host cell used to express the gene. As used herein, "cell", "cell line", and "cell culture" may be used interchangeably and all such designations include progeny. Thus, the words

20 "transformants" or "transformed cells" include the primary subject cell and cultures derived therefrom, without regard to the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. However, as defined, mutant progeny have the same functionality as that of the originally transformed cell.

25 Host cells which may be used in the expression systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the protease polypeptide of interest. Suitable hosts may often include eukaryotic cells. Preferred eukaryotic hosts include, for example, yeast, fungi,

WO 02/00860

PCT/US01/20171

insect cells, mammalian cells either *in vivo*, or in tissue culture. Mammalian cells which may be useful as hosts include HeLa cells, cells of fibroblast origin such as VERO or CHO-K1, or cells of lymphoid origin and their derivatives. Preferred mammalian host cells include SP2/0 and J558L, as well as neuroblastoma cell lines
5 such as IMR 332, which may provide better capacities for correct post-translational processing.

In addition, plant cells are also available as hosts, and control sequences compatible with plant cells are available, such as the cauliflower mosaic virus 35S and 19S, and nopaline synthase promoter and polyadenylation signal sequences.
10 Another preferred host is an insect cell, for example the *Drosophila* larvae. Using insect cells as hosts, the *Drosophila* alcohol dehydrogenase promoter can be used (Rubin, *Science* 240:1453-1459, 1988). Alternatively, baculovirus vectors can be engineered to express large amounts of proteases of the invention in insect cells (Jasny, *Science* 238:1653, 1987; Miller *et al.*, in: *Genetic Engineering*, Vol. 8,
15 Plenum, Setlow *et al.*, eds., pp. 277-297, 1986).

Any of a series of yeast expression systems can be utilized which incorporate promoter and termination elements from the actively expressed sequences coding for glycolytic enzymes that are produced in large quantities when yeast are grown in mediums rich in glucose. Known glycolytic gene sequences can also provide very
20 efficient transcriptional control signals. Yeast provides substantial advantages in that it can also carry out post-translational modifications. A number of recombinant DNA strategies exist utilizing strong promoter sequences and high copy number plasmids which can be utilized for production of the desired proteins in yeast. Yeast recognizes leader sequences on cloned mammalian genes and secretes peptides
25 bearing leader sequences (*i.e.*, pre-peptides). Several possible vector systems are available for the expression of proteases of the invention in a mammalian host.

A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as

WO 02/00860

PCT/US01/20171

adenovirus, bovine papilloma virus, cytomegalovirus, simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, and the like, may be employed.

5 Transcriptional initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical (such as metabolite) regulation.

10 Expression of proteases of the invention in eukaryotic hosts requires the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred eukaryotic promoters include, for example, the promoter of the mouse metallothionein I gene sequence (Hamer *et al.*, *J. Mol. Appl. Gen.* 1:273-288, 1982); the TK promoter of
15 Herpes virus (McKnight, *Cell* 31:355-365, 1982); the SV40 early promoter (Benoist *et al.*, *Nature* (London) 290:304-31, 1981); and the yeast gal4 gene sequence promoter (Johnston *et al.*, *Proc. Natl. Acad. Sci. (USA)* 79:6971-6975, 1982; Silver *et al.*, *Proc. Natl. Acad. Sci. (USA)* 81:5951-5955, 1984).

Translation of eukaryotic mRNA is initiated at the codon which encodes the
20 first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence which encodes a protease of the invention (or a functional derivative thereof) does not contain any intervening codons which are capable of encoding a methionine (*i.e.*, AUG). The presence of such codons results either in the formation of a fusion protein (if the AUG codon is in the same
25 reading frame as the protease of the invention coding sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the protease of the invention coding sequence).

A nucleic acid molecule encoding a protease of the invention and an operably linked promoter may be introduced into a recipient prokaryotic or

WO 02/00860

PCT/US01/20171

eukaryotic cell either as a nonreplicating DNA or RNA molecule, which may either be a linear molecule or, more preferably, a closed covalent circular molecule. Since such molecules are incapable of autonomous replication, the expression of the gene may occur through the transient expression of the introduced sequence.

- 5 Alternatively, permanent expression may occur through the integration of the introduced DNA sequence into the host chromosome.

- A vector may be employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or
10 more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, *e.g.*, antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection.
15 Additional elements may also be needed for optimal synthesis of mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama (*Mol. Cell. Biol.* 3:280-289, 1983).

- The introduced nucleic acid molecule can be incorporated into a plasmid or
20 viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are
25 desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Preferred prokaryotic vectors include plasmids such as those capable of replication in *E. coli* (such as, for example, pBR322, ColEI, pSC101, pACYC 184, π VX; "Molecular Cloning: A Laboratory Manual", 1989, *supra*). *Bacillus* plasmids

WO 02/00860

PCT/US01/20171

- include pC194, pC221, pT127, and the like (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, NY, pp. 307-329, 1982). Suitable *Streptomyces* plasmids include p1J101 (Kendall *et al.*, *J. Bacteriol.* 169:4177-4183, 1987), and streptomyces bacteriophages such as ϕ C31 (Chater *et al.*, In: Sixth International Symposium on Actinomycetales Biology, Akademiai Kiado, Budapest, Hungary, pp. 45-54, 1986). *Pseudomonas* plasmids are reviewed by John *et al.* (*Rev. Infect. Dis.* 8:693-704, 1986), and Izaki (*Jpn. J. Bacteriol.* 33:729-742, 1978).
- Preferred eukaryotic plasmids include, for example, BPV, vaccinia, SV40, 2-micron circle, and the like, or their derivatives. Such plasmids are well known in the art (Botstein *et al.*, *Miami Wnt. Symp.* 19:265-274, 1982; Broach, In: The Molecular Biology of the Yeast *Saccharomyces*: Life Cycle and Inheritance, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470, 1981; Broach, *Cell* 28:203-204, 1982; Bollon *et al.*, *J. Clin. Hematol. Oncol.* 10:39-48, 1980; Maniatis, In: Cell Biology: A Comprehensive Treatise, Vol. 3, *Gene Sequence Expression*, Academic Press, NY, pp. 563-608, 1980).
- Once the vector or nucleic acid molecule containing the construct(s) has been prepared for expression, the DNA construct(s) may be introduced into an appropriate host cell by any of a variety of suitable means, *i.e.*, transformation, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate-precipitation, direct microinjection, and the like. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene(s) results in the production of a protease of the invention, or fragments thereof. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of bromodeoxyuracil to neuroblastoma cells or the like). A variety of incubation conditions can be used to form the peptide of the present invention. The most preferred conditions are those which mimic physiological conditions.

WO 02/00860

PCT/US01/20171

Antibodies, Hybridomas, Methods of Use and Kits for Detection of Proteases

The present invention relates to an antibody having binding affinity to a protease of the invention. The protease polypeptide may have the amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60,

5 SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85,

10 SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109,

15 SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118, or a functional derivative thereof, or at least 9 contiguous amino acids thereof (preferably, at least 20, 30, 35, or 40 contiguous amino acids thereof).

The present invention also relates to an antibody having specific binding

20 affinity to a protease of the invention. Such an antibody may be isolated by comparing its binding affinity to a protease of the invention with its binding affinity to other polypeptides. Those which bind selectively to a protease of the invention would be chosen for use in methods requiring a distinction between a protease of the invention and other polypeptides. Such methods could include, but should not be

25 limited to, the analysis of altered protease expression in tissue containing other polypeptides.

The proteases of the present invention can be used in a variety of procedures and methods, such as for the generation of antibodies, for use in identifying pharmaceutical compositions, and for studying DNA/protein interaction.

WO 02/00860

PCT/US01/20171

The proteases of the present invention can be used to produce antibodies or hybridomas. One skilled in the art will recognize that if an antibody is desired, such a peptide could be generated as described herein and used as an immunogen. The antibodies of the present invention include monoclonal and polyclonal antibodies, as well fragments of these antibodies, and humanized forms. Humanized forms of the antibodies of the present invention may be generated using one of the procedures known in the art such as chimerization or CDR grafting.

The present invention also relates to a hybridoma which produces the above-described monoclonal antibody, or binding fragment thereof. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing monoclonal antibodies and hybridomas are well known in the art (Campbell, Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands, 1984; St. Groth *et al.*, *J. Immunol. Methods* 35:1-21, 1980). Any animal (mouse, rabbit, and the like) which is known to produce antibodies can be immunized with the selected polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of polypeptide used for immunization will vary based on the animal which is immunized, the antigenicity of the polypeptide and the site of injection.

The polypeptide may be modified or administered in an adjuvant in order to increase the peptide antigenicity. Methods of increasing the antigenicity of a polypeptide are well known in the art. Such procedures include coupling the antigen with a heterologous protein (such as globulin or β -galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell

WO 02/00860

PCT/US01/20171

which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175:109-124, 1988). Hybridomas secreting the desired antibodies are cloned and the class and subclass are determined using procedures known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", *supra*, 1984).

For polyclonal antibodies, antibody-containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The above-described antibodies may be detectably labeled. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, and the like), fluorescent labels (such as FITC or rhodamine, and the like), paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art, for example, see Stemberger *et al.*, *J. Histochem. Cytochem.* 18:315, 1970; Bayer *et al.*, *Meth. Enzym.* 62:308, 1979; Engval *et al.*, *Immunol.* 109:129, 1972; Goding, *J. Immunol. Meth.* 13:215, 1976. The antibodies of the present invention may be indirectly labelled by the use of secondary labelled antibodies, such as labelled anti-rabbit antibodies. The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues which express a specific peptide.

The above-described antibodies may also be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10, 1986; Jacoby *et al.*, *Meth. Enzym.* 34, Academic Press, N.Y., 1974). The immobilized antibodies of the

WO 02/00860

PCT/US01/20171

present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as in immunochromatography.

Furthermore, one skilled in the art can readily adapt currently available procedures, as well as the techniques, methods and kits disclosed herein with regard to antibodies, to generate peptides capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides (Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides", In Synthetic Peptides, A User's Guide, W.H. Freeman, NY, pp. 289-307, 1992; Kaspczak *et al.*, *Biochemistry* 28:9230-9238, 1989).

Anti-peptide peptides can be generated by replacing the basic amino acid residues found in the peptide sequences of the proteases of the invention with acidic residues, while maintaining hydrophobic and uncharged polar groups. For example, lysine, arginine, and/or histidine residues are replaced with aspartic acid or glutamic acid and glutamic acid residues are replaced by lysine, arginine or histidine.

The present invention also encompasses a method of detecting a protease polypeptide in a sample, comprising: (a) contacting the sample with an above-described antibody, under conditions such that immunocomplexes form, and (b) detecting the presence of said antibody bound to the polypeptide. In detail, the methods comprise incubating a test sample with one or more of the antibodies of the present invention and assaying whether the antibody binds to the test sample. Altered levels of a protease of the invention in a sample as compared to normal levels may indicate disease.

Conditions for incubating an antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the antibody used in the assay. One skilled in the art will recognize that any one of the commonly available immunological assay formats (such as radioimmunoassays, enzyme-linked immunosorbent assays, diffusion-based Ouchterlony, or rocket immunofluorescent assays) can readily be adapted to employ the antibodies of the present invention. Examples of such assays

WO 02/00860

PCT/US01/20171

can be found in Chard ("An Introduction to Radioimmunoassay and Related Techniques") Elsevier Science Publishers, Amsterdam, The Netherlands, 1986), Bullock *et al.* ("Techniques in Immunocytochemistry," Academic Press, Orlando, FL Vol. 1, 1982; Vol. 2, 1983; Vol. 3, 1985), Tijssen ("Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1985).

5 The immunological assay test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as blood, serum, plasma, or urine. The test samples used in the above-described method will vary
10 based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can readily be adapted in order to obtain a sample which is testable with the system utilized.

A kit contains all the necessary reagents to carry out the previously described
15 methods of detection. The kit may comprise: (i) a first container means containing an above-described antibody, and (ii) second container means containing a conjugate comprising a binding partner of the antibody and a label. In another preferred embodiment, the kit further comprises one or more other containers comprising one or more of the following: wash reagents and reagents capable of detecting the
20 presence of bound antibodies.

Examples of detection reagents include, but are not limited to, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the chromophoric, enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. The compartmentalized kit may be as described
25 above for nucleic acid probe kits. One skilled in the art will readily recognize that the antibodies described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

Isolation of Compounds Which Interact with Proteases

The present invention also relates to a method of detecting a compound capable of binding to a protease of the invention comprising incubating the compound with a protease of the invention and detecting the presence of the compound bound to the protease. The compound may be present within a complex mixture, for example, serum, body fluid, or cell extracts.

5 The present invention also relates to a method of detecting an agonist or antagonist of protease activity or protease binding partner activity comprising incubating cells that produce a protease of the invention in the presence of a compound and detecting changes in the level of protease activity or protease binding partner activity. The compounds thus identified would produce a change in activity indicative of the presence of the compound. The compound may be present within a complex mixture, for example, serum, body fluid, or cell extracts. Once the compound is identified it can be isolated using techniques well known in the art.

10 The present invention also encompasses a method of modulating protease associated activity in a mammal comprising administering to said mammal an agonist or antagonist to a protease of the invention in an amount sufficient to effect said modulation. A method of treating diseases in a mammal with an agonist or antagonist of the activity of one of the proteases of the invention comprising administering the agonist or antagonist to a mammal in an amount sufficient to agonize or antagonize protease-associated functions is also encompassed in the present application.

15 In an effort to discover novel treatments for diseases, biomedical researchers and chemists have designed, synthesized, and tested molecules that inhibit the function of proteases. Some small organic molecules form a class of compounds that modulate the function of protein proteases.

25 Examples of molecules that have been reported to inhibit the function of protein proteases include, but are not limited to, phenylmethylsulfonyl fluoride (PMSF), diisopropylfluorophosphate (DFP) (chapter 3, Barrett *et al.*, Handbook of

- Proteolytic Enzymes, 1998, Academic Press, San Diego), 3,4-dichloroisocoumarin (DCI) (*Id.*, chapter 16), serpins (*Id.*, chapter 37), E-64 (*trans*-epoxysuccinyl L-leucylamido-(4-guanidino) butane) (*Id.*, chapter 188), peptidyl-diazomethanes, peptidyl-*O*-acyl-hydroxamates, epoxysuccinyl-peptides (*Id.*, chapter 210), DAN, 5 EPNP (1,2-epoxy-3(p-nitrophenoxy)propane) (*Id.*, chapter 298), thiorphan (di-3-Mercapto-2-benzylpropanoyl-glycine) (*Id.*, chapter 362), CGS 26303, PD 069185 (*Id.*, chapter 363), and COT989-00 (N-4-hydroxy-N1-[1-(s)-(4-aminosulfonyl)phenylethyl-aminocarboxyl-2-cyclohexylethyl]-2R-[4-methyl)phenylpropyl]succinamide) (*Id.*, chapter 401). Other protease inhibitors 10 include, but are not limited to, aprotinin, amastatin, antipain, calcineurin autoinhibitory fragment, and histatin 5 (*Id.*). Preferably, these inhibitors will have molecular weights from 100 to 200 daltons, from 200 to 300 daltons, from 300 to 400 daltons, from 400 to 600 daltons, from 600 to 1000 daltons, from 1000 to 2000 daltons, from 2000 to 4000 daltons, and from 4000 to 8000 daltons.
- 15 Compounds that can traverse cell membranes and are resistant to acid hydrolysis are potentially advantageous as therapeutics as they can become highly bioavailable after being administered orally to patients. However, many of these protease inhibitors only weakly inhibit the function of proteases. In addition, many inhibit a variety of proteases and will therefore cause multiple side-effects as 20 therapeutics for diseases.

Transgenic Animals.

- A variety of methods are available for the production of transgenic animals associated with this invention. DNA can be injected into the pronucleus of a 25 fertilized egg before fusion of the male and female pronuclei, or injected into the nucleus of an embryonic cell (*e.g.*, the nucleus of a two-cell embryo) following the initiation of cell division (Brinster *et al.*, *Proc. Nat. Acad. Sci. USA* 82:4438-4442, 1985). Embryos can be infected with viruses, especially retroviruses, modified to carry inorganic-ion receptor nucleotide sequences of the invention.

Pluripotent stem cells derived from the inner cell mass of the embryo and stabilized in culture can be manipulated in culture to incorporate nucleotide sequences of the invention. A transgenic animal can be produced from such cells through implantation into a blastocyst that is implanted into a foster mother and
5 allowed to come to term. Animals suitable for transgenic experiments can be obtained from standard commercial sources such as Charles River (Wilmington, MA), Taconic (Germantown, NY), Harlan Sprague Dawley (Indianapolis, IN), etc.

The procedures for manipulation of the rodent embryo and for microinjection of DNA into the pronucleus of the zygote are well known to those of ordinary skill
10 in the art (Hogan *et al.*, *supra*). Microinjection procedures for fish, amphibian eggs and birds are detailed in Houdebine and Chourrout (*Experientia* 47:897-905, 1991). Other procedures for introduction of DNA into tissues of animals are described in U.S. Patent No. 4,945,050 (Sanford *et al.*, July 30, 1990).

By way of example only, to prepare a transgenic mouse, female mice are
15 induced to superovulate. Females are placed with males, and the mated females are sacrificed by CO₂ asphyxiation or cervical dislocation and embryos are recovered from excised oviducts. Surrounding cumulus cells are removed. Pronuclear embryos are then washed and stored until the time of injection. Randomly cycling adult female mice are paired with vasectomized males. Recipient females are mated
20 at the same time as donor females. Embryos then are transferred surgically. The procedure for generating transgenic rats is similar to that of mice (Hammer *et al.*, *Cell* 63:1099-1112, 1990).

Methods for the culturing of embryonic stem (ES) cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using
25 methods such as electroporation, calcium phosphate/DNA precipitation and direct injection also are well known to those of ordinary skill in the art (Teratocarcinomas and Embryonic Stem Cells, A Practical Approach, E.J. Robertson, ed., IRL Press, 1987).

WO 02/00860

PCT/US01/20171

In cases involving random gene integration, a clone containing the sequence(s) of the invention is co-transfected with a gene encoding resistance. Alternatively, the gene encoding neomycin resistance is physically linked to the sequence(s) of the invention. Transfection and isolation of desired clones are carried out by any one of several methods well known to those of ordinary skill in the art (E.J. Robertson, *supra*).

DNA molecules introduced into ES cells can also be integrated into the chromosome through the process of homologous recombination (Capecchi, *Science* 244:1288-1292, 1989). Methods for positive selection of the recombination event (*i.e.*, neo resistance) and dual positive-negative selection (*i.e.*, neo resistance and gancyclovir resistance) and the subsequent identification of the desired clones by PCR have been described by Capecchi, *supra* and Joyner *et al.* (*Nature* 338:153-156, 1989), the teachings of which are incorporated herein in their entirety including any drawings. The final phase of the procedure is to inject targeted ES cells into blastocysts and to transfer the blastocysts into pseudopregnant females. The resulting chimeric animals are bred and the offspring are analyzed by Southern blotting to identify individuals that carry the transgene. Procedures for the production of non-rodent mammals and other animals have been discussed by others (Houdebine and Chourrout, *supra*; Pursel *et al.*, *Science* 244:1281-1288, 1989; and Simms *et al.*, *BioTechnology* 6:179-183, 1988).

Thus, the invention provides transgenic, nonhuman mammals containing a transgene encoding a protease of the invention or a gene affecting the expression of the protease. Such transgenic nonhuman mammals are particularly useful as an *in vivo* test system for studying the effects of introduction of a protease, or regulating the expression of a protease (*i.e.*, through the introduction of additional genes, antisense nucleic acids, or ribozymes).

A "transgenic animal" is an animal having cells that contain DNA which has been artificially inserted into a cell, which DNA becomes part of the genome of the animal which develops from that cell. Preferred transgenic animals are primates,

WO 02/00860

PCT/US01/20171

mice, rats, cows, pigs, horses, goats, sheep, dogs and cats. The transgenic DNA may encode human proteases. Native expression in an animal may be reduced by providing an amount of antisense RNA or DNA effective to reduce expression of the receptor.

5

Gene Therapy

Proteases or their genetic sequences will also be useful in gene therapy (reviewed in Miller, *Nature* 357:455-460, 1992). Miller states that advances have resulted in practical approaches to human gene therapy that have demonstrated positive initial results. The basic science of gene therapy is described in Mulligan (*Science* 260:926-931, 1993).

In one preferred embodiment, an expression vector containing a protease coding sequence is inserted into cells, the cells are grown *in vitro* and then infused in large numbers into patients. In another preferred embodiment, a DNA segment containing a promoter of choice (for example a strong promoter) is transferred into cells containing an endogenous gene encoding proteases of the invention in such a manner that the promoter segment enhances expression of the endogenous protease gene (for example, the promoter segment is transferred to the cell such that it becomes directly linked to the endogenous protease gene).

The gene therapy may involve the use of an adenovirus containing protease cDNA targeted to a tumor, systemic protease increase by implantation of engineered cells, injection with protease-encoding virus, or injection of naked protease DNA into appropriate tissues.

Target cell populations may be modified by introducing altered forms of one or more components of the protein complexes in order to modulate the activity of such complexes. For example, by reducing or inhibiting a complex component activity within target cells, an abnormal signal transduction event(s) leading to a condition may be decreased, inhibited, or reversed. Deletion or missense mutants of a component, that retain the ability to interact with other components of the protein

complexes but cannot function in signal transduction, may be used to inhibit an abnormal, deleterious signal transduction event.

Expression vectors derived from viruses such as retroviruses, vaccinia virus, adenovirus, adeno-associated virus, herpes viruses, several RNA viruses, or bovine papilloma virus, may be used for delivery of nucleotide sequences (e.g., cDNA) 5 encoding recombinant protease of the invention protein into the targeted cell population (e.g., tumor cells). Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors containing coding sequences (Maniatis *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor 10 Laboratory, N.Y., 1989; Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y., 1989). Alternatively, recombinant nucleic acid molecules encoding protein sequences can be used as naked DNA or in a reconstituted system e.g., liposomes or other lipid systems for delivery to target cells (e.g., Felgner *et al.*, *Nature* 337:387-8, 1989). Several other 15 methods for the direct transfer of plasmid DNA into cells exist for use in human gene therapy and involve targeting the DNA to receptors on cells by complexing the plasmid DNA to proteins (Miller, *supra*).

In its simplest form, gene transfer can be performed by simply injecting minute amounts of DNA into the nucleus of a cell, through a process of 20 microinjection (Capecchi, *Cell* 22:479-88, 1980). Once recombinant genes are introduced into a cell, they can be recognized by the cell's normal mechanisms for transcription and translation, and a gene product will be expressed. Other methods have also been attempted for introducing DNA into larger numbers of cells. These methods include: transfection, wherein DNA is precipitated with calcium phosphate 25 and taken into cells by pinocytosis (Chen *et al.*, *Mol. Cell Biol.* 7:2745-52, 1987); electroporation, wherein cells are exposed to large voltage pulses to introduce holes into the membrane (Chu *et al.*, *Nucleic Acids Res.* 15:1311-26, 1987); lipofection/liposome fusion, wherein DNA is packaged into lipophilic vesicles which fuse with a target cell (Felgner *et al.*, *Proc. Natl. Acad. Sci. USA.* 84:7413-

7417, 1987); and particle bombardment using DNA bound to small projectiles (Yang *et al.*, *Proc. Natl. Acad. Sci.* 87:9568-9572, 1990). Another method for introducing DNA into cells is to couple the DNA to chemically modified proteins.

It has also been shown that adenovirus proteins are capable of destabilizing
5 endosomes and enhancing the uptake of DNA into cells. The admixture of adenovirus to solutions containing DNA complexes, or the binding of DNA to polylysine covalently attached to adenovirus using protein crosslinking agents substantially improves the uptake and expression of the recombinant gene (Curiel *et al.*, *Am. J. Respir. Cell. Mol. Biol.*, 6:247-52, 1992).

As used herein "gene transfer" means the process of introducing a foreign
0 nucleic acid molecule into a cell. Gene transfer is commonly performed to enable the expression of a particular product encoded by the gene. The product may include a protein, polypeptide, anti-sense DNA or RNA, or enzymatically active RNA. Gene transfer can be performed in cultured cells or by direct administration
5 into animals. Generally gene transfer involves the process of nucleic acid contact with a target cell by non-specific or receptor mediated interactions, uptake of nucleic acid into the cell through the membrane or by endocytosis, and release of nucleic acid into the cyto-plasm from the plasma membrane or endosome. Expression may require, in addition, movement of the nucleic acid into the nucleus of the cell and
10 binding to appropriate nuclear factors for transcription.

As used herein "gene therapy" is a form of gene transfer and is included
within the definition of gene transfer as used herein and specifically refers to gene transfer to express a therapeutic product from a cell *in vivo* or *in vitro*. Gene transfer
can be performed *ex vivo* on cells which are then transplanted into a patient, or can
15 be performed by direct administration of the nucleic acid or nucleic acid-protein complex into the patient.

In another preferred embodiment, a vector having nucleic acid sequences encoding a protease polypeptide is provided in which the nucleic acid sequence is expressed only in specific tissue. Methods of achieving tissue-specific gene

WO 02/00860

PCT/US01/20171

expression are set forth in International Publication No. WO 93/09236, filed November 3, 1992 and published May 13, 1993.

In all of the preceding vectors set forth above, a further aspect of the invention is that the nucleic acid sequence contained in the vector may include
5 additions, deletions or modifications to some or all of the sequence of the nucleic acid, as defined above.

Expression, including over-expression, of a protease polypeptide of the invention can be inhibited by administration of an antisense molecule that binds to and inhibits expression of the mRNA encoding the polypeptide. Alternatively, expression
10 can be inhibited in an analogous manner using a ribozyme that cleaves the mRNA. General methods of using antisense and ribozyme technology to control gene expression, or of gene therapy methods for expression of an exogenous gene in this manner are well known in the art. Each of these methods utilizes a system, such as a vector, encoding either an antisense or ribozyme transcript of a protease polypeptide of
15 the invention.

The term "*ribozyme*" refers to an RNA structure of one or more RNAs having catalytic properties. Ribozymes generally exhibit endonuclease, ligase or polymerase activity. Ribozymes are structural RNA molecules which mediate a number of RNA self-cleavage reactions. Various types of trans-acting ribozymes,
20 including "hammerhead" and "hairpin" types, which have different secondary structures, have been identified. A variety of ribozymes have been characterized. See, for example, U.S. Pat. Nos. 5,246,921, 5,225,347, 5,225,337 and 5,149,796. Mixed ribozymes comprising deoxyribo and ribooligonucleotides with catalytic activity have been described. Perreault, *et al.*, *Nature*, 344:565-567 (1990).

As used herein, "antisense" refers of nucleic acid molecules or their
25 derivatives which specifically hybridize, *e.g.*, bind, under cellular conditions, with the genomic DNA and/or cellular mRNA encoding a protease polypeptide of the invention, so as to inhibit expression of that protein, for example, by inhibiting transcription and/or translation. The binding may be by conventional base pair

WO 02/00860

PCT/US01/20171

complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix.

In one aspect, the antisense construct is a nucleic acid which is generated *ex vivo* and that, when introduced into the cell, can inhibit gene expression by, without limitation, hybridizing with the mRNA and/or genomic sequences of a protease polynucleotide of the invention.

Antisense approaches can involve the design of oligonucleotides (either DNA or RNA) that are complementary to protease polypeptide mRNA and are based on the protease polynucleotides of the invention, including SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ ID NO:59. The antisense oligonucleotides will bind to the protease polypeptide mRNA transcripts and prevent translation.

Although absolute complementarity is preferred, it is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, means a sequence having sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid. Generally, the longer the hybridizing

WO 02/00860

PCT/US01/20171

nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

5 In general, oligonucleotides that are complementary to the 5' end of the message, e.g., the 5' untranslated sequence up to and including the AUG initiation codon, should work most efficiently at inhibiting translation. However, sequences complementary to the 3' untranslated sequences of mRNAs have been shown to be effective at inhibiting translation of mRNAs as well. (Wagner, R. (1994) Nature
10 372:333). Antisense oligonucleotides complementary to mRNA coding regions are less efficient inhibitors of translation but could be used in accordance with the invention. Whether designed to hybridize to the 5', 3' or coding region of the protease polypeptide mRNA, antisense nucleic acids should be at least six nucleotides in length, and are preferably less than about 100 and more preferably
15 less than about 50 or 30 nucleotides in length. Typically they should be between 10 and 25 nucleotides in length. Such principles will inform the practitioner in selecting the appropriate oligonucleotides. In preferred embodiments, the antisense sequence is selected from an oligonucleotide sequence that comprises, consists of, or consists essentially of about 10-30, and more preferably 15-25, contiguous
20 nucleotide bases of a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID
25 NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID

WO 02/00860

PCT/US01/20171

NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ ID NO:59 or domains thereof.

In another preferred embodiment, the invention includes an isolated,
 5 enriched or purified nucleic acid molecule comprising, consisting of or consisting essentially of about 10-30, and more preferably 15-25 contiguous nucleotide bases of a nucleic acid sequence that encodes a polypeptide that selected from the group consisting of SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ
 10 ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ
 15 ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID
 20 NO:117 and SEQ ID NO:118.

Using the sequences of the present invention, antisense oligonucleotides can be designed. Such antisense oligonucleotides would be administered to cells expressing the target protease and the levels of the target RNA or protein with that of an internal control RNA or protein would be compared. Results obtained using
 25 the antisense oligonucleotide would also be compared with those obtained using a suitable control oligonucleotide. A preferred control oligonucleotide is an oligonucleotide of approximately the same length as the test oligonucleotide. Those antisense oligonucleotides resulting in a reduction in levels of target RNA or protein would be selected.

WO 02/00860

PCT/US01/20171

The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810, published Dec. 15, 1988) or the blood-brain barrier (see, e.g., PCT Publication No. WO 89/10134, published Apr. 25, 1988), hybridization-triggered cleavage agents. (See, e.g., Krol *et al.* (1988) *BioTechniques* 6:958-976) or intercalating agents. (See, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

The antisense oligonucleotide may comprise at least one modified base moiety which is selected from moieties such as 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, and 5-(carboxyhydroxyethyl) uracil. The antisense oligonucleotide may also comprise at least one modified sugar moiety selected from the group including but not limited to arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal or analog thereof. (see also U.S. Pat. Nos. 5,176,996; 5,264,564; and 5,256,775)

In yet a further embodiment, the antisense oligonucleotide is an α -anomeric oligonucleotide. An α -anomeric oligonucleotide forms specific double-stranded

WO 02/00860

PCT/US01/20171

hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gautier *et al.* (1987) *Nucl. Acids Res.* 15:6625-6641). The oligonucleotide is a 2'-O-methylribonucleotide (Inoue *et al.* (1987) *Nucl. Acids Res.* 15:6131-6148), or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) FEBS Lett. 215:327-330).

Also suitable are peptidyl nucleic acids, which are polypeptides such as polyserine, polythreonine, etc. including copolymers containing various amino acids, which are substituted at side-chain positions with nucleic acids (T,A,G,C,U). Chains of such polymers are able to hybridize through complementary bases in the same manner as natural DNA/RNA. Alternatively, an antisense construct of the present invention can be delivered, for example, as an expression plasmid or vector that, when transcribed in the cell, produces RNA complementary to at least a unique portion of the cellular mRNA which encodes a protease polypeptide of the invention.

While antisense nucleotides complementary to the protease polypeptide coding region sequence can be used, those complementary to the transcribed untranslated region are most preferred.

In another preferred embodiment, a method of gene replacement is set forth. "Gene replacement" as used herein means supplying a nucleic acid sequence which is capable of being expressed *in vivo* in an animal and thereby providing or augmenting the function of an endogenous gene which is missing or defective in the animal.

Pharmaceutical Formulations And Routes Of Administration

The compounds described herein, including protease polypeptides of the invention, antisense molecules, ribozymes, and any other compound that modulates the activity of a protease polypeptide of the invention, can be administered to a human patient *per se*, or in pharmaceutical compositions where it is mixed with other active ingredients, as in combination therapy, or suitable carriers or

WO 02/00860

PCT/US01/20171

excipient(s). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition.

A. Routes Of Administration

5 Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intravenous, intramedullary injections, as well as intrathecal, direct intraventricular, intraperitoneal, intranasal, or intraocular injections.

10 Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a solid tumor, often in a depot or sustained release formulation.

Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with tumor-specific antibody. The liposomes will
15 be targeted to and taken up selectively by the tumor.

B. Composition/Formulation

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying,
20 encapsulating, entrapping or lyophilizing processes.

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used
25 pharmaceutically. Proper formulation is dependent upon the route of administration chosen.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration,

WO 02/00860

PCT/US01/20171

penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Suitable carriers include excipients such as, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethylcellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration.

For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of *e.g.* gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

WO 02/00860

PCT/US01/20171

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a cosolvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The cosolvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:D5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, *e.g.* polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose.

Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater

WO 02/00860

PCT/US01/20171

toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may,

5 depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are
0 not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols.

Many of the protease modulating compounds of the invention may be provided as salts with pharmaceutically compatible counterions. Pharmaceutically compatible salts may be formed with many acids, including but not limited to
5 hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

C. Effective Dosage

Pharmaceutical compositions suitable for use in the present invention include
10 compositions where the active ingredients are contained in an amount effective to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount of compound effective to prevent, alleviate or ameliorate symptoms of disease or prolong the survival of the subject being treated.

Determination of a therapeutically effective amount is well within the capability of
15 those skilled in the art, especially in light of the detailed disclosure provided herein.

For any compound used in the methods of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of

WO 02/00860

PCT/US01/20171

the test compound which achieves a half-maximal inhibition of the protease activity). Such information can be used to more accurately determine useful doses in humans.

Toxicity and therapeutic efficacy of the compounds described herein can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD₅₀ and ED₅₀. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See *e.g.*, Fingl *et al.*, 1975, in The Pharmacological Basis of Therapeutics, Ch. 1 p.I).

Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the protease modulating effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data; *e.g.*, the concentration necessary to achieve 50-90% inhibition of the protease using the assays described herein. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%.

WO 02/00860

PCT/US01/20171

In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

D. Packaging

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. The pack or dispenser may also be accompanied with a notice associated with the container in form prescribed by a governmental agency regulating the manufacture, use, or sale of pharmaceuticals, which notice is reflective of approval by the agency of the form of the polynucleotide for human or veterinary administration. Such notice, for example, may be the labeling approved by the U.S. Food and Drug Administration for prescription drugs, or the approved product insert. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition. Suitable conditions indicated on the label may include treatment of a tumor, inhibition of angiogenesis, treatment of fibrosis, diabetes, and the like.

Functional Derivatives

Also provided herein are functional derivatives of a polypeptide or nucleic acid of the invention. By "functional derivative" is meant a "chemical derivative," "fragment," or "variant," of the polypeptide or nucleic acid of the invention, which terms are defined below. A functional derivative retains at least a portion of the function of the protein, for example reactivity with an antibody specific for the protein, enzymatic activity or binding activity mediated through noncatalytic

WO 02/00860

PCT/US01/20171

domains, which permits its utility in accordance with the present invention. It is well known in the art that due to the degeneracy of the genetic code numerous different nucleic acid sequences can code for the same amino acid sequence. Equally, it is also well known in the art that conservative changes in amino acid can be made to arrive at a protein or polypeptide that retains the functionality of the original. In both cases, all permutations are intended to be covered by this disclosure.

Included within the scope of this invention are the functional equivalents of the herein-described isolated nucleic acid molecules. The degeneracy of the genetic code permits substitution of certain codons by other codons that specify the same amino acid and hence would give rise to the same protein. The nucleic acid sequence can vary substantially since, with the exception of methionine and tryptophan, the known amino acids can be coded for by more than one codon. Thus, portions or all of the genes of the invention could be synthesized to give a nucleic acid sequence significantly different from one selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ ID NO:59. The encoded amino acid sequence thereof would, however, be preserved.

WO 02/00860

PCT/US01/20171

In addition, the nucleic acid sequence may comprise a nucleotide sequence which results from the addition, deletion or substitution of at least one nucleotide to the 5'-end and/or the 3'-end of the nucleic acid formula selected from the group consisting of those set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ ID NO:59, or a derivative thereof. Any nucleotide or polynucleotide may be used in this regard, provided that its addition, deletion or substitution does not alter the amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113,

WO 02/00860

PCT/US01/20171

SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 which is encoded by the nucleotide sequence. For example, the present invention is intended to include any nucleic acid sequence resulting from the addition of ATG as an initiation codon at the 5'-end of the inventive nucleic acid sequence or its derivative, or from the addition of TTA, TAG or TGA as a
5 termination codon at the 3'-end of the inventive nucleotide sequence or its derivative. Moreover, the nucleic acid molecule of the present invention may, as necessary, have restriction endonuclease recognition sites added to its 5'-end and/or 3'-end.

10 Such functional alterations of a given nucleic acid sequence afford an opportunity to promote secretion and/or processing of heterologous proteins encoded by foreign nucleic acid sequences fused thereto. All variations of the nucleotide sequence of the protease genes of the invention and fragments thereof permitted by the genetic code are, therefore, included in this invention.

15 Further, it is possible to delete codons or to substitute one or more codons with codons other than degenerate codons to produce a structurally modified polypeptide, but one which has substantially the same utility or activity as the polypeptide produced by the unmodified nucleic acid molecule. As recognized in the art, the two polypeptides are functionally equivalent, as are the two nucleic acid
20 molecules that give rise to their production, even though the differences between the nucleic acid molecules are not related to the degeneracy of the genetic code.

A "chemical derivative" of the complex contains additional chemical moieties not normally a part of the protein. Covalent modifications of the protein or peptides are included within the scope of this invention. Such modifications may be
25 introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues, as described below.

Cysteiny residues most commonly are reacted with α -haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteiny residues also are derivatized by reaction with bromotrifluoroacetone, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysiny and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysiny residues. Other suitable reagents for derivatizing primary amine containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginy residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pKa of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine α -amino group.

Tyrosyl residues are well-known targets of modification for introduction of spectral labels by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

WO 02/00860

PCT/US01/20171

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimide ($R^1-N-C-N-R^2$) such as 1-cyclohexyl-3-(2-morpholinyl)(4-ethyl) carbodiimide or 1-ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to

5 asparaginy and glutaminy residues by reaction with ammonium ions.

Glutaminy and asparaginy residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Derivatization with bifunctional agents is useful, for example, for cross-linking the component peptides of the protein to each other or to other proteins in a complex to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, for example, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl] dithiolpropionimide yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-

15

20 insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the

25 α -amino groups of lysine, arginine, and histidine side chains (Creighton, T.E., Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the stability, solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein complex and the like. Moieties capable of mediating such effects are disclosed, for example, in Remington's

5 Pharmaceutical Sciences, 18th ed., Mack Publishing Co., Easton, PA (1990).

The term "fragment" is used to indicate a polypeptide derived from the amino acid sequence of the proteins, of the complexes having a length less than the full-length polypeptide from which it has been derived. Such a fragment may, for example, be produced by proteolytic cleavage of the full-length protein. Preferably,
10 the fragment is obtained recombinantly by appropriately modifying the DNA sequence encoding the proteins to delete one or more amino acids at one or more sites of the C-terminus, N-terminus, and/or within the native sequence. Fragments of a protein are useful for screening for substances that act to modulate signal transduction, as described herein. It is understood that such fragments may retain
15 one or more characterizing portions of the native complex. Examples of such retained characteristics include: catalytic activity; substrate specificity; interaction with other molecules in the intact cell; regulatory functions; or binding with an antibody specific for the native complex, or an epitope thereof.

Another functional derivative intended to be within the scope of the present
20 invention is a "variant" polypeptide which either lacks one or more amino acids or contains additional or substituted amino acids relative to the native polypeptide. The variant may be derived from a naturally occurring complex component by appropriately modifying the protein DNA coding sequence to add, remove, and/or to modify codons for one or more amino acids at one or more sites of the C-terminus,
25 N-terminus, and/or within the native sequence. It is understood that such variants having added, substituted and/or additional amino acids retain one or more characterizing portions of the native protein, as described above.

A functional derivative of a protein with deleted, inserted and/or substituted amino acid residues may be prepared using standard techniques well-known to those

WO 02/00860

PCT/US01/20171

of ordinary skill in the art. For example, the modified components of the functional derivatives may be produced using site-directed mutagenesis techniques (as exemplified by Adelman *et al.*, 1983, *DNA* 2:183) wherein nucleotides in the DNA coding the sequence are modified such that a modified coding sequence is modified, and thereafter expressing this recombinant DNA in a prokaryotic or eukaryotic host cell, using techniques such as those described above. Alternatively, proteins with amino acid deletions, insertions and/or substitutions may be conveniently prepared by direct chemical synthesis, using methods well-known in the art. The functional derivatives of the proteins typically exhibit the same qualitative biological activity as the native proteins.

TABLES AND DESCRIPTION THEREOF

This patent describes novel protease identified in databases of genomic sequence. The results are summarized in four tables, which are described below.

Table 1 documents the name of each gene, the classification of each gene, the positions of the open reading frames within the sequence, and the length of the corresponding peptide. From left to right the data presented is as follows: "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family", "NA_length", "ORF Start", "ORF End", "ORF Length", and "AA_length". "Gene name" refers to name given the sequence encoding the protease enzyme. Each gene is represented by "SGPr" designation followed by an arbitrary number. The SGPr name usually represents multiple overlapping sequences built into a single contiguous sequence (a "contig"). The "ID#na" and "ID#aa" refer to the identification numbers given each nucleic acid and amino acid sequence in this patent application. "FL/Cat" refers to the length of the gene, with FL indicating full length, and "Cat" indicating that only the catalytic domain is presented. "Partial" in this column indicates that the sequence encodes a partial catalytic domain. "Superfamily" identifies whether the gene is a protease. "Group" and "Family"

WO 02/00860

PCT/US01/20171

- refer to the protease classification defined by sequence homology. "NA_length" refers to the length in nucleotides of the corresponding nucleic acid sequence. "ORF start" refers to the beginning nucleotide of the open reading frame. "ORF end" refers to the last nucleotide of the open reading frame, including the stop codon.
- 5 "ORF length" refers to the length in nucleotides of the open reading frame (including the stop codon). "AA length" refers to the length in amino acids of the peptide encoded in the corresponding nucleic acid sequence.

WO 02/00860

PCT/US01/20171

Table 1 - Proteases

Gene Name	Accession	Size	Spont. Mut.	Group	Family	AA Length	ORF Start	ORF End	ORF Length	AA Length
SPW141	1	81	FL	Protease	Protease	155	1	155	155	24
SPW142	2	81	FL	Protease	Protease	155	1	155	155	24
SPW143	3	81	FL	Protease	Protease	155	1	155	155	24
SPW144	4	81	FL	Protease	Protease	155	1	155	155	24
SPW145	5	81	FL	Protease	Protease	155	1	155	155	24
SPW146	6	81	FL	Protease	Protease	155	1	155	155	24
SPW147	7	81	FL	Protease	Protease	155	1	155	155	24
SPW148	8	81	FL	Protease	Protease	155	1	155	155	24
SPW149	9	81	FL	Protease	Protease	155	1	155	155	24
SPW150	10	81	FL	Protease	Protease	155	1	155	155	24
SPW151	11	81	FL	Protease	Protease	155	1	155	155	24
SPW152	12	81	FL	Protease	Protease	155	1	155	155	24
SPW153	13	81	FL	Protease	Protease	155	1	155	155	24
SPW154	14	81	FL	Protease	Protease	155	1	155	155	24
SPW155	15	81	FL	Protease	Protease	155	1	155	155	24
SPW156	16	81	FL	Protease	Protease	155	1	155	155	24
SPW157	17	81	FL	Protease	Protease	155	1	155	155	24
SPW158	18	81	FL	Protease	Protease	155	1	155	155	24
SPW159	19	81	FL	Protease	Protease	155	1	155	155	24
SPW160	20	81	FL	Protease	Protease	155	1	155	155	24
SPW161	21	81	FL	Protease	Protease	155	1	155	155	24
SPW162	22	81	FL	Protease	Protease	155	1	155	155	24
SPW163	23	81	FL	Protease	Protease	155	1	155	155	24
SPW164	24	81	FL	Protease	Protease	155	1	155	155	24
SPW165	25	81	FL	Protease	Protease	155	1	155	155	24
SPW166	26	81	FL	Protease	Protease	155	1	155	155	24
SPW167	27	81	FL	Protease	Protease	155	1	155	155	24
SPW168	28	81	FL	Protease	Protease	155	1	155	155	24
SPW169	29	81	FL	Protease	Protease	155	1	155	155	24
SPW170	30	81	FL	Protease	Protease	155	1	155	155	24
SPW171	31	81	FL	Protease	Protease	155	1	155	155	24
SPW172	32	81	FL	Protease	Protease	155	1	155	155	24
SPW173	33	81	FL	Protease	Protease	155	1	155	155	24
SPW174	34	81	FL	Protease	Protease	155	1	155	155	24
SPW175	35	81	FL	Protease	Protease	155	1	155	155	24
SPW176	36	81	FL	Protease	Protease	155	1	155	155	24
SPW177	37	81	FL	Protease	Protease	155	1	155	155	24
SPW178	38	81	FL	Protease	Protease	155	1	155	155	24
SPW179	39	81	FL	Protease	Protease	155	1	155	155	24
SPW180	40	81	FL	Protease	Protease	155	1	155	155	24
SPW181	41	81	FL	Protease	Protease	155	1	155	155	24
SPW182	42	81	FL	Protease	Protease	155	1	155	155	24
SPW183	43	81	FL	Protease	Protease	155	1	155	155	24
SPW184	44	81	FL	Protease	Protease	155	1	155	155	24
SPW185	45	81	FL	Protease	Protease	155	1	155	155	24
SPW186	46	81	FL	Protease	Protease	155	1	155	155	24
SPW187	47	81	FL	Protease	Protease	155	1	155	155	24
SPW188	48	81	FL	Protease	Protease	155	1	155	155	24
SPW189	49	81	FL	Protease	Protease	155	1	155	155	24
SPW190	50	81	FL	Protease	Protease	155	1	155	155	24
SPW191	51	81	FL	Protease	Protease	155	1	155	155	24
SPW192	52	81	FL	Protease	Protease	155	1	155	155	24
SPW193	53	81	FL	Protease	Protease	155	1	155	155	24
SPW194	54	81	FL	Protease	Protease	155	1	155	155	24
SPW195	55	81	FL	Protease	Protease	155	1	155	155	24
SPW196	56	81	FL	Protease	Protease	155	1	155	155	24
SPW197	57	81	FL	Protease	Protease	155	1	155	155	24
SPW198	58	81	FL	Protease	Protease	155	1	155	155	24
SPW199	59	81	FL	Protease	Protease	155	1	155	155	24
SPW200	60	81	FL	Protease	Protease	155	1	155	155	24

WO 02/00860

PCT/US01/20171

Table 2 lists the following features of the genes described in this patent application: chromosomal localization, single nucleotide polymorphisms (SNPs), representation in dbEST, and repeat regions. From left to right the data presented is as follows: "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family", "Chromosome", "SNPs", "dbEST_hits", & "Repeats". The contents of the first 7 columns (i.e., "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family") are as described above for Table 1.

"Chromosome" refers to the cytogenetic localization of the gene. Information in the "SNPs" column describes the nucleic acid position and degenerate nature of candidate single nucleotide polymorphisms (SNPs; please see table of polymorphism below). These SNPs were identified by blastn of the DNA sequence against the database of single nucleotide polymorphisms maintained at NCBI (<http://www.ncbi.nlm.nih.gov/SNP/snpblastByChr.html>). "dbEST hits" lists accession numbers of entries in the public database of ESTs (dbEST, <http://www.ncbi.nlm.nih.gov/dbEST/index.html>) that contain at least 150 bp of 100% identity to the corresponding gene. These ESTs were identified by blastn of dbEST. "Repeats" contains information about the location of short sequences, approximately 20 bp in length, that are of low complexity and that are present in several distinct genes.

[illegible]

Table 3 lists the extent and the boundaries of the protease catalytic domains, and other protein domains. The column headings are: "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Profile_start", "Profile_end", "Domain_start", "Domain_end", and

5 "Profile". The contents of the first 7 columns (i.e., "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family") are as described above for Table 1. "Profile Start", "Profile End", "Domain Start" and "Domain End" refer to data obtained using a Hidden-Markov Model to define catalytic range boundaries. The boundaries of the catalytic domain(s) within the overall protein are noted in the

10 "Domain Start" and "Domain End" columns. "Profile" indicates the identity of the Hidden Markov Model used to identify catalytic and other types of domains within the protein sequence. Whether the HMMR search was done with a complete ("Global") or Smith Waterman ("Local") model, is described below. Starting from a multiple sequence alignment of catalytic domains, two hidden Markov models were

15 built. One of them allows for partial matches to the catalytic domain; this is a "local" HMM, similar to Smith-Waterman alignments in sequence matching. The other model allows matches only to the complete catalytic domain; this is a "global" HMM similar to Needleman-Wunsch alignments in sequence matching. The Smith Waterman local model is more specific, allowing for fragmentary matches to the

20 catalytic domain whereas the global "complete" model is more sensitive, allowing for remote homologue identification. These domains were identified using PFAM (<http://pfam.wustl.edu/hmmsearch.shtml>) models, a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. Version 5.5 of Pfam (Sept 2000) contains alignments and models for 2478

25 protein families (<http://pfam.wustl.edu/faq.shtml>). The PFAM alignments were downloaded from <http://pfam.wustl.edu/hmmsearch.shtml> and the HMMr searches were run locally on a Timelogic computer (TimeLogic Corporation, Incline Village, NV). A number of proteins have more than one domain recognized by the HMM searches. For these proteins, the domains have been listed in separate rows.

PCT/US01/20171

Table 3 - Protease Domains, Other Domains

Gene Name	Gen	DBP1	FL1	FL2	FL3	FL4	FL5	FL6	FL7	FL8	FL9	FL10	FL11	FL12	FL13	FL14	FL15	FL16	FL17	FL18	FL19	FL20	FL21	FL22	FL23	FL24	FL25	FL26	FL27	FL28	FL29	FL30	FL31	FL32	FL33	FL34	FL35	FL36	FL37	FL38	FL39	FL40	FL41	FL42	FL43	FL44	FL45	FL46	FL47	FL48	FL49	FL50	FL51	FL52	FL53	FL54	FL55	FL56	FL57	FL58	FL59	FL60	FL61	FL62	FL63	FL64	FL65	FL66	FL67	FL68	FL69	FL70	FL71	FL72	FL73	FL74	FL75	FL76	FL77	FL78	FL79	FL80	FL81	FL82	FL83	FL84	FL85	FL86	FL87	FL88	FL89	FL90	FL91	FL92	FL93	FL94	FL95	FL96	FL97	FL98	FL99	FL100	FL101	FL102	FL103	FL104	FL105	FL106	FL107	FL108	FL109	FL110	FL111	FL112	FL113	FL114	FL115	FL116	FL117	FL118	FL119	FL120	FL121	FL122	FL123	FL124	FL125	FL126	FL127	FL128	FL129	FL130	FL131	FL132	FL133	FL134	FL135	FL136	FL137	FL138	FL139	FL140	FL141	FL142	FL143	FL144	FL145	FL146	FL147	FL148	FL149	FL150	FL151	FL152	FL153	FL154	FL155	FL156	FL157	FL158	FL159	FL160	FL161	FL162	FL163	FL164	FL165	FL166	FL167	FL168	FL169	FL170	FL171	FL172	FL173	FL174	FL175	FL176	FL177	FL178	FL179	FL180	FL181	FL182	FL183	FL184	FL185	FL186	FL187	FL188	FL189	FL190	FL191	FL192	FL193	FL194	FL195	FL196	FL197	FL198	FL199	FL200	FL201	FL202	FL203	FL204	FL205	FL206	FL207	FL208	FL209	FL210	FL211	FL212	FL213	FL214	FL215	FL216	FL217	FL218	FL219	FL220	FL221	FL222	FL223	FL224	FL225	FL226	FL227	FL228	FL229	FL230	FL231	FL232	FL233	FL234	FL235	FL236	FL237	FL238	FL239	FL240	FL241	FL242	FL243	FL244	FL245	FL246	FL247	FL248	FL249	FL250	FL251	FL252	FL253	FL254	FL255	FL256	FL257	FL258	FL259	FL260	FL261	FL262	FL263	FL264	FL265	FL266	FL267	FL268	FL269	FL270	FL271	FL272	FL273	FL274	FL275	FL276	FL277	FL278	FL279	FL280	FL281	FL282	FL283	FL284	FL285	FL286	FL287	FL288	FL289	FL290	FL291	FL292	FL293	FL294	FL295	FL296	FL297	FL298	FL299	FL300	FL301	FL302	FL303	FL304	FL305	FL306	FL307	FL308	FL309	FL310	FL311	FL312	FL313	FL314	FL315	FL316	FL317	FL318	FL319	FL320	FL321	FL322	FL323	FL324	FL325	FL326	FL327	FL328	FL329	FL330	FL331	FL332	FL333	FL334	FL335	FL336	FL337	FL338	FL339	FL340	FL341	FL342	FL343	FL344	FL345	FL346	FL347	FL348	FL349	FL350	FL351	FL352	FL353	FL354	FL355	FL356	FL357	FL358	FL359	FL360	FL361	FL362	FL363	FL364	FL365	FL366	FL367	FL368	FL369	FL370	FL371	FL372	FL373	FL374	FL375	FL376	FL377	FL378	FL379	FL380	FL381	FL382	FL383	FL384	FL385	FL386	FL387	FL388	FL389	FL390	FL391	FL392	FL393	FL394	FL395	FL396	FL397	FL398	FL399	FL400	FL401	FL402	FL403	FL404	FL405	FL406	FL407	FL408	FL409	FL410	FL411	FL412	FL413	FL414	FL415	FL416	FL417	FL418	FL419	FL420	FL421	FL422	FL423	FL424	FL425	FL426	FL427	FL428	FL429	FL430	FL431	FL432	FL433	FL434	FL435	FL436	FL437	FL438	FL439	FL440	FL441	FL442	FL443	FL444	FL445	FL446	FL447	FL448	FL449	FL450	FL451	FL452	FL453	FL454	FL455	FL456	FL457	FL458	FL459	FL460	FL461	FL462	FL463	FL464	FL465	FL466	FL467	FL468	FL469	FL470	FL471	FL472	FL473	FL474	FL475	FL476	FL477	FL478	FL479	FL480	FL481	FL482	FL483	FL484	FL485	FL486	FL487	FL488	FL489	FL490	FL491	FL492	FL493	FL494	FL495	FL496	FL497	FL498	FL499	FL500	FL501	FL502	FL503	FL504	FL505	FL506	FL507	FL508	FL509	FL510	FL511	FL512	FL513	FL514	FL515	FL516	FL517	FL518	FL519	FL520	FL521	FL522	FL523	FL524	FL525	FL526	FL527	FL528	FL529	FL530	FL531	FL532	FL533	FL534	FL535	FL536	FL537	FL538	FL539	FL540	FL541	FL542	FL543	FL544	FL545	FL546	FL547	FL548	FL549	FL550	FL551	FL552	FL553	FL554	FL555	FL556	FL557	FL558	FL559	FL560	FL561	FL562	FL563	FL564	FL565	FL566	FL567	FL568	FL569	FL570	FL571	FL572	FL573	FL574	FL575	FL576	FL577	FL578	FL579	FL580	FL581	FL582	FL583	FL584	FL585	FL586	FL587	FL588	FL589	FL590	FL591	FL592	FL593	FL594	FL595	FL596	FL597	FL598	FL599	FL600	FL601	FL602	FL603	FL604	FL605	FL606	FL607	FL608	FL609	FL610	FL611	FL612	FL613	FL614	FL615	FL616	FL617	FL618	FL619	FL620	FL621	FL622	FL623	FL624	FL625	FL626	FL627	FL628	FL629	FL630	FL631	FL632	FL633	FL634	FL635	FL636	FL637	FL638	FL639	FL640	FL641	FL642	FL643	FL644	FL645	FL646	FL647	FL648	FL649	FL650	FL651	FL652	FL653	FL654	FL655	FL656	FL657	FL658	FL659	FL660	FL661	FL662	FL663	FL664	FL665	FL666	FL667	FL668	FL669	FL670	FL671	FL672	FL673	FL674	FL675	FL676	FL677	FL678	FL679	FL680	FL681	FL682	FL683	FL684	FL685	FL686	FL687	FL688	FL689	FL690	FL691	FL692	FL693	FL694	FL695	FL696	FL697	FL698	FL699	FL700	FL701	FL702	FL703	FL704	FL705	FL706	FL707	FL708	FL709	FL710	FL711	FL712	FL713	FL714	FL715	FL716	FL717	FL718	FL719	FL720	FL721	FL722	FL723	FL724	FL725	FL726	FL727	FL728	FL729	FL730	FL731	FL732	FL733	FL734	FL735	FL736	FL737	FL738	FL739	FL740	FL741	FL742	FL743	FL744	FL745	FL746	FL747	FL748	FL749	FL750	FL751	FL752	FL753	FL754	FL755	FL756	FL757	FL758	FL759	FL760	FL761	FL762	FL763	FL764	FL765	FL766	FL767	FL768	FL769	FL770	FL771	FL772	FL773	FL774	FL775	FL776	FL777	FL778	FL779	FL780	FL781	FL782	FL783	FL784	FL785	FL786	FL787	FL788	FL789	FL790	FL791	FL792	FL793	FL794	FL795	FL796	FL797	FL798	FL799	FL800	FL801	FL802	FL803	FL804	FL805	FL806	FL807	FL808	FL809	FL810	FL811	FL812	FL813	FL814	FL815	FL816	FL817	FL818	FL819	FL820	FL821	FL822	FL823	FL824	FL825	FL826	FL827	FL828	FL829	FL830	FL831	FL832	FL833	FL834	FL835	FL836	FL837	FL838	FL839	FL840	FL841	FL842	FL843	FL844	FL845	FL846	FL847	FL848	FL849	FL850	FL851	FL852	FL853	FL854	FL855	FL856	FL857	FL858	FL859	FL860	FL861	FL862	FL863	FL864	FL865	FL866	FL867	FL868	FL869	FL870	FL871	FL872	FL873	FL874	FL875	FL876	FL877	FL878	FL879	FL880	FL881	FL882	FL883	FL884	FL885	FL886	FL887	FL888	FL889	FL890	FL891	FL892	FL893	FL894	FL895	FL896	FL897	FL898	FL899	FL900	FL901	FL902	FL903	FL904	FL905	FL906	FL907	FL908	FL909	FL910	FL911	FL912	FL913	FL914	FL915	FL916	FL917	FL918	FL919	FL920	FL921	FL922	FL923	FL924	FL925	FL926	FL927	FL928	FL929	FL930	FL931	FL932	FL933	FL934	FL935	FL936	FL937	FL938	FL939	FL940	FL941	FL942	FL943	FL944	FL945	FL946	FL947	FL948	FL949	FL950	FL951	FL952	FL953	FL954	FL955	FL956	FL957	FL958	FL959	FL960	FL961	FL962	FL963	FL964	FL965	FL966	FL967	FL968	FL969	FL970	FL971	FL972	FL973	FL974	FL975	FL976	FL977	FL978	FL979	FL980	FL981	FL982	FL983	FL984	FL985	FL986	FL987	FL988	FL989	FL990	FL991	FL992	FL993	FL994	FL995	FL996	FL997	FL998	FL999	FL1000
1	60	FL	1	140	132	293																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				

WO 02/00860

PCT/US01/20171

Table 4 describes the results of Smith Waterman similarity searches (Matrix: Pam100; gap open/extension penalties 12/2) of the amino acid sequences against the NCBI database of non-redundant protein sequences

- 5 (<http://www.ncbi.nlm.nih.gov/Entrez/protein.html>). The column headings are:
 "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family",
 "Pscore", "aa_length", "aa_ID_match", "%Identity", "%Similar",
 "ACC#nraa_match", and "Description". The contents of the first 7 columns (i.e.,
 "Gene Name", "ID#na", "ID#aa", "FL/Cat", "Superfamily", "Group", "Family") are
 10 as described above for Table 1. "Pscore" refers to the Smith Waterman probability
 score. This number approximates the chance that the alignment occurred by chance.
 Thus, a very low number, such as 2.10E-64, indicates that there is a very significant
 match between the query and the database target. "aa_length" refers to the length of
 the protein in amino acids. "aa_ID_match" indicates the number of amino acids that
 15 were identical in the alignment. "% Identity" lists the percent of amino acids that
 were identical over the aligned region. "% Similarity" lists the percent of amino
 acids that were similar over the alignment. "ACC#nraa_match" lists the accession
 number of the most similar protein in the NCBI database of non-redundant proteins.
 "Description" contains the name of the most similar protein in the NCBI database of
 20 non-redundant proteins.

Table 4
Smith Waterman[illegible]

EXAMPLES

The examples below are not limiting and are merely representative of various aspects and features of the present invention. The examples below demonstrate the isolation and characterization of the proteases of the invention.

5

EXAMPLE 1: Identification of Genomic Fragments Encoding Proteases

Novel proteases were identified from the Celera human genomic sequence databases, and from the public Human Genome Sequencing project
10 (<http://www.ncbi.nlm.nih.gov/>) using hidden Markov models (HMMR). The genomic database entries were translated in six open reading frames and searched against the model using a Timelogic Decypher box with a Field programmable array (FPGA) accelerated version of HMMR2.1. The DNA sequences encoding the predicted protein sequences aligning to the HMMR profile were extracted from the
15 original genomic database. The nucleic acid sequences were then clustered using the Pangea Clustering tool to eliminate repetitive entries. The putative protease sequences were then sequentially run through a series of queries and filters to identify novel protease sequences. Specifically, the HMMR identified sequences were searched using BLASTN and BLASTX against a nucleotide and amino acid
20 repository containing known human proteases and all subsequent new protease sequences as they are identified. The output was parsed into a spreadsheet to facilitate elimination of known genes by manual inspection. Two models were used, a "complete" model and a "partial" or Smith Waterman model. The partial model was used to identify sub-catalytic domains, whereas the complete model was used to
25 identify complete catalytic domains. The selected hits were then queried using BLASTN against the public NRNA and EST databases to confirm they are indeed unique.

WO 02/00860

PCT/US01/20171

Extension of partial DNA sequences to encompass the longer sequences, including full-length open-reading frame, was carried out by several methods. Iterative blastn searching of the cDNA databases listed in Table 5 was used to find cDNAs that extended the genomic sequences. "LifeGold" databases are from Incyte Genomics, Inc (<http://www.incyte.com/>). NCBI databases are from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). All blastn searches were conducted using a penalty for a nucleotide mismatch of -3 and reward for a nucleotide match of 1. The gapped blast algorithm is described in: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402).

Extension of partial DNA sequences to encompass the full-length open-reading frame was also carried out by iterative searches of genomic databases. The first method made use of the Smith-Waterman algorithm to carry out protein-protein searches of the closest homologue or orthologue to the partial. The target databases consisted of Genscan [Chris Burge and Sam Karlin "Prediction of Complete Gene Structures in Human Genomic DNA", *JMB* (1997) 268(1):78-94]] and open-reading frame (ORF) predictions of all human genomic sequence derived from the human genome project (HGP) as well as from Celera. The complete set of genomic databases searched is shown in Table 6 below. Genomic sequences encoding potential extensions were further assessed by blastp analysis against the NCBI nonredundant to confirm the novelty of the hit. The extending genomic sequences were incorporated into the cDNA sequence after removal of potential introns using the Seqman program from DNASTar. The default parameters used for Smith-Waterman searches were Matrix: PAM100; gap-opening penalty: 12; gap extension penalty: 2. Genscan predictions were made using the Genscan program as detailed in Chris Burge and Sam Karlin "Prediction of Complete Gene Structures in Human

WO 02/00860

PCT/US01/20171

Genomic DNA", *JMB* (1997) 268(1):78-94). ORF predictions from genomic DNA were made using a standard 6-frame translation.

Another method for defining DNA extensions from genomic sequence used iterative searches of genomic databases through the Genscan program to predict exon splicing [Burge and Karlin, *JMB* (1997) 268(1):78-94)]. These predicted genes were then assessed to see if they represented "real" extensions of the partial genes based on homology to related proteases.

Another method involved using the Genewise program (<http://www.sanger.ac.uk/Software/Wise2/>) to predict potential ORFs based on homology to the closest orthologue/homologue. Genewise requires two inputs, the homologous protein, and genomic DNA containing the gene of interest. The genomic DNA was identified by blastn searches of Celera and Human Genome Project databases. The orthologs were identified by blastp searches of the NCBI non-redundant protein database (NR). Genewise compares the protein sequence to a genomic DNA sequence, allowing for introns and frameshifting errors.

Table 5. Databases used for cDNA-based sequence extensions

Database	Database Date
LifeGold templates	May 2001
LifeGold compseqs	May 2001
LifeGold compseqs	May 2001
LifeGold compseqs	May 2001
LifeGold fl	May 2001
LifeGold flt	May 2001
NCBI human Ests	May 2001
NCBI murine Ests	May 2001
NCBI nonredundant	May 2001

WO 02/00860

PCT/US01/20171

TABLE 6. DATABASES USED FOR GENOMIC-BASED SEQUENCE EXTENSIONS

Database	Number of entries	Database Date
Celera v. 1-5	5,306,158	Jan 2000
Celera v. 6-10	4,209,980	May 2000
Celera v. 11-14	7,222,425	April 2000
Celera v. 15	243,044	April 2000
Celera v. 16-17	25,885	April 2000
Celera Assembly 5 (release 25h)	479,986	May 2001
HGP Phase 0	3,189	Nov 1/00
HGP Phase 1	20,447	Jan 1/01
HGP Phase 2	1,619	Jan 1/01
HGP Phase 3	9,224	May 2001
HGP Chromosomal assemblies	2759	May 2001

5 **Results:**

The sources for the sequence information used to extend the genes in the provisional patents are listed below. For genes that were extended using Genewise, the accession numbers of the protein ortholog and the genomic DNA are given. (Genewise uses the ortholog to assemble the coding sequence of the target gene from the genomic sequence). The amino acid sequences for the orthologs were obtained from the NCBI non-redundant database of proteins (<http://www.ncbi.nlm.nih.gov/Entrez/protein.html>). The genomic DNA came from two sources: Celera and NCBI-NRNA, as indicated below. cDNA sources are also listed below. All of the genomic sequences were used as input for

10

15 Genscan predictions to predict splice sites [Burge and Karlin, JMB (1997)]

WO 02/00860

PCT/US01/20171

268(1):78-94]. Abbreviations: HGP: Human Genome Project; NCBI, National Center for Biotechnology Information.

SGPr397, SEQ ID NO:1, SEQ ID NO:60

- 5 Genewise orthologs: BAB25826.1, XP_005284.2, NP_065094.1.
Genomic DNA sources: Celera_asm5h 181000001172043
cDNA Sources: Public (gi|9966830|ref|NM_020361.1).

SGPr413, SEQ ID NO:2, SEQ ID NO:61

- 10 Genewise orthologs: gi|6013463, XP_003009.1, P15086.
Genomic DNA sources: Celera_asm5h 300475633

SGPr404, SEQ ID NO:3, SEQ ID NO:62

- Genewise orthologs: BAB31768.1, NP_061355.1, AAH03713.
15 Genomic DNA sources: Celera_asm5h 90000641768196

SGPr536_1, SEQ ID NO:4, SEQ ID NO:63

- Genewise orthologs: BAB18637.
Genomic DNA sources: 90000642234172
20

SGPr414, SEQ ID NO:5, SEQ ID NO:64

- Genewise orthologs: AAF50752.1.
Genomic DNA sources: 90000628114448
cDNA Sources: AK023845.1|AK023845 Homo sapiens cDNA FLJ13783 fis; Incyte
25 399773.5.

SGPr430, SEQ ID NO:6, SEQ ID NO:65

Genewise orthologs: NP_065954.
Genomic DNA sources: 301015601

WO 02/00860

PCT/US01/20171

cDNA Sources:AB046814 Homo sapiens mRNA for KIAA1594.

SGPr496_1, SEQ ID NO:7, SEQ ID NO:66

Genewise orthologs: AAH07196, AAF66953.

5 Genomic DNA sources:90000627702299

SGPr495, SEQ ID NO:8, SEQ ID NO:67

Genewise orthologs: NP_006438.

Genomic DNA sources:90000627041101

10

SGPr407, SEQ ID NO:9, SEQ ID NO:68

Genewise orthologs: BAB27431, AAH03130, NP_057656.

Genomic DNA sources:92000003986525

15 SGPr453, SEQ ID NO:10, SEQ ID NO:69

Genewise orthologs: NP_006528.

Genomic DNA sources:90000640175777

cDNA Sources:AL136825.1|HSM801793 Homo sapiens mRNA; Incyte 428428.1

20 SGPr445, SEQ ID NO:11, SEQ ID NO:70

Genewise orthologs: NP_006438.

Genomic DNA sources:90000627041101

cDNA Sources:9863487 328 bp ubiquitin carboxyl-terminal hydrolase; Incyte
4802789CA2

25

SGPr401_1, SEQ ID NO:12, SEQ ID NO:71

Genewise orthologs: BAB14881, NP_073743, BAB24720.

Genomic DNA sources:92000004473288

WO 02/00860

PCT/US01/20171

cDNA Sources:NM_022832.1| Homo sapiens hypothetical protein FLJ12552
(FLJ12552)

SGPr408, SEQ ID NO:13, SEQ ID NO:72

5 Genewise orthologs: Q24574, AAF50752.

Genomic DNA sources:90000628565543

cDNA Sources:AK027362.

SGPr480, SEQ ID NO:14, SEQ ID NO:73

10 Genewise orthologs: AAF49100, T29010.

Genomic DNA sources:90000640697688

cDNA Sources:EF_hand; CAAX: NP_115971.

SGPr431, SEQ ID NO:15, SEQ ID NO:74

15 Genewise orthologs: AAK26248, BAA92610, Q92353.

Genomic DNA sources:90000642340202

SGPr429, SEQ ID NO:16, SEQ ID NO:75

Genewise orthologs: BAB15591, AAG42764, gi_11358453.

20 Genomic DNA sources:90000642540891

cDNA Sources:AK026930.1|AK026930 Homo sapiens cDNA: FLJ23277.

SGPr503, SEQ ID NO:17, SEQ ID NO:76

Genewise orthologs: AAF40451, AAF46096, AAH04868.

25 Genomic DNA sources:90000642658172

cDNA Sources:BC004868.1|BC004868 Homo sapiens, clone MGC:10702; Incyte
5432879CB1.

SGPr427, SEQ ID NO:18, SEQ ID NO:77

WO 02/00860

PCT/US01/20171

Genewise orthologs: XP_003288, AAC27356, BAA86517.

Genomic DNA sources:181000001646773

cDNA Sources:Incyte 7485896CB1

5 SGPr092, SEQ ID NO:19, SEQ ID NO:78

Genewise orthologs: XP_011971.1, NP_068573.1, AAF80180.1.

Genomic DNA sources:Celera_asm5h 300261795

cDNA Sources:gi|12736016.

0 SGPr359, SEQ ID NO:20, SEQ ID NO:79

Genewise orthologs: 1.) gi|11545845|ref 2.) gi|12006364|gb 3.) gi|3511149|gb|A.

Genomic DNA sources:Celera_asm5h 90000642045264

cDNA Sources:gi|13639688.

5 SGPr104_1, SEQ ID NO:21, SEQ ID NO:80

Genewise orthologs: 1.) gi|7662200|ref.

Genomic DNA sources:HGP_s gi|12039078|4

cDNA Sources:NP_055508.1.

0 SGPr303, SEQ ID NO:22, SEQ ID NO:81

Genewise orthologs: CAA10709.1.

Genomic DNA sources:HGP_s gi|8082389_31

SGPr402_1, SEQ ID NO:23, SEQ ID NO:82

5 Genewise orthologs: A54306, I77530, A45357

Genomic DNA sources:Celera_asm5h 92000004018126

SGPr434, SEQ ID NO:24, SEQ ID NO:83

Genewise orthologs: gi|6755819, gi|6912728, gi|8570164.

WO 02/00860

PCT/US01/20171

Genomic DNA sources:90000628646128, 160000117588372, 165000100269164,
90000628646080

cDNA Sources:gi|6141221, gi|3754092, Incyte 1856589CB1.

- 5 SGPr446_1, SEQ ID NO:25, SEQ ID NO:84
Genewise orthologs: gi_11055972, gi_12839280, gi_13633971.
Genomic DNA sources:90000628646080

- SGPr447, SEQ ID NO:26, SEQ ID NO:85
10 Genewise orthologs: gi|12855280, gi|11055972, gi|8777606.
Genomic DNA sources:90000628729589

- SGPr432_1, SEQ ID NO:27, SEQ ID NO:86
Genewise orthologs: gi_11181573, gi_12832944, gi_13124769, gi_13277969,
15 gi_13632973.
Genomic DNA sources:90000631961624
cDNA Sources: Incyte EST 474674.1.

- SGPr529, SEQ ID NO:28, SEQ ID NO:87
20 Genewise orthologs: NP_002767, AAH02100
Genomic DNA sources:Celera_asm5h 92000003497776
cDNA Sources:gi|4506157.

- SGPr428_1, SEQ ID NO:29, SEQ ID NO:88
25 Genewise orthologs: gi|12838473, gi|12839985, gi|9651113, gi|4165315.
Genomic DNA sources:90000627342893

SGPr425, SEQ ID NO:30, SEQ ID NO:89
Genewise orthologs: gi_12844896, gi_6005882.

WO 02/00860

PCT/US01/20171

Genomic DNA sources:18100004221955

cDNA Sources: Incyte Sequence 400833.1.

SGPr548, SEQ ID NO:31, SEQ ID NO:90

- 5 Genewise orthologs: gi|9957760, gi|5803199, gi|6681654.

Genomic DNA sources:9200000349776, gi|11178143

cDNA Sources:gi|9957759.

SGPr396, SEQ ID NO:32, SEQ ID NO:91

- 10 Genewise orthologs: gi_11055972, gi_12839280, gi_6680267, gi_8393560,
gi_9757698.

Genomic DNA sources:90000632590917

cDNA Sources: Incyte Sequence 7480224CB1.

- 15 SGPr426, SEQ ID NO:33, SEQ ID NO:92

Genewise orthologs: gi_13640890, gi_13646365, gi_7661558.

Genomic DNA sources:90000641479138

cDNA Sources:Incyte Sequence 7481056CB1.

- 20 SGPr552, SEQ ID NO:34, SEQ ID NO:93

Genewise orthologs: gi|7661558, gi|4758508.

Genomic DNA sources:90000641479138

SGPr405, SEQ ID NO:35, SEQ ID NO:94

- 25 Genewise orthologs: gi_7415931, gi_126839, gi_136423, gi_13183572.

Genomic DNA sources: gi|13509126

cDNA Sources:Incyte seqs 7474351CB1 and 134360.1.

SGPr485_1, SEQ ID NO:36, SEQ ID NO:95

WO 02/00860

PCT/US01/20171

Genewise orthologs: gi|9651113.

Genomic DNA sources:90000627342893

cDNA Sources:Incyte Sequence 6026494CA2.

5 SGPr534, SEQ ID NO:37, SEQ ID NO:96

Genewise orthologs: gi|4503135.

Genomic DNA sources:9200004436076, 165000101932709, 92000004433469

cDNA Sources:Incyte ESTs: 1383391.20, 1383391.10, 1383391.13, 7691434H1,
2070278CB1, 741522CA2; NCB1 ESTs: gi|7260671, gi|7260006, gi|7260642,

0 gi|7259962, gi|2018619, gi|7260655, gi|2019751.

SGPr390, SEQ ID NO:38, SEQ ID NO:97

Genewise orthologs: BAB23684

Genomic DNA sources:hCG22693

5

SGPr521, SEQ ID NO:39, SEQ ID NO:98

Genewise orthologs: BAB55604, AAF01139, AAF01139

Genomic DNA sources:HGP_s gi|11178143_10

cDNA Sources:gi|4826949.

10

SGPr530_1, SEQ ID NO:40, SEQ ID NO:99

Genewise orthologs: gi_12314133, NP_033381.1 3, NP_033382.1

Genomic DNA sources:Celera_asm5h 181000001848433

15 SGPr520, SEQ ID NO:41, SEQ ID NO:100

Genewise orthologs: gi|12839535 gi|1352368, gi|4506151.

Genomic DNA sources:90000640807190

WO 02/00860

PCT/US01/20171

cDNA Sources: ESTs gi|13745759, 7472044CB1, 7474338CB1, gi|13703426,
gi|5392427, gi|2142177, gi|2103202, LIB4218-103-R1-K1-H5, LIB4218-085-Q1-
K1-C6, LIB4752-019-R1-K1-H4

- 5 SGPr455, SEQ ID NO:42, SEQ ID NO:101
Genewise orthologs: gi|7512178, gi|7512176.
Genomic DNA sources: 90000641321557
cDNA Sources: Incyte template 987279.1.
- 10 SGPr507_2, SEQ ID NO:43, SEQ ID NO:102
Genewise orthologs: gi|13385812, gi|12854692, gi|2499862.
Genomic DNA sources: 90000642611957
- SGPr559, SEQ ID NO:44, SEQ ID NO:103
- 15 Genewise orthologs: XP_016993, BAB20079
Genomic DNA sources: Celera_asm5h 335001064013332
cDNA Sources: gi|13173471.
- SGPr567_1, SEQ ID NO:45, SEQ ID NO:104
- 20 Genewise orthologs: NP_114435, Q9JIQ8
Genomic DNA sources: Celera_asm5h 90000642045213
cDNA Sources: gi|14042983|ref|NM_032046.1.
- SGPr479_1, SEQ ID NO:46, SEQ ID NO:105
- 25 Genewise orthologs: NP_114154, NP_038949, NP_033382
Genomic DNA sources: 90000624931837
cDNA Sources: EST gi|13997890 and Incyte EST 7480124CB1,.
- SGPr489_1, SEQ ID NO:47, SEQ ID NO:106

WO 02/00860

PCT/US01/20171

Genewise orthologs: gi|7512176, gi|7512178, gi|9757698.

Genomic DNA sources:90000628565500

SGPr465_1, SEQ ID NO:48, SEQ ID NO:107

5 Genewise orthologs: gi|6678293, gi|6678295.

Genomic DNA sources:gi|13431162

SGPr524_1, SEQ ID NO:49, SEQ ID NO:108

Genewise orthologs: gi|12836503, gi|10257390, gi|11415040.

10 Genomic DNA sources:90000626428259

SGPr422, SEQ ID NO:50, SEQ ID NO:109

Genewise orthologs: gi|7661558, gi|4758508.

Genomic DNA sources:90000641479138

15

SGPr538, SEQ ID NO:51, SEQ ID NO:110

Genewise orthologs: NP_110397, Q9ER04, NP_109634

Genomic DNA sources:Celera_asm5h.90000642044035 and 90000642045412

cDNA Sources:gi|13540535.

20

SGPr527_1, SEQ ID NO:52, SEQ ID NO:111

Genewise orthologs: gi|11181573, gi|13277969, gi|10441463.

Genomic DNA sources:90000631961624

cDNA Sources:Incyte 2751509CB1.

25

SGPr542, SEQ ID NO:53, SEQ ID NO:112

Genewise orthologs: gi|1705760, gi|4885369.

Genomic DNA sources:92000004018116, gi|2896799, 92000004013323,

92000004013330, 165000100427031

WO 02/00860

PCT/US01/20171

SGPr551, SEQ ID NO:54, SEQ ID NO:113
Genewise orthologs: BAB23684.1, NP_035306.2, BAB03502.1
Genomic DNA sources: Celera_asm5h 90000643090998

5

SGPr451, SEQ ID NO:55, SEQ ID NO:114
Genewise orthologs: gi_5002340, gi|12018322, gi|1480413.
Genomic DNA sources: 181000000828193

0

SGPr452_1, SEQ ID NO:56, SEQ ID NO:115
Genewise orthologs: gi|13183572, gi|339983, gi|7415931.
Genomic DNA sources: 92000004034678

5

SGPr504, SEQ ID NO:57, SEQ ID NO:116
Genewise orthologs: gi|1633237
Genomic DNA sources: celera_asm5h 92000004018137

0

SGPr469, SEQ ID NO:58, SEQ ID NO:117
Genewise orthologs: BAB30277, CAB41988, XP_016204
Genomic DNA sources: GA_x2HTBKPYW7D

5

SGPr400, SEQ ID NO:59, SEQ ID NO:118
Genewise orthologs: gi|6755819, gi|6912728.
Genomic DNA sources: 90000632590917

5

DESCRIPTION OF NOVEL PROTEASE POLYNUCLEOTIDES

SGPr397, SEQ ID NO:1, SEQ ID NO:60 is 948 nucleotides long. The open reading frame starts at position 1 and ends at position 948, giving an ORF length of 948

WO 02/00860

PCT/US01/20171

- nucleotides. The predicted protein is 315 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Carboxypeptidase, Zn carboxypeptidase. The cytogenetic position of this gene is 8q12. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AV763490.
- SGPr413, SEQ ID NO:2, SEQ ID NO:61 is 1125 nucleotides long. The open reading frame starts at position 1 and ends at position 1125, giving an ORF length of 1125 nucleotides. The predicted protein is 374 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Carboxypeptidase, Zn carboxypeptidase. The cytogenetic position of this gene is 2q35. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.
- SGPr404, SEQ ID NO:3, SEQ ID NO:62 is 1590 nucleotides long. The open reading frame starts at position 1 and ends at position 1590, giving an ORF length of 1590 nucleotides. The predicted protein is 529 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Carboxypeptidase, Zn carboxypeptidase. The cytogenetic position of this gene is 10q26. This nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss1782198_allelePos=201, agaagcctaygaagggg. SNP ss1782198 occurs at nucleotide 612 (aa 58) of the ORF (C or T = silent; AA 204 = tyrosine with either nucleotide). This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AA045748, AA148684, AA047483. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 477 ggagctgctgctgctgctgctg 498.

WO 02/00860

PCT/US01/20171

SGPr536_1, SEQ ID NO:4, SEQ ID NO:63 is 1404 nucleotides long. The open reading frame starts at position 1 and ends at position 1404, giving an ORF length of 1404 nucleotides. The predicted protein is 467 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):

- 5 Protease, Cysteine, papain. The cytogenetic position of this gene is 1p35. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AL542213, AL547246, AL552037. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 480 gctgctgctgctgctgctgcag 501.

10

SGPr414, SEQ ID NO:5, SEQ ID NO:64 is 10062 nucleotides long. The open reading frame starts at position 1 and ends at position 10062, giving an ORF length of 10062 nucleotides. The predicted protein is 3353 amino acids long. This sequence codes for a full length protein. It is classified as

- 15 (superfamily/group/family): Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 2p14. This nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, followed by the sequence surrounding the SNP within the gene): ss16542_allelePos=101, ctaccctagcygaggaaga. SNP ss16542 occurs at nucleotide 9807 (aa 3269, alanine) of
20 the ORF. The SNP is silent. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AU118237, AU131420, AU125083. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 2249 accaccaccaccaccaccaccaccaccacc 2280.

- 25 SGPr430, SEQ ID NO:6, SEQ ID NO:65 is 2943 nucleotides long. The open reading frame starts at position 1 and ends at position 2943, giving an ORF length of 2943 nucleotides. The predicted protein is 980 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 2q37. This

WO 02/00860

PCT/US01/20171

nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss1534585_allelePos=51, tggaaatcctcggac; rs1055687_allelePos=51, tggtaatccgkgtagagg. SNP ss1534585 occurs at nucleotide 538 (aa 180) of the ORF (A or G). The SNP ss1534585 changes amino acid 180. If nucleotide 538 is an adenine, amino acid 180 is a threonine; if it is a guanine, amino acid 180 is an alanine. A second SNP, rs1055687, codes for a G or T at nucleotide 499. rs1055687 changes the amino acid sequence of the gene. Amino acid 167 is a cysteine if nucleotide 499 is a thymidine; amino acid 167 is a glycine if nucleotide 499 is a guanine. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: W87666, AI076108, BG612864.

SGPr496_1, SEQ ID NO:7, SEQ ID NO:66 is 2862 nucleotides long. The open reading frame starts at position 1 and ends at position 2862, giving an ORF length of 2862 nucleotides. The predicted protein is 953 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Cysteine, UCH2b. The cytogenetic position of this gene is Xp11.4. This nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss1029756_allelePos=101, agagaaataygagggtatt. SNP ss1029756 codes for a C or T at nucleotide 351. Amino acid 117 is a tyrosine with either nucleotide, so the SNP is silent. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AW851066, AW851065, AW851076.

SGPr495, SEQ ID NO:8, SEQ ID NO:67 is 2352 nucleotides long. The open reading frame starts at position 1 and ends at position 2352, giving an ORF length of 2352 nucleotides. The predicted protein is 783 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):

WO 02/00860

PCT/US01/20171

Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 6q16. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AL559960, AL530470, AL516184.

- 5 SGPr407, SEQ ID NO:9, SEQ ID NO:68 is 2259 nucleotides long. The open reading frame starts at position 1 and ends at position 2259, giving an ORF length of 2259 nucleotides. The predicted protein is 752 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
 0 Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 2q37. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.

- SGPr453, SEQ ID NO:10, SEQ ID NO:69 is 2139 nucleotides long. The open reading frame starts at position 1 and ends at position 2139, giving an ORF length of
 5 2139 nucleotides. The predicted protein is 712 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
 Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 12q23. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: BG722436, AI927881, BG771888. The nucleic acid contains short repetitive
 0 sequence (the position and sequence of the repeat): 553 gtagtaaaagagaagtaaa 572.

- SGPr445, SEQ ID NO:11, SEQ ID NO:70 is 870 nucleotides long. The open reading frame starts at position 1 and ends at position 870, giving an ORF length of
 5 870 nucleotides. The predicted protein is 289 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
 Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 6q16. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AL559960, AL530470, AL516184.

WO 02/00860

PCT/US01/20171

SGPr401_1, SEQ ID NO:12, SEQ ID NO:71 is 1101 nucleotides long. The open reading frame starts at position 1 and ends at position 1101, giving an ORF length of 1101 nucleotides. The predicted protein is 366 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):

- 5 Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 4q11. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AU124898, AU134553, AJ269069.

SGPr408, SEQ ID NO:13, SEQ ID NO:72 is 3864 nucleotides long. The open reading frame starts at position 1 and ends at position 3864, giving an ORF length of 3864 nucleotides. The predicted protein is 1287 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):

- 10 Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 11p15. This sequence is represented in the database of public ESTs (dbEST) by the following
15 ESTs: BG741190, BF575498, BG170829.

SGPr480, SEQ ID NO:14, SEQ ID NO:73 is 4815 nucleotides long. The open reading frame starts at position 1 and ends at position 4815, giving an ORF length of 4815 nucleotides. The predicted protein is 1604 amino acids long. This sequence codes for a partial protein. It is classified as (superfamily/group/family):
20 Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 17q24. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs:
AU131748, AU120381, BG420766.

- 25 SGPr431, SEQ ID NO:15, SEQ ID NO:74 is 3129 nucleotides long. The open reading frame starts at position 1 and ends at position 3129, giving an ORF length of 3129 nucleotides. The predicted protein is 1042 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 4q31.3. This

WO 02/00860

PCT/US01/20171

sequence is represented in the database of public ESTs (dbEST) by the following
ESTs: BG575871, BG113469, BG112979.

SGPr429, SEQ ID NO:16, SEQ ID NO:75 is 3102 nucleotides long. The open
5 reading frame starts at position 1 and ends at position 3102, giving an ORF length of
3102 nucleotides. The predicted protein is 1033 amino acids long. This sequence
codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 1p36.2. This
sequence is represented in the database of public ESTs (dbEST) by the following
.0 ESTs: AL518266, BG681225, BG217186.

SGPr503, SEQ ID NO:17, SEQ ID NO:76 is 1554 nucleotides long. The open
reading frame starts at position 1 and ends at position 1554, giving an ORF length of
1554 nucleotides. The predicted protein is 517 amino acids long. This sequence
.5 codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 12q24.3. This
sequence is represented in the database of public ESTs (dbEST) by the following
ESTs: BG678894, BG476418, BE264732. The nucleic acid contains short repetitive
sequence (the position and sequence of the repeat): 1534 gagtgcaagtctgaagaatg
10 1553.

SGPr427, SEQ ID NO:18, SEQ ID NO:77 is 3372 nucleotides long. The open
reading frame starts at position 1 and ends at position 3372, giving an ORF length of
3372 nucleotides. The predicted protein is 1123 amino acids long. This sequence
25 codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Cysteine, UCH2b. The cytogenetic position of this gene is 17p13. This
sequence is represented in the database of public ESTs (dbEST) by the following
ESTs: BG831111, AW996553, BE614914.

WO 02/00860

PCT/US01/20171

- SGPr092, SEQ ID NO:19, SEQ ID NO:78 is 786 nucleotides long. The open reading frame starts at position 1 and ends at position 786, giving an ORF length of 786 nucleotides. The predicted protein is 261 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- 5 Protease, Metalloprotease, PepM10. The cytogenetic position of this gene is 11p15. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: BG189720, AW966183, BG198356.
- SGPr359, SEQ ID NO:20, SEQ ID NO:79 is 1452 nucleotides long. The open
- 10 reading frame starts at position 1 and ends at position 1452, giving an ORF length of 1452 nucleotides. The predicted protein is 483 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Metalloprotease, PepM10. The cytogenetic position of this gene is 11q22. This sequence is represented in the database of public ESTs (dbEST) by the
- 15 following ESTs: BG187290.
- SGPr104_1, SEQ ID NO:21, SEQ ID NO:80 is 2298 nucleotides long. The open reading frame starts at position 1 and ends at position 2298, giving an ORF length of 2298 nucleotides. The predicted protein is 765 amino acids long. This sequence
- 20 codes for a full length protein. It is classified as (superfamily/group/family): Protease, Metalloprotease, PepM13. The cytogenetic position of this gene is 3q27. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: BF511209, AW341249, AL119270.
- 25 SGPr303, SEQ ID NO:22, SEQ ID NO:81 is 1257 nucleotides long. The open reading frame starts at position 1 and ends at position 1257, giving an ORF length of 1257 nucleotides. The predicted protein is 418 amino acids long. This sequence codes for a full length catalytic domain. It is classified as (superfamily/group/family): Protease, Metalloprotease, PepM2. The cytogenetic

WO 02/00860

PCT/US01/20171

position of this gene is 17q11.1. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AU138954, BG251083, AW161660.

5 SGPr402_1, SEQ ID NO:23, SEQ ID NO:82 is 2268 nucleotides long. The open reading frame starts at position 1 and ends at position 2268, giving an ORF length of 2268 nucleotides. The predicted protein is 755 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, subtilase. The cytogenetic position of this gene is 19q11. This sequence is represented in the database of public ESTs (dbEST) by the following
10 ESTs: AL041695, AA454137, BG719638.

SGPr434, SEQ ID NO:24, SEQ ID NO:83 is 1176 nucleotides long. The open reading frame starts at position 1 and ends at position 1176, giving an ORF length of 1176 nucleotides. The predicted protein is 391 amino acids long. This sequence
15 codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 3p21. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AW137088, BF593342.

20 SGPr446_1, SEQ ID NO:25, SEQ ID NO:84 is 681 nucleotides long. The open reading frame starts at position 1 and ends at position 681, giving an ORF length of 681 nucleotides. The predicted protein is 226 amino acids long. This sequence codes for a full length catalytic domain. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 3p21. This
25 sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AW243584. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 798 ggtgggcatcatcagctgggg 818.

WO 02/00860

PCT/US01/20171

SGPr447, SEQ ID NO:26, SEQ ID NO:85 is 888 nucleotides long. The open reading frame starts at position 1 and ends at position 888, giving an ORF length of 888 nucleotides. The predicted protein is 295 amino acids long. This sequence codes for a partial protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 16p13.3. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.

SGPr432_1, SEQ ID NO:27, SEQ ID NO:86 is 1887 nucleotides long. The open reading frame starts at position 1 and ends at position 1887, giving an ORF length of 1887 nucleotides. The predicted protein is 628 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is unknown. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: BE264142, BG474605, BF304202.

SGPr529, SEQ ID NO:28, SEQ ID NO:87 is 831 nucleotides long. The open reading frame starts at position 1 and ends at position 831, giving an ORF length of 831 nucleotides. The predicted protein is 276 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 19q13.4. This nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss1550333_allelePos=51, taggggatgaycacctgct; ss1546197_allelePos=51, gccggacsactcgc. SNP ss1550333 codes for a C or T at nucleotide 297. Amino acid 99 is an aspartic acid with either nucleotide, so the SNP is silent. There is another SNP, ss1546197, that codes for a C or G at position 336; amino acid 112 is a threonine when either nucleotide is present, and so this SNP is silent. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: BE898352, BG469321.

WO 02/00860

PCT/US01/20171

- SGPr428_1, SEQ ID NO:29, SEQ ID NO:88 is 858 nucleotides long. The open reading frame starts at position 1 and ends at position 858, giving an ORF length of 858 nucleotides. The predicted protein is 285 amino acids long. This sequence
- 5 codes for a full length catalytic domain. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 8p23. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 473 catgcacctggaaaagctg 491.
- 10 SGPr425, SEQ ID NO:30, SEQ ID NO:89 is 1242 nucleotides long. The open reading frame starts at position 1 and ends at position 1242, giving an ORF length of 1242 nucleotides. The predicted protein is 413 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- 15 Protease, Serine, Trypsin. The cytogenetic position of this gene is 6q14. This nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss674620 _allelePos=201, gagcatctgcVggagagag,, SNP ss674620 codes for a G or A or C at nucleotide 671. If
- 20 the nucleotide is a guanine, amino acid 224 is an arginine; if it is an adenine, amino acid 224 is a glutamine; if the nucleotide is a cytosine, the amino acid at 224 is a proline. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AL551286, AA445948, AA424073. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 1111
- 25 tcagggcaccagtgggtgga 1130.
- SGPr548, SEQ ID NO:31, SEQ ID NO:90 is 963 nucleotides long. The open reading frame starts at position 1 and ends at position 963, giving an ORF length of 963 nucleotides. The predicted protein is 320 amino acids long. This sequence

WO 02/00860

PCT/US01/20171

codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Serine, Trypsin. The cytogenetic position of this gene is 19q13.4. This
sequence is represented in the database of public ESTs (dbEST) by the following
ESTs: none.

5

SGPr396, SEQ ID NO:32, SEQ ID NO:91 is 987 nucleotides long. The open
reading frame starts at position 1 and ends at position 987, giving an ORF length of
987 nucleotides. The predicted protein is 328 amino acids long. This sequence
codes for a full length protein. It is classified as (superfamily/group/family):

10 Protease, Serine, Trypsin. The cytogenetic position of this gene is 4q32. This
sequence is represented in the database of public ESTs (dbEST) by the following
ESTs: none.

SGPr426, SEQ ID NO:33, SEQ ID NO:92 is 1278 nucleotides long. The open
15 reading frame starts at position 1 and ends at position 1278, giving an ORF length of
1278 nucleotides. The predicted protein is 425 amino acids long. This sequence
codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Serine, Trypsin. The cytogenetic position of this gene is 4q13. This
sequence is represented in the database of public ESTs (dbEST) by the following
20 ESTs: none.

SGPr552, SEQ ID NO:34, SEQ ID NO:93 is 666 nucleotides long. The open
reading frame starts at position 1 and ends at position 666, giving an ORF length of
666 nucleotides. The predicted protein is 221 amino acids long. This sequence
25 codes for a full length catalytic domain. It is classified as
(superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of
this gene is 4q13. This sequence is represented in the database of public ESTs
(dbEST) by the following ESTs: none.

WO 02/00860

PCT/US01/20171

SGPr405, SEQ ID NO:35, SEQ ID NO:94 is 2847 nucleotides long. The open reading frame starts at position 1 and ends at position 2847, giving an ORF length of 2847 nucleotides. The predicted protein is 948 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):

- 5 Protease, Serine, Trypsin. The cytogenetic position of this gene is 16p13.3. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.

SGPr485_1, SEQ ID NO:36, SEQ ID NO:95 is 1059 nucleotides long. The open reading frame starts at position 1 and ends at position 1059, giving an ORF length of 1059 nucleotides. The predicted protein is 352 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):

- 10 Protease, Serine, Trypsin. The cytogenetic position of this gene is 8p23. This nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene):ss1532791_allelePos=51 , tggagakaagaacac.
15 ss1532791 codes for a G or a T at position 834. This polymorphism changes amino acid 278. If the nucleotide at 834 is a guanine, amino acid 278 is a glutamic acid (E); if the nucleotide is a thymine, amino acid 278 is an aspartic acid (D). This
20 sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AA781356.

SGPr534, SEQ ID NO:37, SEQ ID NO:96 is 792 nucleotides long. The open reading frame starts at position 1 and ends at position 792, giving an ORF length of 792 nucleotides. The predicted protein is 263 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):

- 25 Protease, Serine, Trypsin. The cytogenetic position of this gene is 16q23. This nucleotide sequence contains the following six single nucleotide polymorphisms (the

WO 02/00860

PCT/US01/20171

- accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene):
- ss1522946_allelePos=51, gctctaccwccacgccc;
- ss1522943_allelePos=51, cgcacctgctcyaccaccac;
- 5 ss1522933_allelePos=51, ctgccagaaggayggagcctgg;
- ss1522931_allelePos=51 total len = 101, gtctgccaraaggacg;
- ss1522930_allelePos=51, gggtgactctgmggccccct;
- ss1522928_allelePos=51, tgcattgggygactctgg;
- SNP ss1522946 codes for A or T at position 721. If 721 is adenine, amino acid
- 10 241 is threonine (T); if 721 is Thymine, amino acid 241 is serine (S).
- SNP ss1522943 codes for C or T at position 717; this SNP is silent (239 = serine).
- SNP ss1522933 codes for C or T at 666; this SNP is silent (222 = aspartic acid).
- SNP ss1522931 codes for A or G at position 660; this SNP is silent (220 =
- 15 glutamine).
- SNP ss1522930 codes for A or C at position 642; this SNP is silent (214 = glycine).
- SNP ss1522928 codes for a C or T at position 633; this SNP is silent (211 = glycine).
- 20 This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AW583018, AW582942, AW960025. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 172 cactctgcggggctccctc 195.
- 25 SGPr390, SEQ ID NO:38, SEQ ID NO:97 is 3387 nucleotides long. The open reading frame starts at position 1 and ends at position 3387, giving an ORF length of 3387 nucleotides. The predicted protein is 1128 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 19q11. This

WO 02/00860

PCT/US01/20171

nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss82431_allelePos=99 , gccgtgaccaccactg; ss1320361_allelePos=225, agcggccascattggcgt. ss82431 codes for an A or G at
5 position 2585. If this nucleotide is an adenine, amino acid 862 is an asparagine (N); if this nucleotide is a guanine, amino acid 862 is a serine. The SNP ss1320361 codes for C or G at position 89. If position 89 is a cytosine, amino acid 30 is a threonine (T). If position 89 is a guanine, amino acid 30 is a serine. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs:
10 C16607.

SGPr521, SEQ ID NO:39, SEQ ID NO:98 is 762 nucleotides long. The open reading frame starts at position 1 and ends at position 762, giving an ORF length of 762 nucleotides. The predicted protein is 253 amino acids long. This sequence
15 codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 19q13.4. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AA542994, BE713379, W58737. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 646 caaggtctgtgtcctctggg 665.

20 SGPr530_1, SEQ ID NO:40, SEQ ID NO:99 is 816 nucleotides long. The open reading frame starts at position 1 and ends at position 816, giving an ORF length of 816 nucleotides. The predicted protein is 271 amino acids long. This sequence codes for a full length catalytic domain. It is classified as
25 (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 9q22. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.

WO 02/00860

PCT/US01/20171

- SGPr520, SEQ ID NO:41, SEQ ID NO:100 is 1737 nucleotides long. The open reading frame starts at position 1 and ends at position 1737, giving an ORF length of 1737 nucleotides. The predicted protein is 578 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- 5 Protease, Serine, Trypsin. The cytogenetic position of this gene is 2q37. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.
- SGPr455, SEQ ID NO:42, SEQ ID NO:101 is 2913 nucleotides long. The open reading frame starts at position 1 and ends at position 2913, giving an ORF length of 2913 nucleotides. The predicted protein is 970 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- 10 Protease, Serine, Trypsin. The cytogenetic position of this gene is 12p11.2. This sequence is represented in the database of public ESTs (dbEST) by the following
- 15 ESTs: AW450155, AW995496.
- SGPr507_2, SEQ ID NO:43, SEQ ID NO:102 is 798 nucleotides long. The open reading frame starts at position 1 and ends at position 798, giving an ORF length of 798 nucleotides. The predicted protein is 265 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- 20 Protease, Serine, Trypsin. The cytogenetic position of this gene is 7q36. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: BG217724, BG219738, BG192709.
- 25 SGPr559, SEQ ID NO:44, SEQ ID NO:103 is 1365 nucleotides long. The open reading frame starts at position 1 and ends at position 1365, giving an ORF length of 1365 nucleotides. The predicted protein is 454 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- Protease, Serine, Trypsin. The cytogenetic position of this gene is 21q22. This

WO 02/00860

PCT/US01/20171

sequence is represented in the database of public ESTs (dbEST) by the following
ESTs: AI978874, AI469095, BF435670

5 SGPr567_1, SEQ ID NO:45, SEQ ID NO:104 is 1614 nucleotides long. The open
reading frame starts at position 1 and ends at position 1614, giving an ORF length of
1614 nucleotides. The predicted protein is 537 amino acids long. This sequence
codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Serine, Trypsin. The cytogenetic position of this gene is 11q23. This
sequence is represented in the database of public ESTs (dbEST) by the following
10 ESTs: BE732381, R78581, AW845106.

SGPr479_1, SEQ ID NO:46, SEQ ID NO:105 is 981 nucleotides long. The open
reading frame starts at position 1 and ends at position 981, giving an ORF length of
981 nucleotides. The predicted protein is 326 amino acids long. This sequence
15 codes for a full length protein. It is classified as (superfamily/group/family):
Protease, Serine, Trypsin. The cytogenetic position of this gene is 1q42. This
sequence is represented in the database of public ESTs (dbEST) by the following
ESTs: BG718703, AA401705, AA398170. The nucleic acid contains short
repetitive sequence (the position and sequence of the repeat): 780
20 tgggaattgtgagctggggccg 800.

SGPr489_1, SEQ ID NO:47, SEQ ID NO:106 is 1671 nucleotides long. The open
reading frame starts at position 1 and ends at position 1671, giving an ORF length of
1671 nucleotides. The predicted protein is 556 amino acids long. This sequence
25 codes for a full length catalytic domain. It is classified as
(superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of
this gene is 11p15. This sequence is represented in the database of public ESTs
(dbEST) by the following ESTs: AW271430, AW237893.

WO 02/00860

PCT/US01/20171

SGPr465_1, SEQ ID NO:48, SEQ ID NO:107 is 894 nucleotides long. The open reading frame starts at position 1 and ends at position 894, giving an ORF length of 894 nucleotides. The predicted protein is 297 amino acids long. This sequence codes for a full length catalytic domain. It is classified as

- 5 (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is unknown. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.

- 10 SGPr524_1, SEQ ID NO:49, SEQ ID NO:108 is 2553 nucleotides long. The open reading frame starts at position 1 and ends at position 2553, giving an ORF length of 2553 nucleotides. The predicted protein is 850 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is unknown. This nucleotide sequence contains the following single nucleotide polymorphisms (the
- 15 accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss2013558_allelePos=201, gacatggawgtggacgac; ss2014128_allelePos=358, acaatttttygagtgccca.ss2013558 codes for a T of C at position 675; this is a silent SNP. Ss2014128 codes for a C or T at 1369; if the nucleotide is a cytosine, amino acid 457 is an arginine; if the
- 20 nucleotide 1369 is a thymine, a stop codon is introduced, truncating the protein to 456 amino acids. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 711
- aaaaaaaagaaaaagaaaaa 734.

- 25 SGPr422, SEQ ID NO:50, SEQ ID NO:109 is 1344 nucleotides long. The open reading frame starts at position 1 and ends at position 1344, giving an ORF length of 1344 nucleotides. The predicted protein is 447 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):

WO 02/00860

PCT/US01/20171

Protease, Serine, Trypsin. The cytogenetic position of this gene is 4q13. This nucleotide sequence contains the following single nucleotide polymorphisms (the accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss1091793_allelePos=101, acatacgcrcgattgtttg; ss448607_allelePos=101, tgggagcrgctcctgcct. SNP ss1091793 codes for an adenine or guanine at position 956. If 956 is guanine, amino acid 319 is arginine (R); if nucleotide 956 is adenine, amino acid 319 is glutamine (Q). The SNP ss448607 codes for an A or G at position 552. This is silent (amino acid 184 = alanine). This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.

SGPr538, SEQ ID NO:51, SEQ ID NO:110 is 1374 nucleotides long. The open reading frame starts at position 1 and ends at position 1374, giving an ORF length of 1374 nucleotides. The predicted protein is 457 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is 11q23. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AL538140, BF934870. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 545 tgggagcctcctggaggag 564.

SGPr527_1, SEQ ID NO:52, SEQ ID NO:111 is 2457 nucleotides long. The open reading frame starts at position 1 and ends at position 2457, giving an ORF length of 2457 nucleotides. The predicted protein is 818 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is unknown. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AW450407, AI190509, AI864473.

WO 02/00860

PCT/US01/20171

- SGPr542, SEQ ID NO:53, SEQ ID NO:112 is 855 nucleotides long. The open reading frame starts at position 1 and ends at position 855, giving an ORF length of 855 nucleotides. The predicted protein is 284 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- 5 Protease, Serine, Trypsin. The cytogenetic position of this gene is 19q13.1. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none.
- SGPr551, SEQ ID NO:54, SEQ ID NO:113 is 2409 nucleotides long. The open
- 10 reading frame starts at position 1 and ends at position 2409, giving an ORF length of 2409 nucleotides. The predicted protein is 802 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- Protease, Serine, Trypsin. The cytogenetic position of this gene is 22q13. This nucleotide sequence contains the following single nucleotide polymorphisms (the
- 15 accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): rs881144 _allelePos=200, ctgcagccctaygccgagagg; rs855791 _allelePos=101, agcgaggcyctatcgcta. SNP rs881144 codes for a C or T at position 1227; this a a silent SNP (409 = tyrosine). SNP rs855791 codes for C or T at position 2180. If the nucleotide at 2180 is
- 20 cytosine, the amino acid at 727 is alanine; if the nucleotide is thymine, amino acid 727 is valine. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: AV693114, N70418, AA609066
- SGPr451, SEQ ID NO:55, SEQ ID NO:114 is 1080 nucleotides long. The open
- 25 reading frame starts at position 1 and ends at position 1080, giving an ORF length of 1080 nucleotides. The predicted protein is 359 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- Protease, Serine, Trypsin. The cytogenetic position of this gene is 12q23. This nucleotide sequence contains the following single nucleotide polymorphisms (the

WO 02/00860

PCT/US01/20171

- accession number of SNP is given, with the allele position, followed by the sequence surrounding the SNP within the gene): ss1881349_allelePos=201, ggggcgcagcaragg; ss1266911_allelePos=101, ccactgcactaaagacrcctag. SNP ss1881349 codes for an A or G at position 217. If the nucleotide at 217 is adenine, amino acid 73 is lysine (K);
- 5 if the nucleotide is guanine, amino acid 73 is glutamic acid (E). The SNP ss1266911 codes for an A or G at position 412. If 412 is guanine, amino acid 138 is alanine (A); if 412 is adenine, amino acid 138 is threonine (T). This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: BG722131, BG722203,
- 10 SGPr452_1, SEQ ID NO:56, SEQ ID NO:115 is 867 nucleotides long. The open reading frame starts at position 1 and ends at position 867, giving an ORF length of 867 nucleotides. The predicted protein is 288 amino acids long. This sequence codes for a full length protein. It is classified as (superfamily/group/family):
- 15 Protease, Serine, Trypsin. The cytogenetic position of this gene is 16p13.3. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none
- SGPr504, SEQ ID NO:57, SEQ ID NO:116 is 135 nucleotides long. The open
- 20 reading frame starts at position 1 and ends at position 135, giving an ORF length of 135 nucleotides. The predicted protein is 44 amino acids long. This sequence codes for a partial length protein. It is classified as (superfamily/group/family): Protease, Serine, Trypsin. The cytogenetic position of this gene is unknown. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none
- 25 SGPr469, SEQ ID NO:58, SEQ ID NO:117 is 138 nucleotides long. The open reading frame starts at position 1 and ends at position 138, giving an ORF length of 138 nucleotides. The predicted protein is 45 amino acids long. This sequence codes for a partial length protein. It is classified as (superfamily/group/family): Protease,

WO 02/00860

PCT/US01/20171

Serine, Trypsin. The cytogenetic position of this gene is unknown. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs:

AW753029, Z19070. The nucleic acid contains short repetitive sequence (the position and sequence of the repeat): 55 gggattgtgagctggggc 72.

5

SGPr400, SEQ ID NO:59, SEQ ID NO:118 is 930 nucleotides long. The open reading frame starts at position 1 and ends at position 930, giving an ORF length of 930 nucleotides. The predicted protein is 309 amino acids long. This sequence codes for a partial length protein. It is classified as (superfamily/group/family):

10

Protease, Serine, Trypsin. The cytogenetic position of this gene is 4q32. This sequence is represented in the database of public ESTs (dbEST) by the following ESTs: none

15

DESCRIPTION OF NOVEL PROTEASE POLYPEPTIDES

SGPr397, SEQ ID NO:1, SEQ ID NO:60 encodes a protein that is 315 amino acids long. It is classified as an Carboxypeptidase protease, of the Zn carboxypeptidase family. The protease domain(s) in this protein match the hidden Markov profile for

20

a Zn carboxypeptidase (PF00246) domain, from amino acid 139 to amino acid 280.

The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 146. Other domains identified within this protein are: Carboxypeptidase activation peptide (PF02244) from amino acid 41 to 120. The pro-segment moiety (activation peptide) is responsible for modulation of

25

folding and activity of the pro-enzyme (see http://pfam.wustl.edu/cgi-bin/getdesc?name=Propep_M14). The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NR/AA) with this protein sequence yielded the following results: Pscore = 3.10E-220; number of identical amino acids = 315; percent identity

WO 02/00860

PCT/US01/20171

= 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_065094.1; the name or description, and species, of the most similar protein in NRAA is: carboxypeptidase B precursor [Homo sapiens].

- 5 SGPr413, SEQ ID NO:2, SEQ ID NO:61 encodes a protein that is 374 amino acids long. It is classified as an Carboxypeptidase protease, of the Zn carboxypeptidase family. The protease domain(s) in this protein match the hidden Markov profile for a Zn carboxypeptidase (PF00246), from amino acid 50 to amino acid 291. The positions within the HMMR profile that match the protein sequence are from profile
- 10 position 1 to profile position 248. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 5.90E-93; number of identical amino acids = 146; percent identity = 49%; percent similarity = 68%; the accession number of the most similar entry in
- 15 NRAA is AAF01344.1; the name or description, and species, of the most similar protein in NRAA is: (AF190274) carboxypeptidase homolog [Bothrops jararaca].

- SGPr404, SEQ ID NO:3, SEQ ID NO:62 encodes a protein that is 529 amino acids long. It is classified as an Carboxypeptidase protease, of the Zn carboxypeptidase
- 20 family. The protease domain(s) in this protein match the hidden Markov profile for a Zn carboxypeptidase (PF00246), from amino acid 91 to amino acid 466. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 248. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of
- 25 amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 502; percent identity = 94%; percent similarity = 98%; the accession number of the most similar entry in NRAA is NP_061355.1; the name or description, and species, of the most similar protein in

WO 02/00860

PCT/US01/20171

NRAA is: carboxypeptidase X2 [Mus musculus].

5 SGPr536_1, SEQ ID NO:4, SEQ ID NO:63 encodes a protein that is 467 amino acids long. It is classified as an Cysteine protease, of the papain family. The protease domain(s) in this protein match the hidden Markov profile for a papain (PF00112), from amino acid 203 to amino acid 456. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 337. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.10E-276; number of identical amino acids = 467; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_071447.1; the name or description, and species, of the most similar protein in
10
15 NRAA is: P3ECSL [Homo sapiens].

SGPr414, SEQ ID NO:5, SEQ ID NO:64 encodes a protein that is 3353 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 1951 to amino acid 2045. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: Ubiquitin carboxyl-terminal hydrolases family 2 (UCH2b, PF00442) from amino acid 1701 to 1731. Ubiquitin carboxyl-terminal hydrolases (EC 3.1.2.15) (UCH) (deubiquitinating enzymes) are thiol proteases that recognize and hydrolyze the peptide bond at the C-terminal glycine of ubiquitin. These enzymes are involved in the processing of poly-ubiquitin precursors as well as that of ubiquitinated proteins. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein
20
25

WO 02/00860

PCT/US01/20171

sequence yielded the following results: Pscore = 0; number of identical amino acids = 1259; percent identity = 99%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_055524.1; the name or description, and species, of the most similar protein in NRAA is: KIAA0570 gene product [Homo sapiens].

SGPr430, SEQ ID NO:6, SEQ ID NO:65 encodes a protein that is 980 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 886 to amino acid 951. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from amino acids 342 to 373. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 930; percent identity = 99%; percent similarity = 99%; the accession number of the most similar entry in NRAA is BAB13420.1; the name or description, and species, of the most similar protein in NRAA is: (AB046814) KIAA1594 protein [Homo sapiens].

SGPr496_1, SEQ ID NO:7, SEQ ID NO:66 encodes a protein that is 953 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 875 to amino acid 935. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from 593 to 694; and a Zn-finger domain (PF02148), found in ubiquitin-hydrolases, from 465 to 534. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the

WO 02/00860

PCT/US01/20171

public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.00E-190; number of identical amino acids = 496; percent identity = 95%; percent similarity = 98%; the accession number of the most similar entry in NRAA is AAF66953.1; the name or description, and species, of the most similar protein in NRAA is: (AF229643) ubiquitin specific protease [Mus musculus].

SGPr495, SEQ ID NO:8, SEQ ID NO:67 encodes a protein that is 783 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 695 to amino acid 781. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from 190 to 221; and Zn-finger in ubiquitin-hydrolases (PF02148) from 465 to 534. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.40E-176; number of identical amino acids = 282; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is AAH05991.1; the name or description, and species, of the most similar protein in NRAA is: (BC005991) Unknown (protein for MGC:14793) [Homo sapiens].

SGPr407, SEQ ID NO:9, SEQ ID NO:68 encodes a protein that is 752 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 481 to amino acid 491. The positions within the HMMR profile that match the protein sequence are from profile position 80 to profile position 90. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of

WO 02/00860

PCT/US01/20171

amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.60E-40; number of identical amino acids = 80; percent identity = 76%; percent similarity = 84%; the accession number of the most similar entry in NRAA is NP_036607.1; the name or description, and species, of the most similar protein in NRAA is: ubiquitin specific protease 23; NEDD8-specific protease [Homo sapiens].

SGPr453, SEQ ID NO:10, SEQ ID NO:69 encodes a protein that is 712 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 615 to amino acid 677. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from 273 to 304; and Zn-finger in ubiquitin-hydrolases (PF02148) from amino acids 29 to 99. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 712; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_115523.1; the name or description, and species, of the most similar protein in NRAA is: hypothetical protein DKFZp434D0127 [Homo sapiens].

SGPr445, SEQ ID NO:11, SEQ ID NO:70 encodes a protein that is 289 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 190 to amino acid 221. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 32. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of

WO 02/00860

PCT/US01/20171

amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 3.60E-185; number of identical amino acids = 289; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is AAH05991.1; the name or description, and species, of the most similar protein in NRAA is: (BC005991) Unknown (protein for MGC:14793) [Homo sapiens].

SGPr401_1, SEQ ID NO:12, SEQ ID NO:71 encodes a protein that is 366 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 292 to amino acid 364. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from amino acids 35 to 66. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 7.30E-254; number of identical amino acids = 366; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_073743.1; the name or description, and species, of the most similar protein in NRAA is: hypothetical protein FLJ12552 [Homo sapiens].

SGPr408, SEQ ID NO:13, SEQ ID NO:72 encodes a protein that is 1287 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 395 to amino acid 475. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from amino acids 100 to 131. The results of a

WO 02/00860

PCT/US01/20171

Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 1287; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is BAB55063.1; the name or description, and species, of the most similar protein in NRAA is: (AK027362) unnamed protein product [Homo sapiens].

SGPr480, SEQ ID NO:14, SEQ ID NO:73 encodes a protein that is 1604 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 1506 to amino acid 1566. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from 734 to 765; and two EF hands (PF00036) from 232 to 260, and from 268 to 296. Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand (see <http://www.expasy.ch/cgi-bin/prosite-search-ac?PDO00018>). This type of domain consists of a twelve residue loop flanked on both side by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. This protein has a putative CAAX motif (CVLQ) which may direct it to the membrane fraction. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 1272; percent identity = 99%; percent similarity = 99%; the accession number of the most similar entry in NRAA is NP_115971.1; the name or description, and species, of the most similar protein in NRAA is: ubiquitin specific protease [Homo sapiens].

WO 02/00860

PCT/US01/20171

- SGPr431, SEQ ID NO:15, SEQ ID NO:74 encodes a protein that is 1042 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 836 to amino acid 948. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from 445 to 476. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.40E-251; number of identical amino acids = 397; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_115946.1; the name or description, and species, of the most similar protein in NRAA is: HP43.8KD protein [Homo sapiens].
- SGPr429, SEQ ID NO:16, SEQ ID NO:75 encodes a protein that is 1033 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 332 to amino acid 419. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from 89 to 120. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.50E-250; number of identical amino acids = 368; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_115612.1; the name or description, and species, of the most similar protein in NRAA is: hypothetical protein FLJ23277 [Homo sapiens]. This protein has a transmembrane domain from amino acid 87 to amino acid 109.

WO 02/00860

PCT/US01/20171

SGPr503, SEQ ID NO:17, SEQ ID NO:76 encodes a protein that is 517 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 432 to amino acid 501.

5 The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from 68 to 99. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of

10 amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 508; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is AAH04868.1; the name or description, and species, of the most similar protein in NRAA is: (BC004868) Unknown (protein for MGC:10702) [Homo sapiens]. This

15 protein has a transmembrane domain from amino acid 35 to amino acid 57.

SGPr427, SEQ ID NO:18, SEQ ID NO:77 encodes a protein that is 1123 amino acids long. It is classified as a Cysteine protease, of the UCH2b family. The protease domain(s) in this protein match the hidden Markov profile for a Ubiquitin

20 carboxyl-terminal hydrolase family 2b (PF00443), from amino acid 648 to amino acid 709. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 72. Other domains identified within this protein are: UCH2b (PF00442) from 101 to 129. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the

25 public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.80E-92; number of identical amino acids = 269; percent identity = 36%; percent similarity = 53%; the accession number of the most similar entry in NRAA is AAF47260.1; the name or description, and species, of the

WO 02/00860

PCT/US01/20171

most similar protein in NRAA is: (AE003465) CG3872 gene product [Drosophila melanogaster].

5 SGPr092, SEQ ID NO:19, SEQ ID NO:78 encodes a protein that is 261 amino acids long. It is classified as a Metalloprotease protease, of the PepM10 family. The protease domain(s) in this protein match the hidden Markov profile for a Peptidase_M10 (PF00413), from amino acid 75 to amino acid 194. The positions within the HMMR profile that match the protein sequence are from profile position 49 to profile position 168. Other domains identified within this protein are: ADAM
10 domain, amino acid 207 to 218. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 4.70E-171; number of identical amino acids = 261; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA
15 is XP_011971.1; the name or description, and species, of the most similar protein in NRAA is: matrix metalloproteinase 26 [Homo sapiens].

SGPr359, SEQ ID NO:20, SEQ ID NO:79 encodes a protein that is 483 amino acids long. It is classified as a Metalloprotease protease, of the PepM10 family. The
20 protease domain(s) in this protein match the hidden Markov profile for a Peptidase_M10 (PF00413), from amino acid 44 to amino acid 212. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 168. Other domains identified within this protein are: 3 x Hemopexin (PF00045) domains from 302 to 403. Hemopexin is a serum
25 glycoprotein that binds heme and transports it to the liver for breakdown and iron recovery, after which the free hemopexin returns to the circulation. Hemopexin-like domains have been found in two types of proteins: - in vitronectin, a cell adhesion and spreading factor found in plasma and tissues and in most members of the matrix metalloproteinases family (matrixins), including MMP-1, MMP-2, MMP-3, MMP-8,

WO 02/00860

PCT/US01/20171

MMP-9, MMP-10, MMP-11, MMP-12, MMP-13, MMP-14, MMP-15, MMP-16, MMP-17, MMP-18, MMP-19, MMP-20, MMP-24, and MMP-25 (see <http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00023>). The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 483; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_004762.1; the name or description, and species, of the most similar protein in NRAA is: matrix metalloproteinase 20 preproprotein; enamelysin [Homo sapiens]. This protein has a transmembrane domain from amino acid 7 to amino acid 29. This may function as a signal peptide.

SGPr104_1, SEQ ID NO:21, SEQ ID NO:80 encodes a protein that is 765 amino acids long. It is classified as a Metalloprotease protease, of the PepM13 family. The protease domain(s) in this protein match the hidden Markov profile for a Peptidase_M13 (PF01431), from amino acid 561 to amino acid 764. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 222. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 765; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_055508.1; the name or description, and species, of the most similar protein in NRAA is: KIAA0604 gene product [Homo sapiens]. This protein has a transmembrane domain from amino acid 61 to amino acid 83.

SGPr303, SEQ ID NO:22, SEQ ID NO:81 encodes a protein that is 418 amino acids long. It is classified as a Metalloprotease protease, of the PepM2 family. The protease domain(s) in this protein match the hidden Markov profile for a

WO 02/00860

PCT/US01/20171

Peptidase_M1 (PF01433), from amino acid 10 to amino acid 397. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 416. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.20E-284; number of identical amino acids = 407; percent identity = 97%; percent similarity = 98%; the accession number of the most similar entry in NRAA is CAA10709.1; the name or description, and species, of the most similar protein in NRAA is: (AJ132583) puromycin sensitive aminopeptidase [Homo sapiens].

SGPr402_1, SEQ ID NO:23, SEQ ID NO:82 encodes a protein that is 755 amino acids long. It is classified as a Serine protease, of the subtilase family. The protease domain(s) in this protein match the hidden Markov profile for a subtilase (PF00082), from amino acid 118 to amino acid 437. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 360. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 513; percent identity = 82%; percent similarity = 89%; the accession number of the most similar entry in NRAA is P29121; the name or description, and species, of the most similar protein in NRAA is: NEUROENDOCRINE CONVERTASE 3 PRECURSOR [Mus musculus].

SGPr434, SEQ ID NO:24, SEQ ID NO:83 encodes a protein that is 391 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a p20-ICE (PF00656), from amino acid 39 to amino acid 46. The positions within the HMMR profile that match the protein sequence are from profile position 129 to profile position 136. The results of a Smith Waterman search (PAM100, gap open and extend penalties of

WO 02/00860

PCT/US01/20171

12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 6.20E-43; number of identical amino acids = 104; percent identity = 42%; percent similarity = 59%; the accession number of the most similar entry in NRAA is NP_036164.1; the name or
5 description, and species, of the most similar protein in NRAA is: transmembrane tryptase [Mus musculus].

SGPr446_1, SEQ ID NO:25, SEQ ID NO:84 encodes a protein that is 226 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease
10 domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 13 to amino acid 227. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 242. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein
15 sequence yielded the following results: Pscore = 2.50E-40; number of identical amino acids = 107; percent identity = 45%; percent similarity = 57%; the accession number of the most similar entry in NRAA is NP_038949.1; the name or description, and species, of the most similar protein in NRAA is: distal intestinal serine protease [Mus musculus].

20 SGPr447, SEQ ID NO:26, SEQ ID NO:85 encodes a protein that is 295 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 33 to amino acid 270. The positions within the HMMR profile that
25 match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.00E-97; number of identical amino acids = 167; percent identity = 60%; percent similarity = 77%; the accession

WO 02/00860

PCT/US01/20171

number of the most similar entry in NRAA is BAB30277.1; the name or description, and species, of the most similar protein in NRAA is: (AK016509) putative [Mus musculus].

- 5 SGPr432_1, SEQ ID NO:27, SEQ ID NO:86 encodes a protein that is 628 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 117 to amino acid 343. The positions within the HMMR profile that match the protein sequence are from profile position 6 to profile position 259.
- 10 The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 3.70E-56; number of identical amino acids = 95; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_076869.1; the name or
- 15 description, and species, of the most similar protein in NRAA is: hypothetical protein IMAGE3455200 [Homo sapiens]. This protein has two transmembrane domains from amino acid 10 to amino acid 29, and from 82 to 99. The region from amino acid 10 to 29 may function as a signal peptide.
- 20 SGPr529, SEQ ID NO:28, SEQ ID NO:87 encodes a protein that is 276 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 184 to amino acid 187. The positions within the HMMR profile that match the protein sequence are from profile position 413 to profile position 416.
- 25 The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.70E-184; number of identical amino acids = 276; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_002767.1; the name or

WO 02/00860

PCT/US01/20171

description, and species, of the most similar protein in NRAA is: kallikrein 10; protease, serine-like, 1 [Homo sapiens].

SGPr428_1, SEQ ID NO:29, SEQ ID NO:88 encodes a protein that is 285 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 24 to amino acid 246. The positions within the HMMR profile that match the protein sequence are from profile position 7 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.90E-58; number of identical amino acids = 92; percent identity = 53%; percent similarity = 73%; the accession number of the most similar entry in NRAA is BAB24215.1; the name or description, and species, of the most similar protein in NRAA is: (AK005740) putative [Mus musculus]. This protein has a transmembrane domain from amino acid 262 to amino acid 284.

SGPr425, SEQ ID NO:30, SEQ ID NO:89 encodes a protein that is 413 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 287 to amino acid 306. The positions within the HMMR profile that match the protein sequence are from profile position 387 to profile position 406. This protein has a putative CAAX motif (CAYG) which may direct it to the plasma membrane. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 5.80E-268; number of identical amino acids = 412; percent identity = 99%; percent similarity = 99%; the accession number of the most similar entry in NRAA is CAC35071.1; the name or

WO 02/00860

PCT/US01/20171

description, and species, of the most similar protein in NRAA is: (AL121939)
 dJ223E3.1 (putative secreted protein ZSIG13) [Homo sapiens].

SGPr548, SEQ ID NO:31, SEQ ID NO:90 encodes a protein that is 320 amino acids
 5 long. It is classified as a Serine protease, of the trypsin family. The protease
 domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089),
 from amino acid 86 to amino acid 313. The positions within the HMMR profile that
 match the protein sequence are from profile position 1 to profile position 259. The
 results of a Smith Waterman search (PAM100, gap open and extend penalties of 12
 10 and 2) of the public database of amino acid sequences (NRAA) with this protein
 sequence yielded the following results: Pscore = 2.60E-168; number of identical
 amino acids = 256; percent identity = 100%; percent similarity = 100%; the
 accession number of the most similar entry in NRAA is AAG09469.1; the name or
 description, and species, of the most similar protein in NRAA is: (AF242195)
 15 KLK15 [Homo sapiens].

SGPr396, SEQ ID NO:32, SEQ ID NO:91 encodes a protein that is 328 amino acids
 long. It is classified as a Serine protease, of the trypsin family. The protease
 domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089),
 20 from amino acid 28 to amino acid 262. The positions within the HMMR profile that
 match the protein sequence are from profile position 1 to profile position 259. The
 results of a Smith Waterman search (PAM100, gap open and extend penalties of 12
 and 2) of the public database of amino acid sequences (NRAA) with this protein
 sequence yielded the following results: Pscore = 1.60E-56; number of identical
 25 amino acids = 111; percent identity = 44%; percent similarity = 61%; the accession
 number of the most similar entry in NRAA is BAA84941.1; the name or
 description, and species, of the most similar protein in NRAA is: (AB018694)
 epidermis specific serine protease [Xenopus laevis].

WO 02/00860

PCT/US01/20171

- SGPr426, SEQ ID NO:33, SEQ ID NO:92 encodes a protein that is 425 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089),
- 5 from amino acid 194 to amino acid 419. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 7.70E-93; number of identical
- 10 amino acids = 181; percent identity = 43%; percent similarity = 61%; the accession number of the most similar entry in NRAA is NP_054777.1; the name or description, and species, of the most similar protein in NRAA is: DESC1 protein [Homo sapiens]. This protein has a transmembrane domain from amino acid 30 to amino acid 52. This region could function as a signal peptide.
- 15
- SGPr552, SEQ ID NO:34, SEQ ID NO:93 encodes a protein that is 221 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089),
- 20 from amino acid 2 to amino acid 222. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 255. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.20E-45; number of identical
- 25 amino acids = 96; percent identity = 42%; percent similarity = 59%; the accession number of the most similar entry in NRAA is NP_054777.1; the name or description, and species, of the most similar protein in NRAA is: DESC1 protein [Homo sapiens].

WO 02/00860

PCT/US01/20171

SGPr405, SEQ ID NO:35, SEQ ID NO:94 encodes a protein that is 948 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 218 to amino acid 406. The positions within the HMMR profile that match the protein sequence are from profile position 60 to profile position 259. Other domains identified within this protein are: two additional trypsin domains, from amino acids 419 to 496, and from amino acids 636 to 761. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.10E-30; number of identical amino acids = 111; percent identity = 54%; percent similarity = 65%; the accession number of the most similar entry in NRAA is P19236; the name or description, and species, of the most similar protein in NRAA is: MASTOCYTOMA PROTEASE PRECURSOR [Canis familiaris].

SGPr485_1, SEQ ID NO:36, SEQ ID NO:95 encodes a protein that is 352 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 68 to amino acid 295. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 7.20E-133; number of identical amino acids = 223; percent identity = 94%; percent similarity = 96%; the accession number of the most similar entry in NRAA is BAB03569.1; the name or description, and species, of the most similar protein in NRAA is: (AB046651) hypothetical protein [Macaca fascicularis].

WO 02/00860

PCT/US01/20171

- SGPr534, SEQ ID NO:37, SEQ ID NO:96 encodes a protein that is 263 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 34 to amino acid 256. The positions within the HMMR profile that
- 5 match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 3.60E-165; number of identical amino acids = 253; percent identity = 96%; percent similarity = 98%; the accession
- 10 number of the most similar entry in NRAA is NP_001897.1; the name or description, and species, of the most similar protein in NRAA is: chymotrypsinogen B1 [Homo sapiens]. This protein has a transmembrane domain from amino acid 2 to amino acid 24. This region could function as a signal peptide.
- 15 SGPr390, SEQ ID NO:38, SEQ ID NO:97 encodes a protein that is 1128 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 896 to amino acid 1122. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259.
- 20 Other domains identified within this protein are: two trypsin domains, from amino acids 264 to 500, and from amino acids 573 to 800. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.60E-53; number of identical amino acids = 135;
- 25 percent identity = 46%; percent similarity = 59%; the accession number of the most similar entry in NRAA is BAB23684.1; the name or description, and species, of the most similar protein in NRAA is: (AK004939) putative [Mus musculus]. This protein has a transmembrane domain from amino acid 28 to amino acid 50. This region could function as a signal peptide.

WO 02/00860

PCT/US01/20171

SGPr521, SEQ ID NO:39, SEQ ID NO:98 encodes a protein that is 253 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089),
5 from amino acid 30 to amino acid 245. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.30E-155; number of identical
10 amino acids = 253; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_005037.1; the name or description, and species, of the most similar protein in NRAA is: kallikrein 7 (chymotryptic, stratum corneum); protease, serine, 6 (chymotryptic, stratum corneum) [Homo sapiens].
15

SGPr530_1, SEQ ID NO:40, SEQ ID NO:99 encodes a protein that is 271 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089),
20 from amino acid 14 to amino acid 255. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.10E-95; number of identical
amino acids = 142; percent identity = 100%; percent similarity = 100%; the
25 accession number of the most similar entry in NRAA is CAC12709.1; the name or description, and species, of the most similar protein in NRAA is: (AL136097) bA62C3.1 (similar to testicular serine protease) [Homo sapiens].

WO 02/00860

PCT/US01/20171

SGPr520, SEQ ID NO:41, SEQ ID NO:100 encodes a protein that is 578 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 73 to amino acid 306. The positions within the HMMR profile that
5 match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.50E-83; number of identical amino acids = 158; percent identity = 73%; percent similarity = 83%; the accession
10 number of the most similar entry in NRAA is BAB24587.1; the name or description, and species, of the most similar protein in NRAA is: (AK006434) putative [Mus musculus].

SGPr455, SEQ ID NO:42, SEQ ID NO:101 encodes a protein that is 970 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 433 to amino acid 674. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. Other domains identified within this protein are: Trypsin, from amino acid 4 to 156;
20 and three 3 x CUB domains (PF00431) from amino acid 175 to 812. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 5.90E-179; number of identical amino acids = 386; percent identity = 41%; percent similarity = 58%; the accession number of
25 the most similar entry in NRAA is T30337; the name or description, and species, of the most similar protein in NRAA is: polyprotein - African clawed frog.

SGPr507_2, SEQ ID NO:43, SEQ ID NO:102 encodes a protein that is 265 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease

WO 02/00860

PCT/US01/20171

domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 42 to amino acid 135. The positions within the HMMR profile that match the protein sequence are from profile position 35 to profile position 148. Other domains identified within this protein are: Trypsin domain from amino acid

5 247 to 258. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.40E-121; number of identical amino acids = 195; percent identity = 73%; percent similarity = 81%; the accession number of the most similar entry in NRAA is

10 NP_080593.1; the name or description, and species, of the most similar protein in NRAA is: RIKEN cDNA 1700016G05 gene [Mus musculus].

SGPr559, SEQ ID NO:44, SEQ ID NO:103 encodes a protein that is 454 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease

15 domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 217 to amino acid 444. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. Other domains identified within this protein are: Low-density lipoprotein receptor domain class A (PF00057), from amino acid 71 to 109. LDL-receptors the class A

20 domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands. The repeat has been shown to consist of a beta-hairpin structure followed by a series of beta turns (see <http://www.expasy.ch/cgi-bin/get-prodoc-entry?PDOC00929>). The results of a

25 Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.40E-288; number of identical amino acids = 454; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_076927.1; the name or description, and

WO 02/00860

PCT/US01/20171

species, of the most similar protein in NRAA is: transmembrane protease, serine 3 [Homo sapiens]. This protein has a transmembrane domain from amino acid 49 to amino acid 71.

- 5 SGPr567_1, SEQ ID NO:45, SEQ ID NO:104 encodes a protein that is 537 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 296 to amino acid 524. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259.
- 10 The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.70E-135; number of identical amino acids = 534; percent identity = 99%; percent similarity = 99%; the accession number of the most similar entry in NRAA is NP_114435.1; the name or
- 15 description, and species, of the most similar protein in NRAA is: mosaic serine protease [Homo sapiens].

- SGPr479_1, SEQ ID NO:46, SEQ ID NO:105 encodes a protein that is 326 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 60 to amino acid 288. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein
- 20 sequence yielded the following results: Pscore = 1.70E-39; number of identical amino acids = 107; percent identity = 42%; percent similarity = 57%; the accession number of the most similar entry in NRAA is NP_114154.1; the name or
- 25 description, and species, of the most similar protein in NRAA is: marapsin [Homo sapiens].

WO 02/00860

PCT/US01/20171

SGPr489_1, SEQ ID NO:47, SEQ ID NO:106 encodes a protein that is 556 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089),
5 from amino acid 56 to amino acid 257. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 227. Other domains identified within this protein are: 2 x CUB domains (PF00431) from amino acids 304 to 503. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences
10 (NRAA) with this protein sequence yielded the following results: Pscore = 2.70E-90; number of identical amino acids = 194; percent identity = 37%; percent similarity = 54%; the accession number of the most similar entry in NRAA is T30338; the name or description, and species, of the most similar protein in NRAA is: oviductin - [Xenopus laevis].
15

SGPr465_1, SEQ ID NO:48, SEQ ID NO:107 encodes a protein that is 297 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 2 to amino acid 240. The positions within the HMMR profile that
20 match the protein sequence are from profile position 12 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.70E-76; number of identical amino acids = 144; percent identity = 48%; percent similarity = 66%; the accession
25 number of the most similar entry in NRAA is NP_033381.1; the name or description, and species, of the most similar protein in NRAA is: testicular serine protease 1 [Mus musculus].

WO 02/00860

PCT/US01/20171

SGPr524_1, SEQ ID NO:49, SEQ ID NO:108 encodes a protein that is 850 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 613 to amino acid 842. The positions within the HMMR profile
5 that match the protein sequence are from profile position 1 to profile position 259. Other domains identified within this protein are: three Low-density lipoprotein receptor domain class A domains (PF00057) from 489 to 603. LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of
10 positively charged sequences in LDLR's ligands. The repeat has been shown to consist of a beta-hairpin structure followed by a series of beta turns (see <http://www.expasy.ch/cgi-bin/get-prodoc-entry?PDOC00929>). The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence
15 yielded the following results: Pscore = 1.30E-79; number of identical amino acids = 193; percent identity = 41%; percent similarity = 55%; the accession number of the most similar entry in NRAA is BAB23684.1; the name or description, and species, of the most similar protein in NRAA is: (AK004939) putative [*Mus musculus*]. This protein has a transmembrane domain from amino acid 77 to amino acid 99.

20 SGPr422, SEQ ID NO:50, SEQ ID NO:109 encodes a protein that is 447 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 216 to amino acid 441. The positions within the HMMR profile
25 that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 4.90E-80; number of identical amino acids = 173; percent identity = 39%; percent similarity = 59%; the accession

WO 02/00860

PCT/US01/20171

number of the most similar entry in NRAA is NP_054777.1; the name or description, and species, of the most similar protein in NRAA is: DESC1 protein [Homo sapiens]. This protein has a transmembrane domain from amino acid 32 to amino acid 54. This region could function as a signal peptide.

5

SGPr538, SEQ ID NO:51, SEQ ID NO:110 encodes a protein that is 457 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 218 to amino acid 448. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 9.1e-315; number of identical amino acids = 457; percent identity = 100%; percent similarity = 100%; the accession number of the most similar entry in NRAA is NP_110397.1; the name or description, and species, of the most similar protein in NRAA is: spinesin [Homo sapiens]. This protein has a transmembrane domain from amino acid 48 to amino acid 70. This region could function as a signal peptide.

10

15

20

25

SGPr527_1, SEQ ID NO:52, SEQ ID NO:111 encodes a protein that is 818 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 47 to amino acid 286. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. Other domains identified within this protein are: two additional trypsin domains, from 323 to 454, and from 564 to 679. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 1.30E-52; number of identical amino acids = 114; percent identity = 42%; percent

WO 02/00860

PCT/US01/20171

similarity = 59%; the accession number of the most similar entry in NRAA is AAH03851.1; the name or description, and species, of the most similar protein in NRAA is: (BC003851) Similar to protease, serine, 8 (prostasin) [Mus musculus].

- 5 SGPt542, SEQ ID NO:53, SEQ ID NO:112 encodes a protein that is 284 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 35 to amino acid 259. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.70E-41; number of identical amino acids = 110; percent identity = 43%; percent similarity = 58%; the accession number of the most similar entry in NRAA is NP_005308.1; the name or description, and species, of the most similar protein in NRAA is: granzyme M precursor; lymphocyte met-ase 1 [Homo sapiens].

- SGPt551, SEQ ID NO:54, SEQ ID NO:113 encodes a protein that is 802 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 568 to amino acid 797. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. Other domains identified within this protein are: three low-density lipoprotein receptor domain class A domains (PF00057) from 447 to 559. LDL-receptors the class A domains form the binding site for LDL and calcium. The acidic residues between the fourth and sixth cysteines are important for high-affinity binding of positively charged sequences in LDLR's ligands. The repeat has been shown to consist of a beta-hairpin structure followed by a series of beta turns (see <http://www.expasy.ch/cgi-bin/get-prodoc-entry?PDOC00929>). The results of a

WO 02/00860

PCT/US01/20171

- Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 0; number of identical amino acids = 675; percent identity = 84%; percent similarity = 90%; the accession number of the most similar entry in NRAA is BAB23684.1; the name or description, and species, of the most similar protein in NRAA is: (AK004939) putative [Mus musculus]. This protein has a transmembrane domain from amino acid 44 to amino acid 66. This region could function as a signal peptide.
- SGPr451, SEQ ID NO:55, SEQ ID NO:114 encodes a protein that is 359 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 89 to amino acid 324. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 9.90E-41; number of identical amino acids = 101; percent identity = 39%; percent similarity = 59%; the accession number of the most similar entry in NRAA is NP_072152.1; the name or description, and species, of the most similar protein in NRAA is: adrenal secretory serine protease precursor [Rattus norvegicus].

- SGPr452_1, SEQ ID NO:56, SEQ ID NO:115 encodes a protein that is 288 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 73 to amino acid 280. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein

WO 02/00860

PCT/US01/20171

sequence yielded the following results: Pscore = 1.40E-81; number of identical amino acids = 142; percent identity = 57%; percent similarity = 72%; the accession number of the most similar entry in NRAA is AAK15264.1; the name or description, and species, of the most similar protein in NRAA is: (AF305425)

5 implantation serine proteinase 2 [Mus musculus].

SGPr504, SEQ ID NO:57, SEQ ID NO:116 encodes a protein that is 44 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089),
 10 from amino acid 1 to amino acid 45. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 52. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.40E-13; number of identical
 15 amino acids = 26; percent identity = 61%; percent similarity = 88%; the accession number of the most similar entry in NRAA is NP_002095.1; the name or description, and species, of the most similar protein in NRAA is: granzyme K precursor; granzyme 3; granzyme K (serine protease, granzyme 3); tryptase II [Homo sapiens].

20 SGPr469, SEQ ID NO:58, SEQ ID NO:117 encodes a protein that is 45 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 1 to amino acid 46. The positions within the HMMR profile that
 25 match the protein sequence are from profile position 210 to profile position 259. The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.20E-17; number of identical amino acids = 32; percent identity = 69%; percent similarity = 84%; the accession

WO 02/00860

PCT/US01/20171

number of the most similar entry in NRAA is BAB30277.1; the name or description, and species, of the most similar protein in NRAA is: (AK016509) putative [Mus musculus].

- 5 SGPr400, SEQ ID NO:59, SEQ ID NO:118 encodes a protein that is 309 amino acids long. It is classified as a Serine protease, of the trypsin family. The protease domain(s) in this protein match the hidden Markov profile for a trypsin (PF00089), from amino acid 133 to amino acid 281. The positions within the HMMR profile that match the protein sequence are from profile position 1 to profile position 198.
- 10 The results of a Smith Waterman search (PAM100, gap open and extend penalties of 12 and 2) of the public database of amino acid sequences (NRAA) with this protein sequence yielded the following results: Pscore = 2.30E-16; number of identical amino acids = 72; percent identity = 38%; percent similarity = 48%; the accession number of the most similar entry in NRAA is NP_036164.1; the name or
- 15 description, and species, of the most similar protein in NRAA is: transmembrane tryptase [Mus musculus].

EXAMPLE 2: Expression Analysis of Mammalian Proteases

Materials and Methods

20 Quantitative PCR Analysis

RNA is isolated from a variety of normal human tissues and cell lines. Single stranded cDNA is synthesized from 10 µg of each RNA as described above using the Superscript Preamplification System (GibcoBRL). These single strand templates are then linearly amplified with a pair of specific primers in a real time

25 PCR reaction on a Light Cycler (Roche Molecular Biochemical). Graphical readout can provide quantitative analysis of the relative abundance of the targeted gene in the total RNA preparation.

DNA Array Based Expression Analysis

WO 02/00860

PCT/US01/20171

DNA-free RNA is isolated from a variety of normal human tissues, cryostat sections, and cell lines. Single stranded cDNA is synthesized from 10ug RNA or 1ug mRNA using a modification of the SMART PCR cDNA synthesis technique (Clontech). The procedure can be modified to allow asymmetric labeling of the 5' and 3' ends of each transcript with a unique oligonucleotide sequence. The resulting sscDNAs are then linearly amplified using Advantage long-range PCR (Clontech) on a Light Cycler PCR machine. Reactions are halted when the graphical real-time display demonstrates the products have begun to plateau. The double stranded cDNA products are purified using Millipore DNA purification matrix, dried, resuspended, quantified, and analyzed on an agarose gel. The resulting elements are referred to as "tissue cDNAs".

Tissue cDNAs are spotted onto GAPS coated glass slides (Corning) using a Genetic Microsystems (GMS) arrayer at 500 ng/ul.

Fluorescent labeled oligonucleotides are synthesized to each novel exon, ensuring they contained internal mismatches with the closest known homologue. Typically oligos are 45 nucleotides long, labeled on the 5' end with Cy5.

Exon-specific Cy5-labeled oligos are hybridized to the tissue cDNAs arrayed onto glass slides, and washed using standard buffers and conditions. Hybridizing signals are then quantified using a GMS Scanner.

Alternatively, tissue cDNAs are manually spotted onto Nylon membranes using a 384 pin replicator, and hybridized to ³²P-end labeled oligo probes.

Tissue cDNAs are generated from multiple RNA templates selected to provide information of relevance to the disease areas of interest and to reflect the biological mechanism of action for each protease. These templates include: human tumor cell lines, cryostat sections of primary human tumors and 32 normal human tissues to identify cancer-related genes; sections of normal, Alzheimer's, Parkinson's, and Schizophrenia brain regions for CNS-related genes; normal and diabetic or obese skeletal muscle, adipose, or liver for metabolic-related genes; and

WO 02/00860

PCT/US01/20171

purified hematopoietic cells, and lymphoid tissues for immune-related genes. To characterize gene mechanism of action, tissue cDNAs are generated to reflect angiogenesis (cultured endothelial cells treated with VEGF ligand, anti-angiogenic drugs, or hypoxia), motility (A549 cells stimulated with HGF ligand, orthotopic metastases, primary tumors with matched metastatic tumors), cell cycle (Hela, H1299, and other cell lines synchronized by drug block and harvested at various times in the cell cycle), checkpoint integrity and DNA repair (p53 normal or defective cells treated with γ -radiation, UV, cis-platinum, or oxidative stress), and cell survival (cells induced to differentiate or at various stages of apoptosis).

EXAMPLE 3: Isolation of cDNAs Encoding Mammalian Proteases

Materials and Methods

Identification of novel clones

Total RNAs are isolated using the Guanidine Salts/Phenol extraction protocol of Chomczynski and Sacchi (P. Chomczynski and N. Sacchi, *Anal. Biochem.* 162:156 (1987)) from primary human tumors, normal and tumor cell lines, normal human tissues, and sorted human hematopoietic cells. These RNAs are used to generate single-stranded cDNA using the Superscript Preamplification System (GIBCO BRL, Gaithersburg, MD; Gerard, GF *et al.* (1989), FOCUS 11, 66) under conditions recommended by the manufacturer. A typical reaction uses 10 μ g total RNA with 1.5 μ g oligo(dT)₁₂₋₁₈ in a reaction volume of 60 μ L. The product is treated with RNaseH and diluted to 100 μ L with H₂O. For subsequent PCR amplification, 1-4 μ L of this sscDNA is used in each reaction.

Degenerate oligonucleotides are synthesized on an Applied Biosystems 3948 DNA synthesizer using established phosphoramidite chemistry, precipitated with ethanol and used unpurified for PCR. These primers are derived from the sense and antisense strands of conserved motifs within the catalytic domain of several proteases. Degenerate nucleotide residue designations are: N = A, C, G, or T; R =

WO 02/00860

PCT/US01/20171

A or G; Y = C or T; H = A, C or T not G; D = A, G or T not C; S = C or G; and W = A or T.

PCR reactions are performed using degenerate primers applied to multiple single-stranded cDNAs. The primers are added at a final concentration of 5 μ M each to a mixture containing 10 mM TrisHCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 200 μ M each deoxynucleoside triphosphate, 0.001% gelatin, 1.5 U AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus), and 1-4 μ L cDNA. Following 3 min denaturation at 95 °C, the cycling conditions are 94 °C for 30 s, 50 °C for 1 min, and 72 °C for 1 min 45 s for 35 cycles. PCR fragments migrating between 300-350 bp are isolated from 2% agarose gels using the GeneClean Kit (Bio101), and T-A cloned into the pCRII vector (Invitrogen Corp. U.S.A.) according to the manufacturer's protocol.

Colonies are selected for mini plasmid DNA-preparations using Qiagen columns and the plasmid DNA is sequenced using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products are run on an ABI Prism 377 DNA Sequencer, and analyzed using the BLAST alignment algorithm (Altschul, S.F. *et al.*, *J.Mol.Biol.* 215: 403-10).

Additional PCR strategies are employed to connect various PCR fragments or ESTs using exact or near exact oligonucleotide primers. PCR conditions are as described above except the annealing temperatures are calculated for each oligo pair using the formula: $T_m = 4(G+C)+2(A+T)$.

Isolation of cDNA clones:

Human cDNA libraries are probed with PCR or EST fragments corresponding to protease-related genes. Probes are ³²P-labeled by random priming and used at 2x10⁶ cpm/mL following standard techniques for library screening. Pre-hybridization (3 h) and hybridization (overnight) are conducted at 42 °C in 5X SSC, 5X Denhart's solution, 2.5% dextran sulfate, 50 mM Na₂PO₄/NaHPO₄, pH 7.0, 50% formamide with 100 mg/mL denatured salmon sperm DNA. Stringent washes are

WO 02/00860

PCT/US01/20171

performed at 65 °C in 0.1X SSC and 0.1% SDS. DNA sequencing is carried out on both strands using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products are run on an ABI Prism 377 DNA Sequencer.

5

EXAMPLE 4: Expression Analysis of Mammalian Proteases

Materials and Methods

Northern blot analysis

Northern blots are prepared by running 10 µg total RNA isolated from 60 human tumor cell lines (such as HOP-92, EKVX, NCI-H23, NCI-H226, NCI-H322M, NCI-H460, NCI-H522, A549, HOP-62, OVCAR-3, OVCAR-4, OVCAR-5, OVCAR-8, IGROV1, SK-OV-3, SNB-19, SNB-75, U251, SF-268, SF-295, SF-539, CCRF-CEM, K-562, MOLT-4, HL-60, RPMI 8226, SR, DU-145, PC-3, HT-29, HCC-2998, HCT-116, SW620, Colo 205, HTC15, KM-12, UO-31, SN12C, A498, 15 CaKi1, RXF-393, ACHN, 786-0, TK-10, LOX IMVI, Malme-3M, SK-MEL-2, SK-MEL-5, SK-MEL-28, UACC-62, UACC-257, M14, MCF-7, MCF-7/ADR RES, Hs578T, MDA-MB-231, MDA-MB-435, MDA-N, BT-549, T47D), from human adult tissues (such as thymus, lung, duodenum, colon, testis, brain, cerebellum, cortex, salivary gland, liver, pancreas, kidney, spleen, stomach, uterus, prostate, 20 skeletal muscle, placenta, mammary gland, bladder, lymph node, adipose tissue), and 2 human fetal normal tissues (fetal liver, fetal brain), on a denaturing formaldehyde 1.2% agarose gel and transferring to nylon membranes.

Filters are hybridized with random primed [α - ^{32} P]dCTP-labeled probes synthesized from the inserts of several of the protease genes. Hybridization is 25 performed at 42 °C overnight in 6X SSC, 0.1% SDS, 1X Denhardt's solution, 100 µg/mL denatured herring sperm DNA with 1-2 x 10⁶ cpm/mL of ^{32} P-labeled DNA probes. The filters are washed in 0.1X SSC/0.1% SDS, 65 °C, and exposed on a Molecular Dynamics phosphorimager.

WO 02/00860

PCT/US01/20171

Quantitative PCR analysis

RNA is isolated from a variety of normal human tissues and cell lines.

Single stranded cDNA is synthesized from 10 µg of each RNA as described above
5 using the Superscript Preamplification System (GibcoBRL). These single strand
templates are then used in a 25 cycle PCR reaction with primers specific to each
clone. Reaction products are electrophoresed on 2% agarose gels, stained with
ethidium bromide and photographed on a UV light box. The relative intensity of the
STK-specific bands were estimated for each sample.

10

DNA Array Based Expression Analysis

Plasmid DNA array blots are prepared by loading 0.5 µg denatured plasmid
for each protease on a nylon membrane. The [γ^{32} P]dCTP labeled single stranded
DNA probes are synthesized from the total RNA isolated from several human
15 immune tissue sources or tumor cells (such as thymus, dendrocytes, mast cells,
monocytes, B cells (primary, Jurkat, RPMI8226, SR), T cells (CD8/CD4+, TH1,
TH2, CEM, MOLT4), K562 (megakaryocytes). Hybridization is performed at 42 °C
for 16 hours in 6X SSC, 0.1% SDS, 1X Denhardt's solution, 100 µg/mL denatured
herring sperm DNA with 10⁶ cpm/mL of [γ^{32} P]dCTP labeled single stranded probe.
20 The filters are washed in 0.1X SSC/0.1% SDS, 65 °C, and exposed for quantitative
analysis on a Molecular Dynamics phosphorimager.

EXAMPLE 5: Protease Gene ExpressionVector Construction25 Materials and MethodsExpression Vector Construction

Expression constructs are generated for some of the human cDNAs
including: a) full-length clones in a pCDNA expression vector; and b) a GST-fusion

WO 02/00860

PCT/US01/20171

construct containing the catalytic domain of the novel protease fused to the C-terminal end of a GST expression cassette; and c) a full-length clone containing a mutation within the predicted polypeptide cleaving site within the protease domain, inserted in the pCDNA vector.

- 5 These mutants of the protease might function as dominant negative constructs, and will be used to elucidate the function of these novel proteases.

EXAMPLE 6: Generation of Specific Immunoreagents to Proteases

Materials and Methods

- 10 Specific immunoreagents are raised in rabbits against KLH- or MAP-conjugated synthetic peptides corresponding to isolated protease polypeptides. C-terminal peptides were conjugated to KLH with glutaraldehyde, leaving a free C-terminus. Internal peptides were MAP-conjugated with a blocked N-terminus. Additional immunoreagents can also be generated by immunizing rabbits with the
- 15 bacterially expressed GST-fusion proteins containing the cytoplasmic domains of each novel PTK or STK.

 The various immune sera are first tested for reactivity and selectivity to recombinant protein, prior to testing for endogenous sources.

- 20 Western blots

 Proteins in SDS PAGE are transferred to immobilon membrane. The washing buffer is PBST (standard phosphate-buffered saline pH 7.4 + 0.1% Triton X-100). Blocking and antibody incubation buffer is PBST +5% milk. Antibody dilutions are varied from 1:1000 to 1:2000.

- 25

EXAMPLE 7: Recombinant Expression and Biological Assays for Proteases

Materials and Methods

Transient Expression of Proteases in Mammalian Cells

WO 02/00860

PCT/US01/20171

The pcDNA expression plasmids (10 µg DNA/100 mm plate) containing the protease constructs are introduced into 293 cells with lipofectamine (Gibco BRL). After 72 hours, the cells are harvested in 0.5 mL solubilization buffer (20 mM HEPES, pH 7.35, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM MgCl₂, 1 mM EGTA, 2 mM phenylmethylsulfonyl fluoride, 1 µg/mL aprotinin). Sample aliquots are resolved by SDS polyacrylamide gel electrophoresis (PAGE) on 6% acrylamide/0.5% bis-acrylamide gels and electrophoretically transferred to nitrocellulose. Non-specific binding is blocked by preincubating blots in Blotto (phosphate buffered saline containing 5% w/v non-fat dried milk and 0.2% v/v nonidet P-40 (Sigma)), and recombinant protein is detected using the various anti-peptide or anti-GST-fusion specific antisera.

In Vitro Protease Assays

In vitro Protease Assay Using Fluorogenic Peptides

Assays are carried out using a spectrofluorometer, such as Perkin-Elmer 204S. The standard reaction mixtures (100 µl) contains 200 mM Tris-HCl, pH8.5, and 200 µM fluorogenic peptide substrate. After enzyme addition, reaction mixtures are incubated at 37 °C for 30 min and terminated by addition of 1.9 ml of 125 mM ZnSO₄ (Brenner, C., and Fuller, R. S., 1992, *Proc. Natl. Acad. Sci. U. S. A.* 89:922-926). The precipitate is removed by centrifugation for 1 min in a microcentrifuge (15,000 × g), and the rate of product (7-amino-4-methyl-coumarin) released into the supernatant solution is determined fluorometrically [(excitation) = 385 nm, (emission) = 465 nm]. Examples of substrates used in the literature include: Boc-Gly-Arg-Arg-4-methylcoumaryl-7-amide (MCA), Boc-Gln-Arg-Arg-MCA, Z-Arg-Arg-MCA, and pGlu-Arg-Thr-Lys-Arg-MCA. Stock solutions (100 mM) are prepared by dissolving peptides in dimethyl sulfoxide that are then diluted in water to 1 mM working stock before use. (Details of this assay can be found in: R. Yosuf, *et al. J. Biol. Chem.*, Vol. 275, Issue 14, 9963-9969, April 7, 2000 which is

WO 02/00860

PCT/US01/20171

incorporated herein by reference in its entirety including any figures, tables, or drawings.)

Protease assay in intact cells using fluorogenic peptides-

5 Calpain activity is measured by the rate of generation of the fluorescent product, AMC, from intracellular thiol-conjugated Boc-Leu-Met-CMAC (Rosser, B. G., Powers, S. P., and Gores, G. J. (1993) *J. Biol. Chem.* 268, 23593-23600). Cells are dispersed, grown on glass coverslips, continuously superfused with physiologic saline solution at 37 °C, and sequentially imaged with a quantitative fluorescence
10 imaging system. At t = 0, Boc-Leu-Met-CMAC (10 µM, Molecular Probes) is introduced into the superfusion solution, and mean fluorescence intensity (excitation 350 nm, emission 470 nm) of individual cells is measured at 60-s intervals. At 10 min, TNF-alpha (30 ng/ml) is added to the superfusion solution with 10 µM Boc-Leu-Met-CMAC. The slope of the fluorescence change with respect to time
15 represents the intracellular calpain activity (Rosser, *et al.*, 1993, *J. Biol. Chem.* 268:23593-23600). For calpain assays in whole cell populations, suspension cultures of cells are loaded with 10 µM Boc-Leu-Met-CMAC, and changes in intracellular fluorescence are measured prior to and after TNF-alpha addition at 37 °C using a FACS Vantage system. Cellular fluorescence of AMC is measured using
20 a 360-nm excitation filter and a 405-nm long-pass emission filter. (Details of this assay can be found in: Han, *et al.*, 1999, *J Biol Chem*, 274:787-794 which is incorporated herein by reference in its entirety including any figures, tables, or drawings)

25 Protease assay using chromogenic substrates

The proteolytic activity of enzymes is measured using a commercially available assay system (Athena Environmental Sciences, Inc.). The assay employs a universal substrate of a dye-protein conjugate cross linked to a matrix. Protease activity is determined spectrophotometrically by measuring the absorbance of the

WO 02/00860

PCT/US01/20171

dye released from the matrix to the supernatant. Reaction vials containing the enzyme and substrate are incubated for 3 h at 37 °C. The activity is measured at different incubation times, and reactions are terminated by adding 500 µl of 0.2 N NaOH to each vial. The absorbance of the supernatant in each reaction vial is

5 measured at 450 nm. The proteolytic activity is monitored using 10 µl (approximately 10 µg) of purified protein incubated with 5 µg of -casein (Sigma) in 50 mM Tris-HCl (pH 7.5) for 30 min, 1 h or 2 h at 37 °C. The reaction products are resolved by SDS-polyacrylamide gel electrophoresis and proteins visualized by staining with Coomassie Blue (Details of this assay can be found in: Faccio, *et al.*,

10 2000, *J Biol Chem*, 275:2581-2588 which is incorporated herein by reference in its entirety including any figures, tables, or drawings).

Protease assay using radiolabeled substrate bound to membranes-

Unlabeled protease is mixed with radiolabeled substrate-containing

15 membranes in buffer (100 mM HEPES, 100 mM NaCl, 125 µM magnesium acetate, 125 µM zinc acetate, pH 7.5) and incubated at 30 °C. Typically, each reaction had a final volume of 80-100 µl. Each reaction is normalized to the same final concentration of lysis buffer components (25 mM Tris, 0.1 M sorbitol, 0.5 mM EDTA, 0.01% NaN₃, pH 7.5) because the amount of membranes added to each

20 reaction is varied. To examine metal ion specificity, reactions are assembled without substrate and pretreated with 1.125 mM 1,10-orthophenanthroline for 20 min on ice. Subsequently, metal ions and substrate-containing membranes are added, and reactions are initiated by incubation at 30 °C; the additions result in dilution of the 1,10-orthophenanthroline to a final concentration of 1 mM. The metal

25 ions are added in the form of acetate salts from 25-100 mM stock solutions (Zn²⁺, Mg²⁺, Cu²⁺, Co²⁺, or Ca²⁺) that are first acidified with 2 mM concentrated HCl and then neutralized with 1 mM HEPES, pH 7.5; this step is necessary to achieve full solubilization of zinc acetate. For analysis by immunoprecipitation, samples are diluted 10-20× with immunoprecipitation buffer (Berkower, C., and Michaelis, S.

(1991) *EMBO J.* 10:3777-3785) containing 0.1% SDS, cleared of insoluble material (13,000 × g for 5-10 min at 4 °C), and immunoprecipitated with substrate-specific antibody. Alternatively, samples are solubilized by SDS (final concentration, 0.5%), boiled for 3 min, and directly immunoprecipitated after dilution with
5 immunoprecipitation buffer. Immunoprecipitates are subjected to SDS-polyacrylamide gel electrophoresis as described, fixed for 7 min with 20% trichloroacetic acid, dried, and exposed to a PhosphorImager screen for detection and quantitation (Molecular Dynamics, Sunnyvale, CA). All of the above reagents can be purchased from Sigma. (Details of this assay can be found in: Schmidt, *et al.*, 2000, *J Biol Chem*, 275:6227-6233 which is incorporated herein by reference in
10 its entirety including any figures, tables, or drawings). Variation of this assay to apply to substrate not bound to membrane is straightforward.

A comprehensive discussion of various protease assays can be found in: The Handbook of Proteolytic Enzymes by Alan J. Barrett (Editor), Neil D. Rawlings (Editor), J. Fred Woessner (Editor) (February 1998) Academic Press, San Diego;
15 ISBN: 0-12-079370-9 (which is incorporated herein by reference in its entirety including any figures, tables, or drawings).

Similar assays are performed on bacterially expressed GST-fusion constructs
20 of the proteases.

EXAMPLE 8a: Chromosomal Localization of Proteases

Materials And Methods

Several sources were used to find information about the chromosomal
25 localization of each of the genes described in this patent. First, the Celera Browser was used to map the genes. Alternatively, the accession number of a genomic contig (identified by BLAST against NRNA) was used to query the Entrez Genome Browser (<http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/MapViewHelp.html>), and the cytogenetic localization was read from the NCBI data. References for

WO 02/00860

PCT/US01/20171

association of the mapped sites with chromosomal amplifications found in human cancer can be found in: Knuutila, et al., Am J Pathol, 1998, 152:1107-1123. Information on mapped positions was also obtained by searching published literature (at NCBI, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) for documented

5 association of the mapped position with human disease.

Results

The chromosomal regions for mapped genes are listed in Table 2.

10 The following section describes various diseases that map to chromosomal locations established for proteases included in this patent application. The protease polynucleotides of the present invention can be used to identify individuals who have, or are at risk for developing, relevant diseases. As discussed elsewhere in this application, the polypeptides and polynucleotides of the present invention are useful in identifying compounds that modulate protease activity, and in turn ameliorate

15 various diseases.

SGPr397, SEQID_1, maps to human chromosomal position 8q12. Chromosomal aberrations in this region are associated with breast cancer: Rummukainen J, et al., *Cancer Genet Cytogenet.* 2001 Apr 1;126(1):1-7.

20 SGPr413, SEQ ID NO:2, maps to human chromosomal position 2q35. This region is highly implicated in osteoarthritis (Loughlin J, et al., Linkage analysis of chromosome 2q in osteoarthritis. *Rheumatology.* 2000 Apr;39(4): 377-81).

25 SGPr404, SEQ ID NO:3, maps to human chromosomal position 10q26. Genomic amplification of this region has been associated with the following cancers (Knuutila): Malignant fibrous histiocytoma of soft tissue.

SGPr536_1, SEQ ID NO:4, maps to human chromosomal position 1p35.

WO 02/00860

PCT/US01/20171

SGPr414, SEQ ID NO:5, maps to human chromosomal position 2p14.

5 SGPr430, SEQ ID NO:6, maps to human chromosomal position 2q37. This region is highly implicated in osteoarthritis (Loughlin J, et al. ,Linkage analysis of chromosome 2q in osteoarthritis. *Rheumatology*. 2000 Apr;39(4): 377-81).

10 SGPr496_1, SEQ ID NO:7, maps to human chromosomal position Xp11. 4. Genomic amplification of this region has been associated with the following cancers (Knuutila): small cell lung cancer and prostate cancer.

SGPr495, SEQ ID NO:8, maps to human chromosomal position 6q16.

15 SGPr407, SEQ ID NO:9, maps to human chromosomal position 2q37. This region is highly implicated in osteoarthritis (Loughlin J, et al. ,Linkage analysis of chromosome 2q in osteoarthritis. *Rheumatology*. 2000 Apr;39(4): 377-81).

SGPr453, SEQ ID NO:10, maps to human chromosomal position 12q23.

20 SGPr445, SEQ ID NO:11, maps to human chromosomal position 6q16.

SGPr401_1, SEQ ID NO:12, maps to human chromosomal position 4q11. Genomic amplification of this region has been associated with the following cancers (Knuutila): Follicular carcinoma.

25

SGPr408, SEQ ID NO:13, maps to human chromosomal position 11p15.

WO 02/00860

PCT/US01/20171

SGPr480, SEQ ID NO:14, maps to human chromosomal position 17q24. Genomic amplification of this region has been associated with the following cancers (Knuutila): Non-small cell lung cancer, and testicular cancer.

- 5 SGPr431, SEQ ID NO:15, maps to human chromosomal position 4q31. 3. Genomic amplification of this region has been associated with the following cancers (Knuutila): Osteosarcoma.

- SGPr429, SEQ ID NO:16, maps to human chromosomal position 1p36. 2. Genomic
10 amplification of this region has been associated with the following cancers (Knuutila): alveolar cancer. .

- SGPr503, SEQ ID NO:17, maps to human chromosomal position 12q24. 3.
Genomic amplification of this region has been associated with the following cancers
15 (Knuutila): Non-small cell lung cancer.

SGPr427, SEQ ID NO:18, maps to human chromosomal position 17p13.

- SGPr092, SEQ ID NO:19, maps to human chromosomal position 11p15.
20

SGPr359, SEQ ID NO:20, maps to human chromosomal position 11q22. Genomic amplification of this region has been associated with the following cancers (Knuutila): Uterine cervix cancer.

- 25 SGPr104_1, SEQ ID NO:21, maps to human chromosomal position 3q27. Genomic amplification of this region has been associated with the following cancers (Knuutila): Squamous cell carcinomas of the head and neck; Malignant fibrous histiocytoma of soft tissue.

WO 02/00860

PCT/US01/20171

SGPr303, SEQ ID NO:22, maps to human chromosomal position 17q11. 1.
Genomic amplification of this region has been associated with the following cancers
(Knuutila): Breast carcinoma and Hepatocellular carcinoma.

- 5 SGPr402_1, SEQ ID NO:23, maps to human chromosomal position 19q11.
Genomic amplification of this region has been associated with the following cancers
(Knuutila): Leiomyosarcoma.

- SGPr434, SEQ ID NO:24, maps to human chromosomal position 3p21. Genomic
10 amplification of this region has been associated with the following cancers
(Knuutila): Bladder carcinoma.

- SGPr446_1, SEQ ID NO:25, maps to human chromosomal position 3p21. Genomic
amplification of this region has been associated with the following cancers
15 (Knuutila): Bladder carcinoma.

SGPr447, SEQ ID NO:26, maps to human chromosomal position 16p13. 3.

- SGPr432_1, SEQ ID NO:27, has not been assigned a chromosomal location.
20

SGPr529, SEQ ID NO:28, maps to human chromosomal position 19q13. 4.
Genomic amplification of this region has been associated with the following cancers
(Knuutila): Breast carcinoma.

- 25 SGPr428_1, SEQ ID NO:29, maps to human chromosomal position 8p23.

SGPr425, SEQ ID NO:30, maps to human chromosomal position 6q14.

WO 02/00860

PCT/US01/20171

SGPr548, SEQ ID NO:31, maps to human chromosomal position 19q13. 4.

Genomic amplification of this region has been associated with the following cancers
(Knuutila): Breast carcinoma.

- 5 SGPr396, SEQ ID NO:32, maps to human chromosomal position 4q32. Genomic
amplification of this region has been associated with the following cancers
(Knuutila): Non-small cell lung cancer.

- SGPr426, SEQ ID NO:33, maps to human chromosomal position 4q13. Genomic
10 amplification of this region has been associated with the following cancers
(Knuutila): Non-small cell lung cancer.

- SGPr552, SEQ ID NO:34, maps to human chromosomal position 4q13. Genomic
amplification of this region has been associated with the following cancers
15 (Knuutila): Non-small cell lung cancer.

SGPr405, SEQ ID NO:35, maps to human chromosomal position 16p13. 3.

- SGPr485_1, SEQ ID NO:36, maps to human chromosomal position 8p23.
20

SGPr534, SEQ ID NO:37, maps to human chromosomal position 16q23. Genomic
amplification of this region has been associated with the following cancers
(Knuutila): Diffuse large cell lymphoma of stomach.

- 25 SGPr390, SEQ ID NO:38, maps to human chromosomal position 19q11. Genomic
amplification of this region has been associated with the following cancers
(Knuutila): Leiomyosarcoma.

WO 02/00860

PCT/US01/20171

SGPr521, SEQ ID NO:39, maps to human chromosomal position 19q13.4.

Genomic amplification of this region has been associated with the following cancers
(Knuutila): Breast carcinoma.

- 5 SGPr530_1, SEQ ID NO:40, maps to human chromosomal position 9q22. Genomic
amplification of this region has been associated with the following cancers
(Knuutila): Non-small cell lung cancer.

- SGPr520, SEQ ID NO:41, maps to human chromosomal position 2q37. This region
10 is highly implicated in osteoarthritis (Loughlin J, et al. ,Linkage analysis of
chromosome 2q in osteoarthritis. *Rheumatology*. 2000 Apr;39(4): 377-81).

- SGPr455, SEQ ID NO:42, maps to human chromosomal position 12p11.2.
Genomic amplification of this region has been associated with the following cancers
15 (Knuutila): ovarian germ cell tumor, testicular cancer and non-small cell lung
cancer.

- SGPr507_2, SEQ ID NO:43, maps to human chromosomal position 7q36. Genomic
amplification of this region has been associated with the following cancers
20 (Knuutila): Ovarian cancer.

SGPr559, SEQ ID NO:44, maps to human chromosomal position 21q22.

- SGPr567_1, SEQ ID NO:45, maps to human chromosomal position 11q23.
25 Genomic amplification of this region has been associated with the following cancers
(Knuutila): Pleural mesothelioma.

SGPr479_1, SEQ ID NO:46, maps to human chromosomal position 1q42.

WO 02/00860

PCT/US01/20171

SGPr489_1, SEQ ID NO:47, maps to human chromosomal position 11p15.

SGPr465_1, SEQ ID NO:48, has not been assigned a chromosomal location.

5 SGPr524_1, SEQ ID NO:49, has not been assigned a chromosomal location.

SGPr422, SEQ ID NO:50, maps to human chromosomal position 4q13. Genomic amplification of this region has been associated with the following cancers (Knuutila): Non-small cell lung cancer.

10

SGPr538, SEQ ID NO:51, maps to human chromosomal position 11q23. Genomic amplification of this region has been associated with the following cancers (Knuutila): Pleural mesothelioma.

15 SGPr527_1, SEQ ID NO:52, has not been assigned a chromosomal location.

SGPr542, SEQ ID NO:53, maps to human chromosomal position 19q13. 1. Genomic amplification of this region has been associated with the following cancers (Knuutila): Small cell lung cancer (highly associated, with 10 of 35 patients tested showing amplification).

20

SGPr551, SEQ ID NO:54, maps to human chromosomal position 22q13. Genomic amplification of this region has been associated with the following cancers (Knuutila): Osteosarcoma.

25

SGPr451, SEQ ID NO:55, maps to human chromosomal position 12q23.

SGPr452_1, SEQ ID NO:56, maps to human chromosomal position 16p13. 3.

WO 02/00860

PCT/US01/20171

SGPr504, SEQ ID NO:57, has not been assigned a chromosomal location.

SGPr469, SEQ ID NO:58, has not been assigned a chromosomal location.

- 5 SGPr400, SEQ ID NO:59, maps to human chromosomal position 4q32. Genomic amplification of this region has been associated with the following cancers (Knuutila): Non-small cell lung cancer.

EXAMPLE 8b: Candidate Single Nucleotide Polymorphisms (SNPs)

10 Materials And Methods

The most common variations in human DNA are single nucleotide polymorphisms (SNPs), which occur approximately once every 100 to 300 bases. Because SNPs are expected to facilitate large-scale association genetics studies, there has recently been great interest in SNP discovery and detection. Candidate
15 SNPs for the genes in this patent were identified by blastn searching the nucleic acid sequences against the public database of sequences containing documented SNPs (dbSNP, at NCBI, <http://www.ncbi.nlm.nih.gov/SNP/snpblastpretty.html>). dbSNP accession numbers for the SNP-containing sequences are given. SNPs were also identified by comparing several databases of expressed genes (dbEST, NRNA) and
20 genomic sequence (i.e., NRNA) for single basepair mismatches. The results are shown in Table 1, in the column labeled "SNPs". These are candidate SNPs – their actual frequency in the human population was not determined. The code below is standard for representing DNA sequence:

G = Guanosine
25 A = Adenosine
T = Thymidine
C = Cytidine
R = G or A, puRine
Y = C or T, pYrimidine

WO 02/00860

PCT/US01/20171

K =G or T, Keto
 W =A or T, Weak (2 H-bonds)
 S =C or G, Strong (3 H-bonds)
 M =A or C, aMino
 5 B =C, G or T (i.e., not A)
 D =A, G or T (i.e., not C)
 H =A, C or T (i.e., not G)
 V =A, C or G (i.e., not T)
 N =A, C, G or T, aNy
 10 X =A, C, G or T

complementary G A T C R Y W S K M B V D H N X
 DNA + - + - + - + - + - + - + - + - + - + - + - + - + - + - + - +
 15 strands C T A G Y R S W M K V B H D N X

For example, if two versions of a gene exist, one with a "C" at a given position, and a second one with a "T" at the same position, then that position is represented as a Y, which means C or T. SNPs may be important in identifying heritable traits associated with a gene.

20

Results

The results of SNP identification are contained in Table 2 above, and in Example 1, under the section entitled DESCRIPTION OF NOVEL PROTEASE POLYNUCLEOTIDES. As discussed above, a variety of SNPs were identified in the protease polynucleotides of the present invention.

25

EXAMPLE 9: Demonstration Of Gene Amplification By Southern Blotting

Materials and Methods

WO 02/00860

PCT/US01/20171

Nylon membranes are purchased from Boehringer Mannheim. Denaturing solution contains 0.4 M NaOH and 0.6 M NaCl. Neutralization solution contains 0.5 M Tris-HCL, pH 7.5 and 1.5 M NaCl. Hybridization solution contains 50% formamide, 6X SSPE, 2.5X Denhardt's solution, 0.2 mg/mL denatured salmon DNA, 0.1 mg/mL yeast tRNA, and 0.2 % sodium dodecyl sulfate. Restriction enzymes are purchased from Boehringer Mannheim. Radiolabeled probes are prepared using the Prime-it II kit by Stratagene. The β -actin DNA fragment used for a probe template is purchased from Clontech.

Genomic DNA is isolated from a variety of tumor cell lines (such as MCF-7, MDA-MB-231, Calu-6, A549, HCT-15, HT-29, Colo 205, LS-180, DLD-1, HCT-116, PC3, CAPAN-2, MIA-PaCa-2, PANC-1, AsPc-1, BxPC-3, OVCAR-3, SKOV3, SW 626 and PA-1, and from two normal cell lines.

A 10 μ g aliquot of each genomic DNA sample is digested with EcoR I restriction enzyme and a separate 10 μ g sample is digested with Hind III restriction enzyme. The restriction-digested DNA samples are loaded onto a 0.7% agarose gel and, following electrophoretic separation, the DNA is capillary-transferred to a nylon membrane by standard methods (Sambrook, J. *et al.* (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory).

20 EXAMPLE 10: Detection Of Protein-Protein Interaction Through Phage Display Materials And Methods

Phage display provides a method for isolating molecular interactions based on affinity for a desired bait. cDNA fragments cloned as fusions to phage coat proteins are displayed on the surface of the phage. Phage(s) interacting with a bait are enriched by affinity purification and the insert DNA from individual clones is analyzed.

T7 Phage Display Libraries

WO 02/00860

PCT/US01/20171

All libraries were constructed in the T7Select1-1b vector (Novagen) according to the manufacturer's directions.

Bait Presentation

- Protein domains to be used as baits are generated as C-terminal fusions to
5 GST and expressed in *E. coli*. Peptides are chemically synthesized and biotinylated at the N-terminus using a long chain spacer biotin reagent.

Selection

- Aliquots of refreshed libraries (10^{10} - 10^{12} pfu) supplemented with PanMix and a cocktail of *E. coli* inhibitors (Sigma P-8465) are incubated for 1-2 hrs at room
10 temperature with the immobilized baits. Unbound phage is extensively washed (at least 4 times) with wash buffer.

After 3-4 rounds of selection, bound phage is eluted in 100 μ L of 1% SDS and plated on agarose plates to obtain single plaques.

Identification of insert DNAs

- Individual plaques are picked into 25 μ L of 10 mM EDTA and the phage is
15 disrupted by heating at 70 °C for 10 min. 2 μ L of the disrupted phage are added to 50 μ L PCR reaction mix. The insert DNA is amplified by 35 rounds of thermal cycling (94 °C, 50 sec; 50 °C, 1min; 72 °C, 1min).

Composition of Buffer

- 20 10x PanMix
5% Triton X-100
10% non-fat dry milk (Carnation)
10 mM EGTA
250 mM NaF
25 250 μ g/mL Heparin (sigma)
250 μ g/mL sheared, boiled salmon sperm DNA (sigma)
0.05% Na azide
Prepared in PBS

WO 02/00860

PCT/US01/20171

Wash Buffer

PBS supplemented with:

0.5% NP-40

25 µl g/mL heparin

5 PCR reaction mix

1.0 mL 10x PCR buffer (Perkin-Elmer, with 15 mM Mg)

0.2 mL each dNTPs (10 mM stock)

0.1 mL T7UP primer (15 pmol/µL) GGAGCTGTCGTATTCCAGTC

0.1 mL T7DN primer (15 pmol/µL)

10 AACCCCTCAAGACCCGTTTAG

0.2 mL 25 mM MgCl₂ or MgSO₄ to compensate for EDTA

Q.S. to 10 mL with distilled water

Add 1 unit of Taq polymerase per 50 µL reaction

LIBRARY: T7 Select1-H441

15

Example 11: Gene Expression based on Incyte and Public ESTsMaterials and Methods

The nucleic acid sequences for the proteases were used as queries in a BLASTN search of the Incyte and public dbEST databases of expressed sequences.

- 20 The tissue sources of the libraries in which the protease was represented are listed below, along with the frequency the gene occurred in specific tissues. The frequency is determined by the number of clones representing the gene within a given tissue source. The Incyte gene identification number or public NCBI accession number is given, followed by the tissue source. A brief summary of the
- 25 tissue specificity is then given for each gene.

Results

WO 02/00860

PCT/US01/20171

SGPr397, SEQID:1,

Incyte 366783.1 Clones: 2 prostate, 3 colon, retina and small intestine

366783.3 1 prostate clone

Selective expression in prostate (3/8 clones) and colon (3/8 clones)

5

SGPr413, SEQID:2,

Incyte 475365.6 5 clones, 3 in small intestine, plus prostate, breast tumor

475365.5 10 clones: 8 in small intestine, plus brain (2)

Highly selective expression in small intestine (11/15 clones)

10

SGPr404, SEQID:3,

Incyte 1129157.1 213 clones, highest in brain (18), m/f genitalia (21/23), breast (14), and digestive (25)

1129157.2 1: mixed

15 Broad expression, some elevation in brain (18/214 clones), digestive tissues (25/118) and male/female genitalia (21, 23 clones) and breast (14 clones)

SGPr536_1, SEQID:4,

Incyte 233762.17 149 clones: no tissue >21 hits

20 233762.15 15 clones, mixed

Broad expression seen in 164 clones

SGPr414, SEQID:5,

Incyte 399773.5 669 clones, 404 libraries, broadly distributed

25 Expressed broadly and strongly (669 clones)

SGPr430, SEQID:6,

Incyte 407823.1 21 clones (4 testis, 3 brain, 3 prostate)

1136483.1 1 Prostate

WO 02/00860

PCT/US01/20171

- 411246.1 2 ea lung tumor, sm intestine, fetal liver, and 1 heart
322700.1 T cells
407823.2 Fetal liver/spleen
Mixed expression, highest in testis (4/31 clones), brain (3/31) and prostate (4/31)
5 and fetal liver/spleen (4/31)
- SGPr496_1, SEQID:7,
Incyte 986031.1 12 clones: 2 ea lung, brain, adrenal tumor
Selective expression in lung (2/12 clones), adrenal tumor (2/12) and brain (2/12)
10
- SGPr495, SEQID:8,
Incyte 350921.2 16 clones: 2 thymus, 3 colon, 3 brain
350921.7 Adrenal tumor
350921.10 14 clones: 2 colon, 1 adrenal tumor
15 350921.6 10 clones: 2 adrenal (1 tumor), 2 brain
350921.1 Adrenal
350921.9 Colon (2) Sm intestine, lung tumor
Selectively expressed in adrenal gland (5/46 clones) and colon (7/46)
- 20 SGPr407, SEQID:9,
No ESTs
- SGPr453, SEQID:10,
Incyte 428428.1 17 clones: 3 lung (2 tumors), 2 prostate, 4 testis, 2 teratoma
25 (hNT2)
428428.5 brain, lung, teratoma
428428.6 teratoma (2), lung, kidney
Highly expressed in hNT2 teratoma cell line (5/24 clones), and selective for lung (5 clones) and testis (4 clones)

WO 02/00860

PCT/US01/20171

- SGPr445, SEQID:11,
 Incyte 350921.7 1 adrenal tumor
 350921.10 14 incl 1 adrenal tumor
 5 350921.6 10 clones: normal and tumor adrenal (1 ea) colon tumor (2)
 350921.1 Adrenal gland
 350921.8 2 prostate tumor, 1 retina
 Highest in adrenal gland (5/28 clones), indicates a possible involvement in adrenal hormone processing
- 10 SGP401_1, SEQID:12,
 Incyte 232414.1 169 clones: 69 NS, 18 male genitalia, (10 prostate), 11 female genitalia, 11 respiratory system, 5 kidney, 9 in one glioblastoma library.
 232414.2 testis
- 15 Selective for nervous system (69/170 clones), especially glioblastoma
- SGPr408, SEQID:13,
 Incyte 233660.2 357 clones, 248 libraries: 54 brain, 26 hemic/immune 24/21 f/m genitalia, 21 digestive, 20 cardiovascular
 20 233660.11 13 clones, broad expression.
 233660.10 7 clones, broad expression
 Expressed broadly and strongly (377 clones)
- SGPr480, SEQID:14,
 25 Incyte 1326256.3 274 clones, broad but highest in NS (59), hemic (35) genitalia (24/10, m/f). 6 clones in one pituitary gland library
 1326256.8 4 mixed clones
 1326256.1 26 clones, 13 in male genitalia
 1326256.10 10 clones mixed

WO 02/00860

PCT/US01/20171

Broad, strong expression (over 300 clones)

SGPr431, SEQID:15,

Incye 236368.1 151 clones, 110 libraries, highest in digestive, nervous, hemic

5 (18, 17, 18). 5, 4 hits each in two fetal liver/spleen libraries

236368.2 1 fetal heart

236368.14 7: mixed

Broad and moderately strong expression (159 clones total)

10 SGPr429, SEQID:16,

Incye 890540.9 41 clones, broad

890540.1 125 clones, broad

890540.8 15 clones, broad

Broad and moderately strong expression (181 clones)

15

SGPr503, SEQID:17,

Incye 1447357.3 107 clones highest in NS (15), male genitalia (11) and digestive tissue (15)

1447357.1 Dendritic cells

20 245045.1 16 mixed

Broad expression (124 clones), highest in nervous system (16), male genitalia (11) and digestive tissue (15)

SGPr427, SEQID:18,

25 Incye 903092.31 41 clones, 35 libraries; 9 clones in prostate, otherwise very broad

903092.23 1 brain

Expression elevated in prostate (9/42 clones)

WO 02/00860

PCT/US01/20171

- SGPr092, SEQID:19,
 Incyte 339251.1 4/5 uterus, 1 mixed tissue
 339251.2 1/1 uterus; Highly selective expression in uterus (5/6 clones)
- 5 SGPr359, SEQID:20,
 Incyte 391133.1 mixed tissue (fetal lung, testis, B-cell)
 gi|7280399 = same as Incyte
 Mixed tissues, one EST only
- 10 SGPr104_1, SEQID:21,
 Incyte 12/23 clones in brain
 232015.5 6/7 clones in brain
 232015.2 2/4 clones in brain
 232015.1 1/1 brain
- 15 232015.6 1/1 brain
 Brain specific
- SGPr303, SEQID:22,
 Incyte 323846.15 38 samples, highest in brain(8) breast(4), uterus and ovary(7)
 20 323846.1 304 clones, high in nervous sys(72) and genitalia (28 f, 24 m),
 other tissue
 414048.34 45 clones, highest in NS
 323846.11 8 clones
 Broad expression
- 25 SGPr402_1, SEQID:23,
 Incyte 244407.4 25 clones, highest in testis (8) brain (4), uterus (2; 1 tumor)
 244407.2 uterus
 244407.1 testis

WO 02/00860

PCT/US01/20171

244407.6 uterus tumor
244407.5 fallopian tube tumor
244407.9 mixed tissue incl tumor, nasal tumor
Enriched in genital samples.

5

SGPr434, SEQID:24,

Incyte 110154.4 no clone origin

110154.6 2 prostate

110154.12 1 prostate, 1 pituitary

10 110154.11 1 pituitary

110154.8 fallopian tube tumor (2), mixed (1)

110154.7 Pituitary

110154.5 Thigh muscle (2) – tissue-specific splicing

110154.10 3 heart, 2 brain, pituitary (61/62 match)

15 Selective expression in prostate (3/17 clones), 4 pituitary gland (4/17 clones) and
fallopian tube tumor (2/17 clones). May indicate a role in hormone processing in
pituitary and prostate.- hormone processing.

SGPr446_1, SEQID:25,

20 Incyte 1040641.1 Heart, Muscle

1388371.1 2 Heart

Specific for muscle (3/3 clones) especially heart muscle (2/3 clones)

SGPr447, SEQID:26,

25 Incyte 1352932.1 pancreas tumor

Single clone from pancreas tumor

SGPr432_1, SEQID:27,

Incyte 474674.15 29 clones, mixed

WO 02/00860

PCT/US01/20171

- 474674.30 90 clones, mixed
 474674.1 82 clones, mixed
 Broad and strong expression (201 clones total)
- 5 SGPr529, SEQID:28,
 Incyte 988019.3 71 clones, 23 in f genitalia. 9 from 1 ovary tumor library, 2
 from another, 2 from another, and one from yet another (no normal ovaries). 5 from
 one pancreatic tumor line, 4 from pancreas tumor library
 988019.1 breast skin
- 10 Selective expression in pancreas (4/72 clones from one pancreatic tumor library and
 5 from a pancreatic tumor line) and ovary (14 from ovary tumors, none from normal
 ovary).
- SGPr428_1, SEQID:29,
 15 Incyte 891146.1 4: brain, pituitary, blood, thymus
 Broad, low-level expression (4 clones all from different tissues)
- SGPr425, SEQID:30,
 Incyte 400833.1 25 clones, mixed (<4 from any tissue, except 5 from 'fetus')
- 20 Expressed broadly but not strongly (25 clones total)
- SGPr548, SEQID:31,
 Incyte 971236.1 2 clones from mixed testis, fetal lung, B cells
 Rare transcript, just two clones from a mixed library of testis, fetal lung and B cells
- 25 SGPr396, SEQID:32,
 Incyte 209051.1 Lung (1)
 889126.1 Brain (1)
 Only 2 ESTs – lung and brain

WO 02/00860

PCT/US01/20171

SGPr426, SEQID:33,
Incyte Nô ESTs

- 5 SGPr552, SEQID:34,
Incyte 1510512.1 tonsil, spinal cord
1511222.1 tonsil
406221.1 83 clones: 16 in NS, 10 in hemic/immune, 9 in male genitalia,
and several other tissues. 1 tonsil,
10 981355.3 8 clones, 2 ovary tumor, 1 tonsil, varied
Of 94 clones, see some selectivity in tonsil (3/94, but tonsil not usually seen as an
expression source), and nervous system (17 clones)

- SGPr405, SEQID:35,
15 Incyte 134360.1 1 kidney
One clone, from kidney

- SGPr485_1, SEQID:36,
Incyte 180576.2 5/5 clones in testis
20 Testis specific (5/5 clones)

- SGPr534, SEQID:37,
Incyte 1383391.20 112/114, matches well at start (103-165 = perfect match) but
maybe template artefact
25 1450812.1 1/1 pancreas, few mismatches are N's
1383391.13 5/5 pancreas
1045834.1 1/1 pancreas
Almost completely pancreas-specific (118/120 clones from pancreas)

WO 02/00860

PCT/US01/20171

SGPr390, SEQID:38,

Incyte 199428.9 Bone tumor, small intestine

199428.3 382 clones: 41 brain, 34/23 genitalia (m/f), 22 hemic/immune,
27 digestive

- 5 Broad tissue distribution, highest in brain (41/382 clones), male and female genitalia
(34 and 23 clones, respectively) and digestive system (27 clones)

SGPr521, SEQID:39,

Incyte 427826.1 28 clones, most in sm intestine tumor (5, 1 library), neonatal

- 10 keratinocytes (3 ea from 2 libraries), 8 ovary tumors, 5 breast skin

Selective expression in ovarian tumors (8/28 clones), neonatal keratinocytes (6/28),
breast keratinocytes (5) and in a small intestine tumor library (5 clones from one
library)

- 15 SGPr530_1, SEQID:40,
No ESTs

SGPr520, SEQID:41,

Incyte 405947.1 4/4 clones adrenal tumor (pheochromocytoma) (3 from one

- 20 library, 1 from another)

1338652.1 1/1 clones from adrenal tumor (pheochromocytoma)

1477189.1 1/1 clones from adrenal (mixed normal and
pheochromocytoma)

- 25 Specific to pheochromocytoma (adrenal gland tumor): 4/5 clones from
pheochromocytoma and 1/5 from mixed normal adrenal gland and
pheochromocytoma.

SGPr455, SEQID:42,

Incyte 1115833.1 mixed fetal lung/testis/Bcell

WO 02/00860

PCT/US01/20171

987279.1 Brain (1), mixed tissues incl tumor (1)
Three clones, only one (brain) with a specific source

SGPr507_2, SEQID:43,

- 5 Incyte 403891.1 10 clones: 6 in testis and 6 in mixed (testis, lung, Bcell)
403891.2 1 brain
Testis-selective: 6/11 clones from testis and 5/11 from mixed libraries including
testis samples

10 SGPr559, SEQID:44,

Incyte 475100.1 35 clones, 11 in f genitalia, 8 in digestive: 7 uterus tumors
(none normal), 4 in breast, 2 ovary tumors, 1 HeLa cervical tumor
475100.6 Th1 cells, HeLa cells
Selective expression in tumors of the uterus (7/37 clones), ovary (2/37) cervix (2/37
15 from HeLa cervical tumor cell line), as well as breast (4)

SGPr567_1, SEQID:45,

- Incyte 981355.3 Mixed (2/8 clones from ovary tumor library, 1 ea from tonsil,
brain, lung tumor, heart, placenta, dorsal root ganglion)
20 Rare broad expression (8 clones from 7 different tissues).

SGPr479_1, SEQID:46,

- Incyte 219214.1 1 testis
Single EST, expressed in testis

25

SGPr489_1, SEQID:47,

Incyte 338956.1 2 kidney, 1 placenta
1042306.1 1 mouth tumor, 2 fallopian tube tumor
1384824.1 Sm intestine, kidney

WO 02/00860

PCT/US01/20171

Rare but broad expression, selective to kidney (3/8 clones) and fallopian tube tumor
(2/8 from one library)

SGPr465_1, SEQID:48,

5 No ESTs

SGPr524_1, SEQID:49,

Incye 952182.3 1 testis

952182.2 1 testis

10 952182.4 1 prostate

Specific to male genitalia (2/3 clones in testis, 1/3 in prostate)

SGPr422, SEQID:50,

Incye 1511284.1 1 tonsil

15 1351259.1 1 brain

Rare transcript seen only in tonsil (1/2 clones) and brain (1/2 clones)

SGPr538, SEQID:51,

Incye 903092.29 4 clones, 3 brain, 1 breast

20 903092.19 1 brain

903092.22 2 brain

903092.28 38 clones, 24 in brain

903092.24 2 brain, 1 sm intestine

Selective expression in nervous system (32/48 clones)

25

SGPr527_1, SEQID:52,

Incye 65450.1 1 prostate tumor

103554.2 mixed tissues incl tumor

228456.2 11 clones: mixed (2 brain, 2 blood)

WO 02/00860

PCT/US01/20171

103554.1 Mixed (3)
Broad low-level expression (16 clones)

SGPr542, SEQID:53,

- 5 Incyte 244085.1 Expression is selective to hemopoetic cells: All 11 clones are from hemopoetic tissues: 6 from fetal liver/spleen, 4 of which are mast cells, 2 from umbilical cord blood, 1 from CD34+ bone marrow, and two clones from leukemias: 1 from AML blast cells and one from CML

- 10 SGPr551, SEQID:54,

Incyte 319529.1 22 clones: 7 liver, 1 fetal liver/spleen, 3 lung

319529.2 1 liver

319529.3 2 mixed tissues incl testis, 1 testis (tissue-specific splice)

Selectively expressed in liver (9/26 clones), may have a testis-specific splice form

- 15 (3/3 clones of one template)

SGPr451, SEQID:55,

Incyte 1471541.1 1 mixed)

- 20 SGPr452_1, SEQID:56,

Incyte 446374.1 Mixed (melanocytes, uterus, fetal heart)

No expression data (single EST from mixed tissues)

SGPr504, SEQID:57,

- 25 Incyte 244085.1 Expression is selective to hemopoetic cells: All 11 clones are from hemopoetic tissues: 6 from fetal liver/spleen, 4 of which are mast cells, 2 from umbilical cord blood, 1 from CD34+ bone marrow, and two clones from leukemias: 1 from AML blast cells and one from CML

WO 02/00860

PCT/US01/20171

SGPr469, SEQID:58,
Incyte 110154.10 7 clones: 3 heart, 2 brain, 1 pituitary
110154.3 heart, muscle, testis
Selective expression in heart (4/10 clones)

5

SGPr400, SEQID:59,
Incyte 889126.1 Brain
Only one EST, in brain

10

CONCLUSION

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, treatments, molecules, specific compounds described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of" and "consisting of" may be replaced

WO 02/00860

PCT/US01/20171

with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.

In view of the degeneracy of the genetic code, other combinations of nucleic acids also encode the claimed peptides and proteins of the invention. For example, all four nucleic acid sequences GCT, GCC, GCA, and GCG encode the amino acid alanine. Therefore, if for an amino acid there exists an average of three codons, a polypeptide of 100 amino acids in length will, on average, be encoded by 3100, or 5 x 1047, nucleic acid sequences. Thus, a nucleic acid sequence can be modified to form a second nucleic acid sequence, encoding the same polypeptide as encoded by the first nucleic acid sequences, using routine procedures and without undue experimentation. Thus, all possible nucleic acids that encode the claimed peptides and proteins are also fully described herein, as if all were written out in full taking into account the codon usage, especially that preferred in humans. Furthermore, changes in the amino acid sequences of polypeptides, or in the corresponding nucleic acid sequence encoding such polypeptide, may be designed or selected to

WO 02/00860

PCT/US01/20171

take place in an area of the sequence where the significant activity of the polypeptide remains unchanged. For example, an amino acid change may take place within a β -turn, away from the active site of the polypeptide. Also changes such as deletions (*e.g.* removal of a segment of the polypeptide, or in the corresponding
5 nucleic acid sequence encoding such polypeptide, which does not affect the active site) and additions (*e.g.* addition of more amino acids to the polypeptide sequence without affecting the function of the active site, such as the formation of GST-fusion proteins, or additions in the corresponding nucleic acid sequence encoding such polypeptide without affecting the function of the active site) are also within the
10 scope of the present invention. Such changes to the polypeptides can be performed by those with ordinary skill in the art using routine procedures and without undue experimentation. Thus, all possible nucleic and/or amino acid sequences that can readily be determined not to affect a significant activity of the peptide or protein of the invention are also fully described herein.

15 The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited
20 herein.

Other embodiments are within the following claims.

WO 02/00860

PCT/US01/20171

What is claimed is:

CLAIMS

1. An isolated, enriched or purified nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence that:
 - 5 (a) encodes a polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 and biological domains thereof;
 - (b) is the complement of the nucleotide sequence of (a); or
 - (c) hybridizes under stringent conditions to the nucleotide molecule of (a) and encodes a protease polypeptide.
2. The nucleic acid molecule of claim 1, further comprising a vector or promoter operatively linked to the nucleotide sequence.

WO 02/00860

PCT/US01/20171

3. The nucleic acid molecule of claim 1, wherein said nucleic acid molecule is isolated, enriched, or purified from a mammal.
- 5 4. The nucleic acid molecule of claim 3, wherein said mammal is a human.
- 10 5. The nucleic acid molecule of claim 1 comprising a nucleic acid comprising a nucleotide sequence which hybridizes under stringent conditions to a nucleotide sequence encoding a protease polypeptide having an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.
- 15
- 20
- 25

WO 02/00860

PCT/US01/20171

6. An isolated, enriched, or purified protease polypeptide, wherein said polypeptide comprises an amino acid sequence at least about 90% identical to a sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118 and biological domains thereof.
7. The protease polypeptide of claim 6, wherein said polypeptide is isolated, purified, or enriched from a mammal.
8. The protease polypeptide of claim 7, wherein said mammal is a human.
9. An antibody or antibody fragment having specific binding affinity to a protease polypeptide or to a domain of said polypeptide, wherein said polypeptide comprises an amino acid sequence selected from the

WO 02/00860

PCT/US01/20171

- group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.
10. A hybridoma which produces the antibody of Claim 9.
11. A kit comprising an antibody which binds to a polypeptide of claim 6 and a negative control antibody.
12. A method for identifying a substance that modulates the activity of a protease polypeptide comprising the steps of:
- (a) contacting a protease polypeptide substantially identical to an amino acid sequence selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID

WO 02/00860

PCT/US01/20171

- NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72,
 SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID
 NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79,
 SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID
 NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86,
 SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID
 NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93,
 SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID
 NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100,
 SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID
 NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107,
 SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID
 NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114,
 SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ
 ID NO:118 with a test substance;
- (b) measuring the activity of said polypeptide; and
 (c) determining whether said substance modulates the activity of
 said polypeptide.
13. A method for identifying a substance that modulates the activity of a
 protease polypeptide in a cell comprising the steps of:
- (a) expressing a protease polypeptide in a cell, wherein said
 polypeptide comprises a sequence substantially identical to an
 amino acid sequence selected from the group consisting of
 those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID
 NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65,
 SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID
 NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72,
 SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID

WO 02/00860

PCT/US01/20171

- 5 NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79,
SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID
NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86,
SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID
10 NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93,
SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID
NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100,
SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID
NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107,
15 SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID
NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114,
SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ
ID NO:118;
- (b) adding a test substance to said cell; and
- 15 (c) monitoring a change in cell phenotype, cell proliferation, cell
differentiation or the interaction between said polypeptide and
a natural binding partner.
14. A method for treating a disease or disorder by administering to a
20 patient in need of such treatment a substance that modulates the
activity of a protease substantially identical to an amino acid
sequence selected from the group consisting of those set forth in SEQ
ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID
NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID
25 NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID
NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID
NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID
NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID
NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID

WO 02/00860

PCT/US01/20171

- NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.
- 5
- 10
- 15
15. The method of claim 14, wherein said disease or disorder is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, metabolic disorders and inflammatory disorders.
- 15
16. The method of claim 15, wherein said disease or disorder is selected from the group consisting of cancers of tissues; cancers of blood or hematopoietic origin; cancers of the breast, colon, lung, prostate, cervical, brain, ovarian, bladder or kidney.
- 20
17. The method of claim 15, wherein said disease or disorder is selected from the group consisting of central or peripheral nervous system diseases, migraines; pain; sexual dysfunction; mood disorders; attention disorders; cognition disorders; hypotension; hypertension; psychotic disorders; neurological disorders and dyskinesias.
- 25
18. The method of claim 15, wherein said substance modulates protease activity *in vitro*.

WO 02/00860

PCT/US01/20171

19. The method of claim 18, wherein said substance is a protease inhibitor.
20. A method for detection of a protease polypeptide in a sample as a
5 diagnostic tool for a disease or disorder, wherein said method comprises:
(a) contacting said sample with a nucleic acid probe which hybridizes under hybridization assay conditions to a nucleic acid target region of a protease polypeptide having an amino acid sequence
10 selected from the group consisting of those set forth in SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and
25 SEQ ID NO:118, said probe comprising the nucleic acid sequence encoding the polypeptide, fragments thereof, and the complements of the sequences and fragments; and

WO 02/00860

PCT/US01/20171

(b) detecting the presence or amount of the probe:target region hybrid as an indication of the disease.

21. The method of claim 20, wherein said disease or disorder is selected from the group consisting of cancers, immune-related diseases and disorders, cardiovascular disease, brain or neuronal-associated diseases, metabolic disorders and inflammatory disorders.
22. The method of claim 21, wherein said disease or disorder is selected from the group consisting of cancers of tissues; cancers of hematopoietic cancers of blood or hematopoietic origin; cancers of the breast, colon, lung, prostate, cervical, brain, ovarian, bladder or kidney.
23. The method of claim 21, wherein said disease or disorder is selected from the group consisting of central or peripheral nervous systems disease, migraines, pain; sexual dysfunction; mood disorders; attention disorders; cognition disorders; hypotension; hypertension; psychotic disorders; neurological disorders; and dyskinesias.
24. The isolated, enriched or purified nucleic acid molecule of claim 1 comprising a nucleic molecule encoding a biological domain of a protease polypeptide having a sequence selected from the group consisting of SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID

WO 02/00860

PCT/US01/20171

NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86, SEQ ID
 NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90, SEQ ID
 NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94, SEQ ID
 NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98, SEQ ID
 NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102, SEQ ID
 NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID NO:106, SEQ
 ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ ID NO:110,
 SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113, SEQ ID
 NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and
 SEQ ID NO:118.

25. The nucleic acid molecule of claim 1 comprising a nucleic acid
 sequence encoding a protease polypeptide having an amino acid
 sequence that has least 90 % identity to a polypeptide selected from
 the group consisting of those set forth in SEQ ID NO:60, SEQ ID
 NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID
 NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID
 NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID
 NO:73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID
 NO:77, SEQ ID NO:78, SEQ ID NO:79, SEQ ID NO:80, SEQ ID
 NO:81, SEQ ID NO:82, SEQ ID NO:83, SEQ ID NO:84, SEQ ID
 NO:85, SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, SEQ ID
 NO:89, SEQ ID NO:90, SEQ ID NO:91, SEQ ID NO:92, SEQ ID
 NO:93, SEQ ID NO:94, SEQ ID NO:95, SEQ ID NO:96, SEQ ID
 NO:97, SEQ ID NO:98, SEQ ID NO:99, SEQ ID NO:100, SEQ ID
 NO:101, SEQ ID NO:102, SEQ ID NO:103, SEQ ID NO:104, SEQ
 ID NO:105, SEQ ID NO:106, SEQ ID NO:107, SEQ ID NO:108,
 SEQ ID NO:109, SEQ ID NO:110, SEQ ID NO:111, SEQ ID

WO 02/00860

PCT/US01/20171

NO:112, SEQ ID NO:113, SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID NO:117 and SEQ ID NO:118.

26. The nucleic acid molecule of Claim 1 wherein the molecule
5 comprises a nucleotide sequence substantially identical to a sequence
selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2,
SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID
NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID
NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID
10 NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID
NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID
NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID
NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID
NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID
15 NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID
NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID
NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID
NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID
NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID
20 NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, and SEQ
ID NO:59.
27. An isolated, enriched or purified nucleic acid molecule consisting
essentially of about 10-30 contiguous nucleotide bases of a nucleic
25 acid sequence that encodes a polypeptide that is selected from the
group consisting of SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62,
SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66,
SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70,
SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:74,

WO 02/00860

PCT/US01/20171

- SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:77, SEQ ID NO:78,
 SEQ ID NO:79, SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82,
 SEQ ID NO:83, SEQ ID NO:84, SEQ ID NO:85, SEQ ID NO:86,
 SEQ ID NO:87, SEQ ID NO:88, SEQ ID NO:89, SEQ ID NO:90,
 5 SEQ ID NO:91, SEQ ID NO:92, SEQ ID NO:93, SEQ ID NO:94,
 SEQ ID NO:95, SEQ ID NO:96, SEQ ID NO:97, SEQ ID NO:98,
 SEQ ID NO:99, SEQ ID NO:100, SEQ ID NO:101, SEQ ID NO:102,
 SEQ ID NO:103, SEQ ID NO:104, SEQ ID NO:105, SEQ ID
 NO:106, SEQ ID NO:107, SEQ ID NO:108, SEQ ID NO:109, SEQ
 10 ID NO:110, SEQ ID NO:111, SEQ ID NO:112, SEQ ID NO:113,
 SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:116, SEQ ID
 NO:117 and SEQ ID NO:118.
28. The isolated, enriched or purified nucleic acid molecule of Claim 27
 15 consisting essentially of about 10-30 contiguous nucleotide bases of a
 nucleic acid sequence selected from the group consisting of SEQ ID
 NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5,
 SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID
 NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID
 20 NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID
 NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID
 NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID
 NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID
 NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID
 25 NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID
 NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID
 NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID
 NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID
 NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID

WO 02/00860

PCT/US01/20171

NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID
NO:58, and SEQ ID NO:59.

- 5
29. A recombinant cell comprising the nucleic acid molecule of claim 1.
30. A method for detecting the presence or amount of protease polypeptide in a sample comprising
- 10 (a) contacting the sample with the antibody of Claim 9 under conditions suitable for protease-antibody immunocomplex formation; and
- (b) detecting the presence or amount of the antibody conjugated to the protease polypeptide.

WO 02/00860

1/68

PCT/US01/20171

FIG. 1A

>SGPR397_SEQID_1
ATGAAGTGTCTCGGGGAGCGCAGGGGCCAGGCAGCTGCTTTCCTGCTCTTTTCTGCTGCTCTTTTGAAGATTCT
GCAACCGGGGACAGCCACCTTTATAACACCGCTATGCTGTGTAAAGTGAAGATTATTTCCCAAAACAGA
AGAGGAAGCATATGCACTGAAGAAATATCCTATCAACTTAAGTGTGACCTGTGGCAGCCAGCAGTATCTCCTA
TGATCAGAGGGAACAGTTACTGTCCATATCCCCCAAAATGGTTCCCGAGCCCTGTAGCCCTTTCTACAGGA
AGCCAACTCCAGTACCAAGGTCTCTAGAGATCTTCAGAAACACTTGGAGAGGGGAGCAGCTTGCACACCC
AGAGAAACCGAAGATCCCTCTCTGGATATAATTAAGTTTATCACTCTTAGAAGAAATTCAAAATTTGGATGCA
TCATCTGAATAAACTCACTCAGGCCCTCATCCACATGTTCTCTATTGGAAGATCATATGAGGGAAGATGCTTTTT
ATTITAAAGCTGGGAGAGCATCAGCACTCAAAGAGCTGTTTGGATAGACTGTGGTATTATGCAAGAGATGG
ATTGGTCTGCTGCTTTTGTGAGTGGTTTGTAAAGAGAGCTCTTCAACATATAAGAGTGACCCAGCCATGAGAAAA
TGCTGAATCATCTATATTTCTATATCATGCTGTGTTTAACTGCTGAGTACCAATTTAGTTGGACCAATGATCGA
TTTTGGAGAAAAAAGGTCAAGGAACCAAGGTTTGGCTGCCGTGGAGTGCAATAGAACTGGAAAGT
GAAGTGTGTGTAAGTTTGGGACCACTGGATCCAGATCCAAAGGTTTCTGCAGGTTTACTCTGCAAAATAT
GAGTCCAGAGGACTCTCATGGGAGACTCATGTTTTTCTGTATGTGA

>SGPR413_SEQID_2
ATGAAGCTCTGCTTGAAACCCCTTATCTTTTGGGATGCTGTTTCTGGAGGGCTGGGATATGATAGATCCTTA
GCCAACACAGACAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAACATATAC
CACCCATGGGAGAGATCTATGAGTGGATGAGAGATCAAGTGAAGTACAAAGGAAGTGGTGACACAGCATTT
CCTAGGAGTGACCTATGAGACCCACCCCATGTTATCTGAAGATCAGCCACCATCTGGTAAATCCCAAGAAAT
CATTTGGATGGAGTTGGAATTCACGCCAGAGATGGAATGCTCTGCTTTTGGCAATGGTTCTGCTCAAGAAAT
TCTACAAACCATAAAGACACTCAAGTATACGCAAGCTCTTAGAACCTGGACTTATGCTCTTCCAGTTCTT
AACATAGATGGTTATATCTACACTGGACACTGATCGTCTTGGAGGAAATCCCGTTACCCCATATAATGGCA
CATGTTTTGGGACGGATCTCAATCGAAATTTCAATGCTTGGTGTAGTATTGGTCTCTAGAACTGCCAAGA
TCAACATTTCTGTGGACAGGGCCAGTGTCTGAACACAGAGACTAAAGCTGTTGCCAGCTTCTAGAGGCAAGA
AGGATGATATTTTGTCTTCTGACCATGCACTCTTATGGGCAATTAATTTCTCACACTTACGGCTACACCAAAA
TAAATCAAGTAACCAACCCAGAAATGATTCAGTTGGACAGAGCGCAATGCAATTGAAAGCAAGTATGGAAC
CAATTAGAGTTGGATCGAGTGCAGATATTTATGCTCTCAGGGTCTTCAAGAGATTGGGCCCGGAGACAT
TGGGATTTCCCTTCTCATATACGTTTGAGCTGAGGGACAGTGGAAACATATGGGTTTGTCTGCCAAGCTCAGAT
CCAGCCACCTGTGAGGAGACCATGGAGGCTGTGCTGTCAGTCTGATGTGATGCGAAACACTGGCACT
CGGACAGTGTCTGGAAGGGTGACATCTGCCACTATGCTGCTGGGCCCTGCTGGTCTGCTCTTCTCTCTAA

WO 02/00860

2/68

PCT/US01/20171

FIG. 1B

>SGPR404_SEQID_3
ATGGTGAAGCAATGACAGCCACACCGTGGGTCACTGTTAAGAATGGATGCTGGAGACATGATATTTGAGGGAACAG
TGAGAAAGGAGATCCCTGTTCTCAATGAGTACCGTCCCATGCTGGCCGCTGACATCCCATAAACCCCTCAGT
CCTGGTTTGAATATGGAGCATCTGCATGAGATGGAGATCCTGGGCTGCCACTGCCAGATCCCTAATAATTAAT
ATCACGCCGGAAGAGATGACACCATGATGACCTGGATTTTAAGCACCAATTTAAGAGAAATGCGCCAGT
TGATGAAGTTGTGAATGAATGTGTTCCCAATATACCAAGATTTACAACATTTGAAAAAGCCACAGGGCTGA
AGCTGTATGCTGTGGAGATCTCAGATCACCTGGGAGCATGAAGTCCGTTGAGCCCGAGTTCCACTACATCGCG
GGGGCCACGCAATGAGGTGCTGGCCGGAGCTGCTGCTGCTGTGTGTCAGTTCTGCTGTCAGGAGTAC
TTGGCCCGAATGCGCGCATGCTCACCTGGTGGAGAGACGGATTCAGTCTCCCTCCCTCAACCCCG
ATGGCTACGAGAAAGCCCTACGAAAGGGGCTCGGAGCTGGGAGGCTGCTGCTGGGAGCTGGACCCAGATG
GAAATTGACATCAACAACACTTTCTGATTTAAACACGCTGCTGGGAGCAGAGGATCGACAGATGTCCCA
GGAAGTTCCCAATCACTATATTGCAATCCCTGAGTTTCTGCGGAAATGCCACGGTGGCTGCCGAGACC
AGAGCATCATAGCCTGGATGGAATAATCCCTTTTGTGCTGGCGGCAACCTGCAGGGCGCGGAGCTGGTGG
TGGCGTACCCCTAGCACCTGGTGGCTCCCTGGAAGACGACACACCCACCCCGGAGGAGCCAGT
GTTCCGCTGGCTGGCTACTCTATGCTCCACACCCGCTCATGACAGACCCCGAGGAGGTGTGCCAC
ACGGAGGACTTCAGAAAGAGGAGGCACTGCAATGGGCTCTGCGCACACCGCTGCTGGAGTGTGAACG
ATTCAGCTACCTTCATACAACTGCTTCGAATCTGCACTACGCTGGGCTGTGATAAATACCCACATGAGAGCC
AGCTGCCGAGGAGTGGAGATACCGGGAATCTCTGATCGTTTCATGGAGCAGGTTTCATCGGCATTAA
GGCTTGGTGAAGATTACATGGAAGGAATCCAAACGCCATTTCTCGTAGAAGGATTAACCATGACATC
CGAACAGCAACGATGGGATTACTGGCGCTCTGAAACCTGGAGAGTATGTTGACAGCAAGGCGGAAG
GTTTCACTGCTCCACCAAGAACTGTATGGTTGGCTATGACATGGGGCCCAAGGTGTGACTTCACACTTAGCA
AAACCAACATGCCAGGATCCGAGAGATCATGGAGAAGTTTGGGAAGCAGCCCCGTACGCTGCCAGCCAGGG
GCTGAAGCTGCCGGGGCGGAAGAGACAGCGTGGGTGA

>SGPR538_1_SEQID_4
ATGTGGCGATGCCACTGGGGCTACTGCTGTTGCTGCCGCTGGCTGGCCACTTGGCTCTGGGTGCCACGAGG
GTCGTGGCGCCGGAGCTAGCACCGGCTGTCACCTGCGGGGATCCGGGACCGCGGAGCGCGGTACTGCTC
CAGGAGCAGACCTGTGCTGCCGCGCGCTGCCACGACTGTGCCCTGCCCTACCTGCCGCGCCATCTGTTACT
GTGACCTCTTTGCAACCGCAGCGGTCTCCGACTGCTGCCCTGACTTCTGGGACTTCTGCTGCGCGTCCACCC
CCTTTTCCCGGATCCAAAGATGTATGATGGAGGTGATCTATCAAGTCTTGGGAACGTAAGTGGGAGCACTGT
AACCGTTGCACTGCCAGGAGACAGGAGTGGCAGTGTGACCAAGAACCATGCTGCTGGGATCCAGACATGA
TCAAAGCCCATCAACCAAGGCCAATATGCTGGCTGGCAGGCTGGGAACACACAGCGCTTCTGGGGCATGACCCCTGGA

WO 02/00860

3/68

PCT/US01/20171

FIG. 1C

TGAGGCGATTGCTACCGGCTGGGCGACCATCGGCCCATCTTCTCGGTGATGAACATGCAATGAATTTATACAGT
GCTGAACCCAGGGAGGTGCTTCCACAGCCTTCGAGGCGCTCTGAGAGTGGCCCAACCTGATTCATGAGCCT
CTTGACCAAGGCAACTGTGCGAGCCTCTGGGCTTCTCCACAGCAGCTGTGGCATCCGATCGTGTCTCAATCCCA
TTCTCTGGGACACATGACGCTGTCTGTGGCCGAGAACCTGCTGTCTGTGACACCCACACGACGAGGCGCT
GGCGGCTGGGCGTCTGATGGTGGCTGGTGGTCTGCGCTGGCGGAGGGGTGGTGTCTGACCACTGCTACC
CCTTCTGGGCGCTGAACAGACAGGCTGGCCCTGGCCCGCTGTATGATGACAGCCGAGCCATGGGTGCG
GGCAAGCGCGAGCGCACTGGCCCACTGGCCCAACAGCTATGTTAATACACATCTACAGGTCACCTCCTCG
TCTACCGCTCGGCTCCACGACAGGAGATCATGAGGAGCTGATGGAGATGGCCCTGTCCAGCCCTCATG
GAGGTGCATGAGGACTTCTTCTATACAGGGAGGACATCTACAGCCACGCGGAGCTTGGGAGGCCAG
AGAGATACCGCGCGGCGATGGGACCCACTCAGTCAAGATCAAGATGGGAGGAGAGACGCTGCCAGATGGAAG
GACGCTCAAAATCTGGAAGTGGGCGCACTCTGGGCGCCGCTGGGCGAGAGGCGCCACTTCCGCACTCGT
GGCGGCGTCAATGAGTGCAGATCGAGAGCTTCTGCTGGGCGCTCTGGGCGCGGCTGGGCGATGAGGACAT
GGGTCACTACTGA

>SGPR414_SEQID_5
ATGTGCGAGAACTGCGCAGACCTGGTGGAGGTGTTAAATGAATATCAGATGTAGAAGTGGTGTGAGTGGCA
GCTCAGAAAGGAACACTACTCTCAAAATATTTACTACATCAATTCCTGGACACAGGCAATGTCTATGCTGCTTC
AAGGAATATAGCATTTGGAGATTTTAATCAAGTAGTGTGCGACTTATTAACCTTAGTGTGCCCCAAGTTCAAGT
GCTCGGGACCGAGCTTTGTAAACATTTGACTACATTAACATAGATTCACGTTGGCAGATGAGAGTAATCAAGC
AGAGAACCACTGAATATAGATAGAGAGTGAATGAAGGAGATCAAGAGAGCAAAATCAATAGAAAAAATCA
AACTCTACAAGAAATTTGTAATCTGAGTGAAGGAGATCTTCAAGAGTTCTGATCCTTTTATGTTGAGTACAG
ATGAGAAGGAAAACTCTTACTATGTGTGGCAAAATTTTCAAAATTCAGTTCCCTTATATCTGCTTACAAGCAT
AATACTACCCCTACTATTGAGATATCACTCAAGAAAGTAACATATTAGGGGCACTCTGTGATGATGATGATG
TAGAGTACCATTTGCTTGGTATGATGTTTGTGGGAAAAATGGCCCTTCTCTCATGAGGATTTG
CTTTGAATATGGAACCTCTGAACTTTGCCATTTCTTATAGCATGCGTTTATACAGTTGTGTCTAATATAGAA
TATGGCTACATATTCGCGCTGTCAGCACAATTAACCTTTTAGGACCTATGTTATAGGTTGTCTAATATGCAAGCT
CTCGGATCAGGAGTTACGACAGAGTGCAGCTGTAACATGGCTGACTTAATGTGGAGCACAGTCAAGAACCAT
TGGATACACATATGCTTTGATAAGAAAGCCTAGATCTTGCAATTAAGTACTTTATGACCTACTTTGACTATG
AGGTTGGCTGGATTGAGTCAAGATAAATCAACTCCACTGCTTCAATGATGTGCAATTAATGAATATTAGTAT
CGGACACAGAAAGCTCCATTGCAAAAGAACTTCGAGCTGGCTTTATAGCAAAATGTGGTGGAGCATATTTG
GACCAATTTACATATTGAGATTATCAACAGTGCAGAGTGAATTTTGGCAGAGAGGGGCGACTGAG
TACTCAACATATTGACTGATTTGGGCTGCAGCACAGTTGAACATTTGATGCGGTATATACATGACTTATTCCTT

FIG. 1D

CACTCATCAAGAAATTTGGATCCCGTACCAGCTTAGACATCTACTTAATCTGGTCTCAGCTCTTGAGCCAAAGTGTTCAT
TACTGAACAGACACTGTACTTGGCATCCATGTTAATTAAGCACTGTGGAATACGCACGTAGCAGCTAAAGGCTCA
GTTATCTAAACAGAGTCTTTTGGCATCTTTAATAACTAATATCCCATTTGGAATTAAGAAAGGGAAGAGAGC
TTAGAAGAACAGCTCCATCACTTGGTACCTGACGTAGTCTCAAGCAGTGAATATAGGATACACATCAAA
GTGGAGGTAGTGACATTTGAATGGATGAGCAACTTATTAATAGAACCAACATGTGTGAACAGCACTTTTCAGACA
CAGAGGAATCCATGCGAGGAATTTCTGACGAACCTGCCAACAGTGGTGAAGATGGAAGCAGTGGTCTGGTAG
CAGTAGTGGCATAGTATGATGATCAGACATTCAGAGTTAATTTAGCCACGCAAGCCAGCTCAGCTGGGAGCCCTG
GCAGTGAAGGTACAGTCAAGACATTCAGATATTGAAGCCCTCAAGAGGAAGATGAAGACGATGATCATGGT
CATATCTCCCAAAAGCAGTTGGTACAGATCTTCGGAATGAAGAGTTAGAGAGTCAAGCAGGCAATTTGCCTG
GGGACCTCCCAAGCCGCGTCAAGAAATGGGCAAGCAGCGGAACAAGGAAGACCTGGTTTTTAACACTG
AATCATTTGCCATCAGTAGATAATCGAATGCGAATGCTGGATGCTTTTCACACTCTGAAGACCCAGAACATGATAT
TTAGGGGAATGAATGCTACTCATATAGCACAAAGGCTCTCAGGAGTCTTGTACACAGCACTGGGAGCTTCCT
TGGGAGACTATTGGGAATGAATTTAATTTGTCACAAATTTATGGTCCACAGCATCACACCCACCCACA
CCATCACCCACCACCGATGGGCAATGGTTGATGATGCTAAGTGCAGATGTCAGTTGTAGTGTCTCCCA
GGTTAGTGCAGAAATCAGAAAAAATATGGCTGAATTTTGAATGGTGAAGAAATCTGGATGTGAAGAGAGGCTAGTTCA
GATTAATTCACATGCGGAATGACATCTCACCTCCAAACACATCTCCCAATTTAGCTTCCATTTTACCATGAACAT
CTTAGTCAAGGACCTGTAGTTTCAAAACATCAATTCACAGTAATGCTGTACAGACATTAATTTGGATAATGTTG
CAAGAAAGGAATACTTTGTTGGGATAGTCCAAAGATGAAGATGSCAGTTAATCTTCTGAAGGATTAATAAT
GAAGCAGAGAACTCTTTGCTGTAGTATGTTGGTTTACAGATAGACAAATTCGAATGAGATTCATTGAAGGTT
GCCITTGAAACCTTGGGAAACACAGATCAGTAGTAATTTTCACTTCGCTCTTCTTCCAAACTATTTGGTACTTTTCAG
CAGTTTGGAGCAGTTACGATACACACTGGATACCAATGAGGAGAAAGAACTGAACACATGACATGAAGCTTTTC
TTTGATAATTTGGTATACTACATTCAACTGTAGAGAAAGAAACAAACATGCACTGTACGCCATAGTGCCTG
AAGTTCAAGTTGCTCTCAATTTCTGACTGTGATTTTCACTCTGGGATCACCTGATCATTTTCAGGTTAAGTTTA
GAGCAAGTTGACATCTTATGGCAATTTTAGTAGAAGATTTGAAATGTTATGATGATGCATCCATTTGGTTTTTAA
TCAAGTTGCAAGTAAAGATCAACATGCTATGGGTATGGAACCTACAAACATCTTTCTGGAGAGATGCCCA
GCTAAACCTGAAACAAATAGCATGACTGGCTTAAACCTGTTTCCGATCTGCTGCTGATTTGGCTAC
AGTGCTATGATGGTTGTTCAAAATCTGAGCTGTGTGATGACCAAATTTTGGGGCAATTTGTTAAGAGCACA
TCTGGTATGTCAGTCGACGCTATCCAGTATTAATCTCTATATTAATGTTAAACAGGTTTGGAGAGG
AGCAAGAAATTTATAGTAAGTGCATGGAGAGTCTTATGATGCTTCTAGCAGTCTTGAACAGGAATCACATCAAG
TCTCATGGTTATAGAAAGAGGACTCCTTATGCTGAAGACACATCTGAGAGGCTTTAGGAGAGGTTTGCATATCA
TCTGAGACAGTGGCAAAATGAAGGCACTGGTATTAGTAGTCAATTTGAAGCAGCTGAGTGAACAAACAGCTCTGCCC
GCTAAGGGTTGTATGCCAGCCAGCTGGACTTCCTGACAAGATGACTATTGAATATGATCTCTAGTGAACCAAGGTAGC

FIG. 1E

AGATCTTAGGGCTGAAGTAACTCAATTGGTATGAAATTTACAGAAAGAAACAATAAATCAACAGCTCAGCTTCAG
GAGTTTGGTCAAAAGCAACCGAAGAGAGTTTCTCTGAGGCTCATGGAGCTGTCAAGATGATTTTCATCTCG
ACACGAGTTAACACAGATTATGATGAAAAGCACCTTCATGAGCTTGGTTTAAAGGATGACAGATGGTATTTGTGA
TCCTTGGGTGCACCAAGAGAGAGCGGAAAGGGAGGTGTTCAGCTGCCAGCATCTTCCCTCCCAACCCCTC
AGAAAGACAACATCCAAATGCTTTTCTTTACAGAGCTCTTAACTACTCTCTTTTATTTATAGAGATGCTT
GCATCATTTAAACCACTCAGGAAAGTGGAGTGATGATGAGAGCTTACGATGTGAAGAACTTCATCTT
CATGCAAAATCTGTCTAGGGGTCTGGAGCTACTGATGCTCTTCTACATGCTCTTAATATGTTGATGGCA
TTCAGAAATCTCAGATGACAGAGTTTAAAGCTCAGTCTGATCACAGCTTAGACATGAAGTTTCACATTAT
CAATGTGGCTCTTGGTGGTGGCTCATTTGCTGTCTTTAGTGAATCTAGCCTTAGCCTGATAGCGATCTTTACA
AGATTGGCTAAAGAAATTGACTCTCTCTTATCTCTGAGACTGCATTCGATGAATCATGAGTGGTCTCTTAAG
TTATCCCTGTCAAGGCTGGATGGAGGAGACTCAATCAATCGTTCTTTCTGCTATTGGCTGCCCTCAACATTTGA
AATTTCTCTGATGCTCAAGCACTCAACCTTATGAGATGATGATGAGAAAGAACCAATATTAACACAGG
ATGTAAGAGTATTTTGGTTGTTATGCAATTAAGTTGACAACATACATATAAGGACGCTAGTCAGACAACGCTC
CTCGACTTAGATGCCCTTGGCAAGACATTTGGCTGACTGTATTCGAAGTAGGAGATCCTTTGATCATCAGGATGGT
AATGTAGAAAGATGATGGGCTTACAGGACTCTTAAGGCTTGCAACAGTGTGTTAAACACAAACCCCTTTAAAT
TTTCAAGGAAGGACAGGAATTTTGGAGATATCTTCAATCTCTGTTTGTGCGCAAGTCTAAAGGACCGACA
ACAGCCAAAGTGCAATACATCTTCAAGAGCTGCCCTTACGATTTGTTAGTAGAGTGGTAAAGGGTCTGT
TGAGAACTACAGCTAATACACAACCTGGGTATGGCACACACATGCAGTCCCATGCACCTTATAATGGGATTA
CTGGCCTCATGAAGATGCTCGTCTGATGTAGATTTGTTGGCCTTACTAACCTTGGAGCTTGTACTTACTGTAGCT
TCTACTATTCAGCACTTTATGATACCTGAGGCAAGACAGGCTGCTTCTCACTGCCAAGTATTCAGAGGATATGA
AGCACAAGACCACTCTCTGGAGCTTCAGAAATGTTACATATTTAATGAGAGTGAATGCAAGCATATAATCC
TAGACCTTCTGTAACACATACCCATGATAGGAGCCCTCTGAATCTGGGGAACAGAAAGATATGACAGATT
TTTTACTGATCTAATACCAAAATCGAAGAAATGCTCCCGAATGAAAATACCGTCAAAAGTTTATTGGAGGT
GTAATTACAAACAATGTTGTTATCTCTTGGATTGTGAACATGTTAGTCAAACTGCTGAAGAGTTTTATCTGTGAGGT
GCCAAGTGGCTGATATGAAGAACTTTATGAATCTCTTGATGAAGTTACTATAAAGACACTTTTGAAGGTGATAA
CATGTACTTGTCTCAATGTGGGAAGAAAGTACGAGCTGAAAAGGGCATGTTTAAAGAAATGCTCGGCATTT
TNAGTTCAATCTATGAGATACACATTTAATATGGTCAGGATGAAAGAGAAAGTGAATACACACTTTCCCTT
CCCATACGTTTGGACATGACGCCCTATACAGAGATTTCTATGGGAAAGAGTGAAGGAAAGAGTTTAA
AGAACTCAGTGATCATTTAAAGACTCAGAGAGCTATGAATATGACTTGTAGGAGTGACTGTTTCACAGGAAC
GGCAGATGGTGGACACTATATAGCTTTATCAGAGATATGAAATCCCATGCTTATAAACAATAAATGGTAT
CTTTTAAATGATGCTGAGGTAACCTTTTGAATCTGCTCAACTTGCATCTGAATGTTTGGTGGAGATGACGA
CCAAGACCTATGATCTGTTACAGATAAAATTTATGGACTTCTCTTTGAAAAGACACACAGATGTCATATGCTGTTT

FIG. 1F

TACAACGGCATGGAAACAGAGGAAAGAAATGGCAGAGAAATACAAATTTGATGTTTCGTGACAGATTACTAGAGTGG
 ATTTGGCATGATAACATGCAGTTTTCGAAGACAAACATTTTGAACATACATATTTGGATTTATGTGGCAAT
 GTGTAGTGTATTTCCAGTACATTACCAGATCCTAAAGCTGTGCTCTTAATGACAGCAAGTTAAGCACTTCCTTT
 GTCTAGAGACATTATTCATTTCTAAGAAAGCCACAGATGCTGAGTGATTTGAATGAGTTCACGAAACAGATT
 ATATAGTCAGCAGCTTTGAGTGGTTTTAGATCGATGAGGATGAGGACTGGTGGCCAAATGCAGTACTAA
 TTAAGTGGCCCTAATCAAAATTTGAGACAGATTTTCAGCGTTTGTGATCCATGTTATTCAGAGGCTGAGACCTGT
 GCATGCTCATCTCTATTTGCAGCCAGAAATGGAAGATGGGTCAGATGATGGATACCTCAGTAGAAGATATTGG
 TGGTCGTTTCATGTCACCTCGCTTTGTGAGAACCTGTATTAAATTATGAAACATGGTGTAAACCTCAGATAA
 CATCTACAGATATTTGCCCTTCCTTTACGAATTTGCAAAATGGGTGGAAGAGAGCCAAATTTTGGCTTTGATT
 GCAAGCTATATCTACAATGGTACATTTTACATGGGACAAAGAGCCCTGAAATCCTCAAGTTGAAGTGTATCA
 GAGGAAGAGGGGAGAGAGAGAGAGAGAGATATCCTCTCTGCGAGAGAAATACAGGCCAGCTG
 CCTTGAAGAGATGATAGCTTTAGTTGCTTTTGGTTGAACAGTCTCGATCAGAAAGGCAATTGACATTATCACA
 GACTGACATGGCAGCATTAACAGAGGAAAGGATTCCTCTCTTTTCAACATATTCGTGATGGCATCAATATA
 AGACAACTTTGTAATCTGATTTTCAGCCCTGTGTCGATACATATCGACTTTCGAGAACATATGATCTATGCTTTT
 CACATCAATAGCAAGTTGACTCTCGAGGAGCCCAATCCTTTCTTGAAGTTTGAATGCTTAATGGAGTTTGC
 GGTGACCTCCAGGAATGCCCTCTTTCATCTTATCTGCAAGGATATGGAGGTGATTAACATCAATCCT
 TCTCAGTGTCTAGATTGGTTGGCAGTGCAGACACCCGAAATAAGTGGCACACAGCTGGTCTTACAGAAATG
 GAAACTGGGTGAGCGGTTCTTTGGCTCACAATATCCTAGAGTGAGGACTCTGCAGCTTATCTCTGGTG
 TCCCTTATACCAAGCAATTCATTCGTCAGATGTTCCGGTCAACAGGCTTTTGCACATCCCAACCCGTGACCTT
 CCACTCAGTCCAGACACACAGATGCTACATCAGGTCTACACGCTGCTCTGTTGCTCTCAAGGCCAA
 CTTTATGTGATGCTGTTCATGGCCTACAAAGCTAGTGCCCTATTATAGCTTATGACTTACTGTTTAAATTC
 CAAAACTGAGAACTGATGTTTCCACATATTTCAATGATTTGTGAACTTTTCCAGCTTAACTTTCTGAGCCA
 GCAATAGCTACAAATCACAATAACAGGCTTTGCTTTCATTTGATACAATGCTGCTGACTGCTCAGAGAAAT
 TCCGCCCTTATTTGTCAGAACCCAGTGGTAACCAAGAACATTTGCCCTCAATTAATCCTTGCAGCCATGATCA
 GATGTTGCTCTTTTAAACCGTGGGATGCTGCCAGCTACTATGCGCTCTGAGGCTCTGCTGAGCAGCTC
 CTGCATTCACAGCACTGCTTCTCACCAACATCCAGTGGCCCTTAAAGATCTTACCCACATGCCAGCC
 AATCCCTGGAGCAGTAGAAGACTGTTTAACTGATGACGCTGTTTATAGCTCAGAGGCCAGATAGAGAG
 AAGAATTAGAGATATTAACAGTTCAAGAAACCAACCAATAGTTGTTACTACGTTGCTAGAGGCCGCTCCTG
 CTGGACTCTTTTAAAGTGCCCTCAGAACTATTAAGATCTGATGAAGACAGACTTCTGTTGTTATTTAATCGAG
 GATTGATTTCTAATGACAGTCTTTCAACACTTTGCATGATGATCAGAGGCTACAGCTTGCCATGTGACTGG
 AGATTTAGTAGAACTTCTGTCATATTTCTTCCGTTTGAAGTCTACAGCCCTTATCTTACAGAAAGATGTA
 AACAGCATTAAATCCAGTGGCAGGAGCGAATTTGAAATTTGCCCATAACTGTTAACTCTCTTAAATTCCTATAGTCC

FIG. 1G

TCAGAACTTAGAAATGCGCTGTATAGATGCTCTCAGAAGCACTGTACTTTTGAATCCCCATGATTTTCTTCATCACT
CTGGTCCCTTTTACAAAGCTTTGATCTTACCATCAGATTAATACCAATGTCTCTGGACCTTAATTTCCG
TTCTGGAAGAAATTAACGCTATAGGAGGAAAGCAATATAGGCCCTGCGCCCTGAACCTAATATGTGGC
TTGCCACAAATGTTGGAACCAAGTGAAGCAAGTACGTTTATGCTGATGCTGAGCACTACTTCTTTCT
TTATCAAGTGTGCATCTATTATGCGAGTTGCAATCAACTGTGAATAATTTACTGAAACATTAAGTAGAAGCTGA
GTGCTCTAGTTGCTATGAAGTTTGCCACTTCATCTTGCACTGTCCGCCCAACTTTTGAAGTGAAGTATGCCAGA
CTACGTCTGCTATGTCAAAAACCTGCATCAACGCTTTTGTGGAAGATCTGTTTCCGAGAAATATTAAGTTATGTC
CTAATGATGAAGCAAGCTTTTTAAACAAACAATGTGTACAGCTTCATGACATCTCTCTCTAAAGGTTCTCAAG
TCAAGTGTCTTTGGAAGCAAACTGTGCCAATTTGATCAGCACTTATTAACAACCTTATAGGCCATATAGAAAC
CTACAGTCTGATTTCTCCACCAAGTTGAAATTTCCAGCAAGTGTCTTTAAATGGGCACTGAGGCGCACTC
GTTTGCTGCTGTCTGAGTACACTGCCAACAAGTTAAACCCAGCTGTAATCCAACCTTCAAGAGCTTTTAAGCA
ATTGCAAGCACTTGTCTGCACAGAGAACAACACTCAACAGCAGCAAGGCCAAGAAAGAAACTTAAGAGATGAG
AAGCAAGCTCCCAATTAAGAGGCGGCTGTAGCAAGTAGAGGAGCAAGCTAGACAGCTGATCAGTGTGAG
ATTAAACAGAAACACGAGGAGTGTGACCCCAAGCAAGCTTGTGCAATAGCAGCAAGAGCTCTCAATATT
GATCCAGCAAGCTAGCAAGATCTCTCTCCCTGAAATAGTTCTGTTAAGAAATACCCAGATGGAATGCTCATCTT
CGTTTTCAGAAGACCTGTCAAAATATCAGGTGCACAGATCGAAGAACAGCTCCAACATGGTAGATATAGCAAT
TGAAAGAAATTAAAGCACTCCCACTGTCTCAAGGATCTACCCCTAGCTGAGAGAANTCTGATTCCTCTACTTC
TATCTCTGCAGTCTGTCTGACTTAGCTGACTTGAGAGCTGTGATGGCCAAAGCTTTGCCCTCCCAAGCAACCTGAG
GTTGTCTTATCTCTGATTTGCTGAGTCCCAAGAGCACTTATGATCATGACGAACAATGACATTTTAGATATCC
CTGTGTGAGGACCACTTGAATCTACAATGCATCTGCTGCACAAGGATATCTGGCAAGAAACCAAGCTCTTCTTGA

>SPR430_SEQID_6

ATGTCCTCTCTGAAGGTACATGCTTATCAGAAATCTGAAGTATGCAGACTGGGATTCAAGAGTGGAAAGGA
TCTCTTGAATTTGTAGAAAAGAGAAATAGAACTCAGCTAGTGTACACTACAATCTGAGAGGATCCCAAGGATAT
TCAAGCTGAACTATACATTTAAATATTTGGTGTCTGCAGGCTGGAGCGAAACAAGCGGCTAATGTTAACTC
TGAGAAATACAGCTTGTCTGTATGCAAGATGCCAATGAAGGCTGAGAGAAGATTAAGGTTGTCTTAGATG
CAGTGCATTAACACGATCTCTCTGCAGCTGTGAACCGCTGACGGGCTGTGTTGTTGGAAGCCCTCTGGGG
AGCAAGCACTCAGAGAAGAACCAAGCAAGCGTCTTCTATCAGCAATCAGAGCTTCCAGGCTGTCAAAAGAGGA
TTTGGAACTTAAGATGATATCCATTTCCGAAAGCTTCTGTGATACGCTGGGTAGAGCTTGATTAAGAGCTGAGCA
GGAATGGAAATAGCTGGAAGATCTCTTTTGACATCTACTTCAACACCTCTTAGATCAAGGGTGTGCTAGAAAT
GCTACTGAAAGAGGAAGAAATGATATCAACTGGCTCAGAAATGGAATGAAGATATCCCTAAGGAAGAAATGATGAT
CATTCAGAACAAAGGCAATGACATCCCTCCAGCAAGATTTAAATGAGCAAGTGAAGCAAGCAAGATTTGATG

FIG. 1H

AACAGTCAGAAGAGAAATAGGACATCAGGTGGGCTTTACCTTTACAGTCATCATCTCTTTTATGGTAGCAGAGCTG
GATCCAAAGAACACTCTCTGGTGGCACTAATAGACAGGACTAATGTTTCAAGCCAGACTCCCTCTGCCAAA
GAAGTTTGGGATTTCTCTCAGCCAGTTCTCTTTCTGTTAAAACTGAGGTGTAAACCAGGATTACACTGGCTG
GAATAACCAAGAGTGGCCCTTTCCTCTCACCACAGCAGCACTGCAGGGCTTCTCCAATTTGGGAATACCTG
CTATATGAATGCTATTCTACAATCTCTATTTTCACTCCAGTCAATTTGCAATGACTTTTAAACAGGATCCCAT
GGAGAAAATTCACCTCAATGCATTTATCAGACGCTTGCACACTTCTGTTTAAAGATATCTGTAATTCAGA
GACCAAAAGGATTTACTCAAGAGGTTAAATGCCATTTCCAGCTACAGCAGAGATTCTCTGGTTATATGCGA
AATGATGCTCATGAATTTTAAGTCAGTGTGGACCACTGAAAGAGATGGAAGAAATTAATAAACTTGA
AGACTGAACCTGTTCTGGAGAAGAAATTCACCAGATATTTCAAGCTACAGACATACACTTGCCTGTTATAC
TAATTTGGAGTTTGAAGTTTCAAGCACTCCATCATTTGTAAAGCATGTGGAGAGATTAATCCCAAGAGAACAGTTT
AATGACCTCTCTATTGACCTTCTGTAGGAAAACCCACTCCCTCTCTGTTTCAATTCAGATTTCTTTGATCTTT
CTTTAGGCCGGAAGAACTGGAGTATCTGTGAGAGTGTGGGAGTGTGCTTGTCAAGGACAAATTA
CAGGCTTCTAGGCTCTCATCTCCATTTGAACCGATATAGCTTCAATGTGCTCTCTGCTTAAACATAGATT
GGCAGCAAGTCATCTCAAGATACCTGACCTGTCTCATCTGACCTGAAATACAAACCCCTTTTACCC
TTGGTTGGAGTGCAATATGGCAATGTCTAGACCATTTGAAAGCTCTCAATGGTGAATTCCTGCATCACCGCC
CTTCTACACCTTCAAGAAATTCACCTTCAATCCAGAGCTCTTGGCTTTATGCTTGAATTCAGACAGTGAAG
TGAGCTAAACCGTTCTGTGGCCCTCAGCCAGAGACTTTGTGAATGTAGGCAACGACAGCAGGAGGAGCC
TGGAAAAGATTCAAAATTTGCCCAATAGAGCCTGACAAGTCTGAATTTGGAACCTCAGGATTTGACAGATTGA
GCGAAGAGCTCTTAGCAGCTGCTTTGGAGATAGTAGAGAGATGCTTCAACCATCTGAGTCATGAGATG
ATGATAAGCCAACTAGCAGCCAGATACCGGATTTGCAAGAGATGATATTCAGAAATGCCAGAAATCCAGACA
CTATGAAACTGAGAAAGCCCAACAACTACAGAGCTGGATCCTGCCAGTTTACTGAGATAAAGACTGTG
ATGAGATAAGAAACAACTCCAGAGGATCTCAGGAGAGTTGATTGGCTCCAGCAGTATGATGGAGC
GTAAAGGGAAGAGAGCTTCAGCAGGCACTGCTCAGAGCTTCAAGAGCAAGGCTTGGGAACAGAA
AGAAGATGATGACCTCAAGAGCTACCGAGTTAAGTCTTCAAGAGTTTAAACCTCTTTGTGATGCTTGGG
TTCTGATGAGGACTCTGGAATGAGGATGTTTTGATGAGTACAGAGCTGAAGCTGAGGAACTGAAAAG
AAATGCTGAGACAGGAAATCTGCTCATTTGACCCGCTCATCAGTTTGTGATGCTGATGCTGAGGATCTTC
TTCAGGTCAATTACATTAGTGTATATGACATTAAGAGCAAGCGTGTCTTACTTACAATGACCTGGAGGTATCA
AAATCCAAAGAGGCTGCCGTGACAGGTGATCGAGATCGAGTGGCTACATCTCTTTTATATGCACAGGAGATC
TTTGATGAGCTGCTGGAAAACAGAAAGAACTCTCAGTCTAGCAGGAAAGTGGGGAAGACTACCCGTCAGGC
CTCGTGTA

[illegible]

FIG. 1J

CCGAGTATAGCCCATGCTGGGACATTAGTTTGGACTTGCCTGGCTCTTGCACCTCTCTTGCCTCATGAGCCACGCA
CGGAGTAGAGCAGTGTGAACGGGAAAGCCACATACAGGAATACACCTCATCGGACTGCTTGGGAGGT
TTACGGGCCAGAGCATTTAGAGAGCAGTGCACAAATCAAAATGTGTAGTTGCCAAAGCTACAGAAATCTATCC
AAACGATCACTAATGAATATCTATCTGTCTGTCTTTTCTTATTCAAACGGCTTTGAATCTTCAAGTACAGACAG
GGCGAAGTCACTATACATATCTCTTCTCTGGAGCTGGATGACCGGTTTATGGCCTCAAGTAAAGAGA
GCAGATGAATTGGACATTCGAGCTGCCAACCAATATGGAACACAGAAATAGTATCCCTGTGTCTGTGG
TTATCACCAGGACCTTGAGAGTGGCCATATACAGACTTCATCCGGCACACAGAAATAGTATCCCTGTGTCTGG
TTGATCATGCCCTCTACAGGCCAGTATTAAGGACGTCTAGGCAGTGAAGGATATTTACTGTCTGTCTACAC
AAACAGGTGCTAGAACATGAGTCTGACAAAAGTGAAGAAATGAACACACAGCCCTACTGA

>SGPR495_SEQID_8

ATCGGGTGAAGATCCAACTTAAGCTTACCTGAGAAAGCCAAAGAAAGTAAAGGCCCTACTGTACCTCATGAT
GAAGACTTCCAGATGATATTCGTGTAGGTTTAAGTTGCCAACATGAAGTCATGCTATCAGCGTGAATCATGTAA
AGAGACATAGCTGAGAAATCTGTGGTCAGTTGCTCAAGATGTTTAGAAGAAAGAAATCTTATGTGTGGGACG
TAGCTACTTACTGTGATATTTGGTTGTGCTCAAGTGTGGCTTCACAGGATGTGTAAACATCAGAAAGCCAC
ATTATTAAGACCTATTAAAGATTCCAGAGATCCAGACGCCCATTTATTTATTAATTAATCTGAGCACATGGATTATG
GTGTTAAGATGTTGAAGAAATTTCAACGCAATGTGAAGAAGTTTGGCTCAGATGTGTGAATCTTCTCCAG
AAACCTGCTCTTAACACACAAACAGTGCAATTTCTAGACTGATGAACTTTGTGAAGAAATGTGAACAKAGTG
AAATACAGAGGAGAAATAATGCAAAATTTCTGTAGTGAAGTATCAAAATTAGAAATACCTGCTTTTTTAAT
GCAGTCTGCAAGATCTGGCACAGACTTACTCTAGTCAATGATGAATGAAGATACAGAAAGATGATGACAAAG
TCAAGATTTGCTTCTTCAGACTCTACGTGCGCCATTTGTGTGGAACTTTCAAGCGCTGGACACTGACTGAC
CAGCTTTGCTCTTCTTCACAGACGAAGAGACTGAAAGAGACCACTTTCTCCCTAAAGTTCTTTTAATCA
CTTTGTGACAGGACCTCGAATTAAGATTTTCACACAGGACAGTCAAGAGGCTTCTCATTTCTCTCGAT
GCAGTGGAGCAGAAACAAACGCAATACAAGCTAGCATTAAGCTTTCAAGCATTTAACACCCCACTACTAAACT
GCTGATGATGAACATAGAAAAAAGCAAGCACTCCACGGTGAAGAGTCCATCTGATATTTCACTCTATAA
TCGACAAAGGGTTTCAAACCTTTACTTTGGGGAAGATGAATTAAGTATGAGAAATCTACGGGACAGACATGTA
TGAGACAGTGGCAATTTACTATAGAAATTTCTCAACCTAGAGCTGCCAAGAACCTTCTCATCTAAAGAT
AAGAGCTCAACTCATGATCCAGAAATGATTAGAAATTTGCTCTTGGAGAACTCTCACATCCAGAAAGAT
AAAGCTTTGAATTAATGGGGATCTTAATGTTTGGCCAGCTCATGAATCTTGAGTCACTGTGAATGAAGAGCG
TACTGTACAGTGTGAAGAAAGCCAGCCATCTGGAAGCAATGTTGATGCTGACAGTGAAGCTTTCAGAACTGA
ATAGTCTTCAAGCAGACTGGGCTTGCAGTCCAGTGTGATCCGGTGTGGACGAGATGGACCCCTTTTACC
CTGTGACAGAGTTAACTGCTGTACCAACAGAGACTGACAGTGTGAAGAAATGGCAAGCTATTCT

WO 02/00860

11/68

PCT/US01/20171

FIG. 1K

GAACCTTCGTTTGAGCAGCAGCTGTAACTGGGGATCAAGATTTTGACAGAGAAATCAGCCACTAAATATTTGAAATA
ATTATGTTTTTTAGAGGGGAGCATTGGAGTCTTATAGTCCCAAAATGCTTTTCAGACCTTTCTCAGAGCTA
TATACTACTTCTAAAGAAATGTTCAATTTACGTCTCTCTACCACTTACATCTATGGAATTAATAATGGGGAATA
ATAAGCTTCTATGTGAGAATTGTACTAAACACAGAGTACCAAGAAACCAAGTTTCAGAAAAAGAAAGT
AGAAGGAGTTTACTAATGGCCAGGAACAAATTTGCTCATTTTCGCTGTCAGCTGTCCTTATCCACCTGAA
AGATTTTCATCAGGCTGGCTTGGTTCGTAAGTAACAGACATGTAGTTTCCACTTATGCTCGATTAGCAC
CATTCTGCTCTGCTACTTGTAAAGATGCAAGTGGGAGATAAGTTCTCTACGGTCTCTATGCTAGTAGTGAAC
ATAGTGGCTCGATGAGAGAGGCCACTACACTGCTTATGTGAAGTGAACACACCCCTCCAGGAAATATCGGAA
CATAACTAAAGAAAGAAATGTGCTGGTTTGAAGCGGCTGATGTGAATCAGACGCCAGTGGGTCCATGTT
AGTGACACTTACTTACAGGTGGTTCCAGAATCAAGAGCATTAGTGCAACAAGCCCTACCTTCTTTCTATGAAGAG
TATTATAA

>SGPR407_SEQID_9
ATGGAGTATCCAGTCCCATACTTTAGATCCCCGAAACAGGACTCTGATCCCAAGAGAGAAATTTGGTCAAAACCCATT
CITGCTTTGGTGATCGCATACAGACTGTGAGTTGGCCAAAGACAGCAGCTGCTTGCAAAAGCAAGCTTAATAAGTGG
ATGCCCTTTTGATACCGAGCAAAACCTTGCCATGGACCCCACTGGAACTCAAGATTTGTTATCAGCAAAATCGC
CCATATCCCTCCCGACCCATCAACTTTCCTACCTTTACGCTGTCTGAATGCTTCTCTGACGCTGTCTCT
ATCTCCACAGCCCTCATGSCATAAGCCCGAGGGCTTAAAGCCAGCAGATACCAAGAGTTCCTGACATTCT
TATGGAGCGGCTACACTTGAATCAACACAGGAGGCCCGGGCTCCACAGTCCCTGCCCATGGTCCAGC
TCCCTCTCCACCCCAAGGGAGTGTCTCTGCTAAAGAGTCTGAGTTAAATGATGCTGAGTGGGCCAACCTA
ATGTGAAGCGTTATCTGGAAGAACAGAGACAGCAAGATGGTGGATGTTTTTTTGTGACCTCTCCCTGACCATC
TCTCAAGTGCAGGCTGTGGTACCCTCTATGACCTCAAGGTTTTTTTTTTGTGACCTCTCCCTGACCATC
CCCAAGAAAGGATTTGCTGGGGCAAGGTGCTCTGCGGATTTTAAAGCCTTTTACCAAGGAAGAGAGCT
AGAGTTAGAGAAATGCCCTAGGGACTTTGCCAGTGACAAAGTCGGAAGTCTCTACCAAGCTGTGTCCTTTTG
GAACCACTCAGGCAGCATCCACTGTGCCACTTACAAACCTGTGCCAGTGCAGAGCTGGTGGCACCTTTTACA
ATGACTTTGTCTCCCTTAACACAGCTGCGGACAGACAGAAATAGAACTACAGTTATGAAGGCTCTAGT
CTAGACTTCTGTTCAAAGCTTCCACAGATATTTTATTAATCACTCCAGCTGGGAACAAATGGAGGA
AGTTACCAGAACTGGAGTTTGGAAAGAAACATGAAGCTGAGACTGAGACTCTGAAGGAGAGTACCAT
TGGCTGGTGTGGTGCTGTGAAGCTGACGGAAGTCCCAAGATGGAGGCCCAAGGAAGGGCGCTGGCCA
GCTGCAGCTGGTGTCTCCAAAGGGTCAACCTGAGCGGTTATGGGTGTGAGGCAAGAGCTGGAAACAGGAA
CCAGATGCTGCTGCTGGGGCAAGACCTGTGATAGTGTATACAGTCAAGCAAGCCAGCACTAGGGACAG
GCTTGACAGACGGCCACCTTCTACTCTGTCTTACACAGTCCCTCTCTGGGCAATGCTGGATCTCTCTCT

WO 02/00860

12/68

PCT/US01/20171

FIG. 1L

TTCTATGGACACCCAGTCATATTGGATGAAGGCCACCCTTAATGACCTCATTTTAAGGAGGGGCCCTGTGACACAA
ATGGCCACAGATTTTACTGGGGTTTTCTGCTGGAGGAACCTTGCTCTTGAATAATGCTGCTGATGGACCA
GCACCAAGGACGTTTCTCAGAAGAAAGCTGTCTTTCCCTGTCTCTACATTTTTCATAAGGCAGGTA
AACTCTCCAGCGCTGATGCTCATGGATTCTAGTGAAGAAGTTCACTCCCAACAAAGGGGCATCGTGTTCAT
GGAACCAAGACAGCTGGGTGGGAAGGCTCCCTGTCAAACTCCAAAGCTGTGCACTGGGAGGAATGAAC
AGTGGGATGGAGCCACAGAGTCTGCACATTTCAGACAGGAAGGCTGGCCCTCTCTCTCTGTCGAA
CCTGAGATTCAAACTACAGTTTACAAATTTGAGGAAGCTTTGGTCCAGGGCAGGCTTGGGGAAGAAAGTG
ACAACCACTCATCTAGGCAGATGCCCTGGGGTCCCTGGGTGGCATGCCAGCATCCATGTAACTGCCCA
AATTGTCGAGATTGACACCTCCAAATTTGCAATTTGGTTTCCTGAACACAGTTGAGAGTTGAGTACTTCTACT
TCCCTGTCTCAGTTTTTCCCTCAATGATTTCAACCAAGAGGAACATACCTCTCAATCCCTGCTCCCGGGTCC
CCAAGGACAAATTCATTTCCCAAGGACAAATTTGCCCAAGGACAAATTTGAAGGTGATATTGTCCCTGCTGACA
ATGTATGAACCTAGACCGATTATTT

>SGPR453_SEQID_10
ATGCTAGCAATGGATACGTGCAAAACATGTTGGGACGCTGCAGCTTGCTCAAGACCAATTCAGCCTCAACCTCA
GAAATGGCACTGTGTGACTGCAACACGACGAGTCCATTTGGGCTTGCTTAGCTGCTCCCATGTTGCCCTGTG
GAAATATATTGAAGACATGCACCAAGCATTCAAGAAAGCAGTCATCTCTTGCTTGGAGTGAATGAGA
TGTACGTTTTTGTACCTTTGTGATGATTATGTTCTGAATGATACGCACTGGAGACCTGAAGTTACTAGACG
TACATTAAAGTCCCAAAAGTCAAAATTTCACTGCACACTCGTAGTGGGAGGTTTTACGGTCCATGGGTACA
GGTGATGATTTCTTTTACATGACGGTGCCCAATCTCTGCTTCAAGTGAAGATCAACTGTATAGTCTCTT
GGCAGAGGAGGAGTACTAATGGGTAAATCTTTGAAACATGGTTTGAAACATCACCATTTGGAAGAAAGC
AAGAAGACCATTTCCAGGAGAAATAGTAGTAAAGAGAGTAAAGAAAGCGGAGGAATGGAGTATCAAG
TTAAGCAGAAATGGAAAGTATGCCCTCCAAAGAGTTTACGTTTACAAGGGCTCGCTCAGTCGACCATATAG
AAATAGTTTCTGTTCAAGGTCCAGCACAAAGCGCAGCATCCAGCAAAAGATAAGTACTCTCTACCTCAGAA
ATGAAATATCTCAAAAGTCAAGTGACTCTCAGTTAAACGAAAGGCCAATGTAACCTCGTGGTGAACGGAATGA
GAAATTTGGGAAATACTTGTCTATGAAATCTGTTCTTCAAGGTGTGAGTCAATTTCTTTCGACAAATGTTTT
TAAAGCTTGATCTGAACCAATGGCTGCTATGCTGCTAGCGAAGACAAAGATCTGTAAGCATCCACAGTCA
CAGATACAGTAGTATATCAATGAATGATGTCAGGAAAGATACAGTTTTGTTGCTCCAGACATCAAGTCT
GTACATCAGGACTAAGTGGTGAGCATCAAGGATGAAGATGGAACATTTACGCCAAGGAGGCCAATTCACA
GTACATTTCTCTTTGTCAATGAAATGTCATCTTTGTCAAAGTCAATGGTGGGAGGCTGGGCTGTCACCA
TTTGCTATGCTACACTCAGTGTGGAGACTCATTCGCTTTGCTGGTTACGCCCAACAGACGCTCAGGAATTT
CTTTGTGAACCTTTTAGATAAATACACGCTGAATTAGAGACAACTGTTACCAAGTTTACCAGCTCTTATCCCCACTT

WO 02/00860

13/68

PCT/US01/20171

FIG. 1M

CTCAAAGGAACTCATCAAACAAGTTCTGAATGTTGTAATAACATTTTTGATGGACAACCTTCTTAGTCAAGTTACA
TGCTTTGCATGTGACAAACAATCAAAATACATAGAACCTTTCTGGGAGCTTGCTATGGAGTTTCCAGAAAGGTATC
AATGCAGTGGAAAGATATTGCTTCCAGCCATGTCTGGTTACTGAAATGTTGGCCAAATTTACAGAAACTGGAAG
CTTTAGAGGAAATCTACGTATGTGACCAAGTAACTCAAAAGCGTAGAAGGTTTCTCTCAAACAGTTGTACT
CACAGAAGCCGAGAAACAATTAATATATGCCACCTACCTCAGGTTCTCAGCACTGCCACTCAAAACGATTCAGGTG
GTCAGGCGTAATACCGAGAGAGATTTGGTTTCTATGTTGGCTTTAGGAAATCTTAACATGGAGCCCTATTG
CTGAGGGAGACCTGAAATCCCTCAGACCAAGTGTCTATCTATGCTTGCCTGCGCGGTGGTGTGCACCATG
GGAAGGATTTGGCTCAGGGCACTACACTGCTACTGCTATAATTTCTGAAGGAGGTTCTGGGTACACTGCAAT
GATCCAAACTAAGCATGTGCACATGTGATGAAAGTATGCAAGGCTCAAGCTTATATCTGTTTTTATACCAACGAG
TTACTGAGAAATGGACATTTCAAACCTTTTGGCTCCAGAGCTCCTGTTGGGGAGCCCAACATCCCAATGAAGACGCTG
ATACCTCGTCTAATGAAATCCTTAGCTGA

>SGPR445_SEQID_11
ATGCGGGTGAAGATCCAACTAAAGCTTTACCTGAGAAAGCCAAAGAAAGTAAAGGCCCTACTGTACCTCATGAT
GAAGACTCTTCAGATGATATTGCTGTAGGTTTAACTTGCCAAACATGTAAGTCATGCTATCAGCGTGAATCATGTAA
AGAGAGCAATAGCTGAGAATCTGTGGTCAAGTTTCTCAGAAATGTTAGAGAAAGAAAGATTTCTATGATGGCGAGC
TAGTACTTACTTCTGATATTGGTTGTGCTCAAGTGTGGCTCCAGGGATGGTAAAACTCAGAAAGCCAAAC
ATTCAATTGAAGCACTTTAAGAGTTCCAGAACAGAGCCCATTTGTTATTAATTAATCTGAGCACATGGATATATG
GTGTTATGAATGTGATGAAAAATTAACAGCATTTGTAATAAGAGGTTTGGCTCAGATAGTTGATTTTCTCCAG
AAACATGCTTCTAAACACAAACAAGTGCATTTCTAGAACTCATGAAATTTGTAAGAAATTTGTAACAGATG
AAATACAGAGGGAGGAAATGCAGAAATTTATCTGTAAGAGGAAATTCAAATTTAGGAAATTTCTGCTTTTAAAT
GCAGTCAAGCAACTTGGCAGACTTACTCTTACTGATCTGATGATGAGATCAAGAAAGTATGACAAAC
TCAAGATTTTCTCTCCTCAGACTCTCAGCTGGAGCCATTGGTGGTGGAACTTCAAGGCCCTGGACCACTGACCT
CAGCCTTGTCTCTGTTTCTTCCAGCATGAAGGAGACTGAAAGAGGACCCTTTCTCCTAAAGTTCTTTTAAATCA
GCTTTGTGAGAAAGCGGGTGCATCTACATTTAATAA

>SGPR401_1_SEQID_12
ATGACTGTCCGAAACATCGCCTCCATCTGTATATGGGCAACCAATGCCTCTGCTCTGGAAAAAGACATTGGTCCA
GAGCAATTTCCAAATCAATGAACACTATTTGGGATTTGGTCAATTTTGGAAACACATGCTACTGTAACCTCGTGTTC
AGGCATTTGACTTTGCGCTGCCATTTCCGGGAGAAATGTTGGCAACAGAGCCCAAGCAAGAAAGAAAGAAAC
TTGCTGAGCTGCTTGGCGGACCTTTTCCAGCATTTGCCACAGAGAAAGGTTGGCGTCTATCCACCAAA
GAAGTTCAATTCAGGCTGAGAAAGAGAAATGATCTCTTTGATACTACATGACGAGGATGCTCATGAATTTT

FIG. 1N

AATTATTTTCTAAACACTATTTCGGGCATCTCTTCAGGAGGAGAAGAAACAGGAAACAAACAAATTCGAAATTTAAACAAATGCACATTAAGCAGCACTTCGGGAAATATAAACAGAACTCACTTCGGTCCATAGATATTCTTCAGGGAACGCTATACCAATGAAATCTGGGAGCATTCGGGAAATGTTTAGTAGCAAGCTTATCTTGACCTTCTCTGTGGAAGTGTGGAGCAGAAATACATCTTACCCACTGTTAGAGACTTCAGCAACACAGAAACACTGTGTAGTGACAAACAAATATATTGTGAAACACTGCTGCAGCAACCAAGAAAGAGTAGGAGGTAAGAAAGCTGCCCATCTGGTCTGGCCCTGCACCTAAAGCGGCTTCAGTACATGAGCAGCTTCGCGCAGATACACCAAGCTGCTCTAACCGTGTGGTCTTCCTCTGGAACTCCCGGCTTCCTCAACCTCCAGTGATGCAAGCTTCGAGCGCATGATGATCTTGTTGGGTGTGGTGTCTCACTGTGGCAGTGGTCTTAATCGTGGGCATTTATCACTATTGTGAAAGATGACGGCTTCGGCTCTGTTGTTGATGATGACATTGTAGAGAAATATAGCTCTCAAGCTATTAGAGAAATTTCTATGCGCTGACCGTCAGCATATATCAAAAATTCAGAACTCTGGATATTTTATTTCTATCAGTCAAGAGATTA

>SPR408_SEQID_13

ATGTGGCCGGCGGAGAGAAACCAACTGTTGCCGAAAGAGGCAACCACTGGATCATACCAAGTGACAAAGTCACTCTTCAGCTCGACCTAATTTTAGCCGAGAGAAAGAAACATTTCTGATTTGACAGATTAAGATGTGAACCAACCTCAACATATCTGTGAGATTTCCAGCTGGGAGAGACAGCTGTTTCATGACAGGTTTATAGTCCGCTTCCAAAGAAAGCTTGCTGGGTTCTCCAGCTGATTATTCAGCAACAGCTCATCTCACTCATCTTTTGAATTAATCAACCACTGGATATGTGGAATAGTAAACCAAGCATGACTGCCTATTTGAATAGGCTTTTGCAAACACTTTTATGACTCGAATTTAGGAATGCATTATAGTGGGAATTTGAAGAACTCTGAAGAAATCCAGTGCAGCATATCCATCAACCACTTCMAAGGCATTGTGTTTGTTCACCAACAGAAAGAGACAACTTGAACCAACAGATGTTTGAACGAGCTTTGGATGGGATAGTAGTGAGGCTTTGGCAGCAGCATGATGACAAAGCATTTCCAGACTCATGTTTGAATCTTTGGAACAGAAATGGGAACCAACAGCGCTTTAATTAAGAGCTATTATCAAGCAAGCTGAAGCACTAGTGTGCTGGAATGTGGTATTAGGGCTGGCGAATTCAGACATCTTGAATCCCATGGTTCACGCACTTATGGTTCGACGCAAGCTATGCTGATGTGAGAAAGCAATGTCATGCAATTTATCAGCAGAGATCTGGAATGCCCAAACTCAGTATTTTGTGAACCTTTGAAGAAAGTGTGATCAGCGAAGGGCTTCGGTTTGTGCACTTCCTATCTGCTGCCTTACAGCTGAAAAGATTCGATTTGATTTACACCATGCTAGAGATAAATCAATGAATGACGATGCAATTCOCGAGCAAGATATGAGTATGAGTCTTTTATGATTTGAAGATGAAAATCTCCAGACTGAAAGTGGCACTGACATGGAGAGCTGAATGAAGTGGTGTTCACAGTATCAGATGAGCAACATTTCCCAATGATGGTGTGTGAAGAGCACTGCTGGAACCAATAGTGGAACTGAAAGACTCMAAATCTGGAACTTGAAGAAATCTCTTGATCTATGAACCTTTGCTGTATGCTCATCTGGAGCGCTGCTGGTGTCTATTATGCAATGATAAAGTCAATCAGTGAGCAGTGGTAGCAGTCTGGATGATCAACATGTCAGCAGGATACCAAGAGCAATGAAGAAATGAGGATCTTCAGGAAGCAGGAATATTTCTAGTGCCTTTCCGAAATCCCAAACTGATCTGATCTGATCTGAGTGAAGAGCTGACGAGAAATTCMAAATTTCTAGAGTGGTGGTAATCCCAAGCATATAAATCACTTGTGTSCAAGAAAGAGAGA

FIG. 10

GTGGAAGAACCAAGAAAGAGACAACGAGAAATTTGAGCGCAATACATCCAGATAAAATTTCTGTTTGCATCCT
ACAAAACAGTAATGATGAAATAAATTTGGAGTTTCAAGGATTGCTGTCGCTTGTAAATATGATGAGTTTCATGATTA
ATAAGATGATGATTTAGAAGAGTAAATACCCCTGGATTGCTGTCGCTTGTAAATATGATGAGTTTCATGATTA
TCTAGAACCGTCATTTGAAGGAGAAAGAGATACACCAATGGGGCTTCTTAGTGGCGTCAAGTCAACATATAT
GTTTGATCTGCTGTTGGAGCGAGAAAGCCGTGACAGGTTTTTCAATCTTATAAACCCTGGAGAAAGTGAATGGTGA
AGTTTCAATGTTTGTGATTTAAAGCGCAATCTGTAGCTGCTCTTAACCTTTCGCTTACTTAAATCAGACAGTT
ACAGAAATCAAAACAACCTGAATTTCAAAGGCCATCCATTACCTGCTGAAACAATGAGAATAGTCTGGAACGCTGC
TACAATGATTTGCGCTTCTCAGTCTCCAGTAAACCCCTGAAGCTGAAGGATTTTTAGAAAGTAAACAAGGTGT
TTGTTGAAGCTCCGAGACTTTGGATTACAGATGGCCCTTTGACAGCTCTCATTTATGGAACCTCCTGGATCGGC
ATGCAATACAAATCAGATTATTTGTTTGTACCTGAACAAATCCCAGATCTTTATTTCCAAAAGGACAGCATACCA
GAAAGCTGGAGGCGATTTCTGGTAATGGATGAGCTGTGAAGAGTCAAAAGCTGTCACTGCAGCAACAGCAGATGGAGTA
CTGTGGAAGCTATTCTAGAAGAAAGCAGCTGAAAACCTCAAAAGCTGTCACTGCAGCAACAGCAGATGGAGATA
ATGGGACAGCAGCAAAAGTACTGAGACAAGTGAAGTGTGAAAACATCGAATCACCTCTCAATGAGAGGAGCTCT
CAGCATCAGTGGATTAAGAGAACTTGAACAGCATATTCAGACTTCTGATCCAGAAAATTTTCAGTCTGAAGAACG
ATCAGACTCAGATGTGAATAATGACAGGAGTACAAGTTCAAGTGGACAGTGATATTTTAGCTCCAGTCATAGCAG
TGATACTTTGTGCAATGCAGACAATGCTCAGATCCCTTTGGCTTAATGGCTTGAAGTCTCACTCTCAGATACACAAAGT
AGAGAACGAAAGCAATGAAGGAAAGAAAGAAACATGGGATACAGCAAGAAAGACTCTGGAAGTGAATGTA
ATATGATGAGAGTGGCAAGTAGGGGAGAAATGCAGTACATGATTTCAAAGCTGAACCTTATGCTGCAGATGA
AGGTTCTGGGGAAGGACATAATGGTTGATGGTGATGTTGATAAAGAAATTAATTCCTCAGCTCAGGGAGCAATGGT
TTTAGAGCCCTTTTGGAGTTTGTCCCTCTCACTTCAAGGCTTTTCGAGTGTATGCCAGCAATCAAGAGTTTGAG
AGCGTCCGGCTGAATGAGACATTTTCATCATTTTCTGATGACAAATGAAGTACAAATAGACTGGGAGAGCACTT
AAAAAGGAGAAATACAGAGTTAAAGTATACCACTTTTGGTCAATGAACAAGAGCCATGCAAGTTTCTGCTAGAT
GCTGTGTTGCTAAAGGAATGACTGTACGGCAATCAAAAGAGGAATTAATTCCTCAGCTCAGGGAGCAATGGT
TTAGAGCTCAGTATTGACAGGTTTGTCTAAGGAAAGAAATCGGAAGAAATCCTGGCAGCTGCTTTTGGATTATC
ATATTTATGAAGAGATATTAATTTCCAGCACTGGGAGGTTTTCCTTGAAGTTCTTGTGGGTAGAGAGAT
GAACTCCATGTCACAGCTTGCAGTTTGTCAAGAGCGGTGGAAGCTTCAGAGATGAAGTTGGATCCCTTCCAGG
AGGTTGATTTGGAAGCAGTGTGTGACGAAATTCGAGAGAAAGCTTATGAAATCAGTGGGATCCCTTGGATG
ATATTGAATTTGCTAAGGCTAGAGGAACATTTCCCTGTGATTTCTGCTTGAATTTATCAGGATTTAGACTG
GAATCCTAAAGTTTCTACCCGTAATGCTGGCCCTTTTATCTGTGATGATGGTGGGTCATATTTTATAGGGAT
AAAAACAGAAAGAAATTAATGGATTTGACAGATGAGCAAGAAATGAGTGAAGAAAGAGCAAGTGCAGCTCCAG
AAGACTGGACATCGTGTAAACATACTCACCTCGTAAAGAGAAAGCACTAAAAATATATCTGGATGGAGCACCAGAT
AAAGATCTGACTCAAGACTGA

FIG. 1P

>SGPR480_SEQID_14
 ATGGGTGCAAGGATCGGATCGATTCTCAGCTACGAGGAGGCGCTGAGGAGATTACAGATGTAGAGC
 TAAACGACTGAAGGATGCTTTCAAGAGGACCTGTGGACTCTCATATTACATGGGCCAGCACTGCTTCATCGGG
 AAGTGTCTGGGATGGAGTGCCTCAAGGTTTGGAGTGATTTACTGTCTCTTTGGTGGAACTCCAAAGGG
 CTGCACTTCATTAATTAAATTTGAGTCTGCTCTTACAAGAGGCAAGATGAAGAGCAAAATACATTT
 TTAGTCTTTTCAAGTGAATCTGGGAATATGTTATACGGGAAGATGGAAGATGCTCCACGTGGTGATG
 GTAAAGTCCAGATACACTCAGGAAGTGTCTCAGAGGGTGAAGGTAAACTATGAAGTTAGAAATTGGC
 TTTTCTAAACAAGATGCTTTTACTCTCTCGATGGCTCTATCTGGAGGTGTATGTACCTCACTGATGAT
 AGTGATACCTACTTTCTACCAAACTCTGGCTGGAGTCAACATTTGGAGGAATCAGACATCAITGATCTTGAGA
 AACGCTATTGGTTATTGAAGGCTCAATCCCGACTGGACGATTTGATTTAGAGACATTTGGCCCAATTGGTTTAC
 CACTATTGCTCATCTCTAAGTGAAGTTTGTAACTTTTATGATAATCGTGACATACATAGATTTAAG
 GAGATACCTGTGGTTATCAGCTGTTCAGGGGACCCCTGGCTGAAGACAAAATTTGCTTCAAGTATTT
 GATGTGACCGTATGAGTCTCTCAGGGTTGAACGTGAGAGACATGGTGGTTGGCATTGTAGAACTCTGGAA
 GGACAACCGCACTGATGATATTCCTGAATTACATATGATCTCTCTGATATTGTAAGGCACTACTGAATGCACAT
 GACACCAAAAGATGGGTCACTTACTCTGGAAGACTATCAGATCTGAGTGTGAAAATTTCTTGCCAATGAG
 TTTTGAACCTCTCTTCCAGGTGTGCACATAGTTCTGGGGTTAAGCACAGCTACTCCGGGAAGAGGACAA
 ATTATAGAGGATGTTAGAACGAGAGCAGGTATGGTCTGCAAGCAGGACACAACTGGTTTATCATCTCCATG
 CAGTGTGGCAACAGTGGAAAGATATGTCAATACGATGCCACCTGTGGTAAATTGAGCCATCATCTGTTTG
 AATGGAGAAATACTCATTTGGAACTGCAAGGCTATCTATGAGCAGGTGGAAGTGAAGTGGAGCGCT
 CAGTTACGTGAATACAGAGAGAAATTTTCAACAACATTTCTACTGCACTGGAAGCTCAGAACTGCTGGC
 AGCGGCTTCTGTATTTGCCACACCGGGCAGATGTTGCTTTGCTCGACACATAACACTTCTGACAATAAC
 AACAGTGTCTGGAGCCATGGGAATTTTGTGCACCTTAACCTCAGAAACCAGGGGCTATTGATAAT
 CAGCCATTAGTAACCAAGAACAGTAAAGGCTACATCTTAACACTAGAGGAGGAGATTAAACGAACCTCCA
 CAGCTGATTCATGGAAGAGACTATGAATGGTCCAGAACCTGTGGAGGACCTTTATCTGATGGAGCA
 AACCTGGGCTTACCTAGACCACTTATCAAGACAGCAAGACAGACTCCAGAGCTGGAAATTTTCCCGCTAT
 CTCTCTTCCTGAGACAGCAGCTGCCACTCGGACACAGCAGTCTAACATCTGGGTGAATATGGAAATGTACCT
 TCTCGAATGCACCTTAAAGCGGTTATTAGCCTATACAGGCTGTTTATGTCGAATGCAGACCATCAAGGAAAT
 CAGGAATATCTATCTAAAGGCTGGCATTAAAGAGGAAGATATGCCCTGTGGCTATACACAGTGAGAACTAC
 CTTACTCTCTGGATGAGGATCAAAATTTGGAATTTTGAATCCAGGATGAACACACCTGGTAAATTTGAAG
 TTCGAACAAGATATGAGTTGGCCTGAGGAGATGCTTTTATAGCAATAGTAAATAGATAGACACAAGGT
 TCCACAGAAAAGGGAGCCACAGGTCTAAGCAATCTGGGAACACATGCTTCATGAACCTCAAGCATCCAGTGTG

FIG. 1Q

TTAGTAACACACAGCCACTGACACAGTATTTTATCTCAGGGAGACATCTTTATGAACTCAACAGGACAAATCCCAT
TGGTATGAAGGGGCATATGGCTAAATGCTATGGTATTTAGTGGAGGAATCTTGGAGTGGAACTCAAGAAGTGT
TGCCCATTTAAAGCTTCGGTGGACCATAGCAAAATATGCTCCAGGTTTAAATGGGTTTTCAGCAACAGGACTCCCA
AGAAGTCTGGCTTTCTCTGGATGGCTCTCATGAAGATCTTAATCGATGCCATGAAGCCATATGTGGAAGTG
AAGGACAGTGTGGGGACACAGTGGGAAGTAGCTGACAGAGCCGTGGGACAAACCATCTAAGAAAGAAATAGAT
CAATTTGTTGGATTTGTTCCATGGCAGCTAAGATCTCAAGTAAATGCAAGCATGTGGGCATATAAGTGTCC
GAATTGACCTTTCAATTTTGTCTTGCCACTACCAATGGACATTTATAGCACTTAGAAATACAGTATTAAG
TTAGATGGTACTACCCCTGACGGTATGGAATAAGCTGAATGGATGAAGTACACAGGTTTAAAAACAG
CTGAGTGATCTCTGTGGACTTAATCGAACAAATCTCTAGCAGAAGTACATGGTTCCACATAAAGAACTTTTC
CTCAGGACAAACAAAAGTACGACTCTCAGTGGTGGATTTTGTGTCATCTCAGAAATCTCTGCTCCCTGTGTCTCC
AATTTACGCTTCTAGTCCAACACAGACAGATTTCTCTCTCGCCATCTACAAATGAAATGTTCCCTTAAGTACC
AATGGGACCTACCCGACCAATATTCTCCCAATGGAATGCCAAACACAGTGTGTGCCATGTGGAACCTGAGAA
GAAGTTCACAAATGGAAATGGTTAATGGTCCATGCCATCTCTCTGACAGCCCTTTACAGGTTACATCATTTGCA
GTCCACCGAAAAATGATGGGACAGAACTGTATTTCTGTCTCATCTCAGAAATCGCCCCAGCCCTTTTGGAAATG
CCATTTGATTTTCCATGTACTGTGTCATACCCGGAAGAAAGACCTATATGATGGGTTTGGATTCAGGTATCCCGG
TTAGCGAGCCCTCCACCTCAGGAAGCTAGTAATCATGCCAGGATGTGACGAGTGTGACGACAGTATGGCTATCAATAT
CCATTCACCTACGAGTTGTGCAAGAGTGGAACTCTGTGCTGGTCCCATGGTATAGATTTTGCAGAGGC
TGTAAATGATTTGGGAAGACAGAGCTTTTCATTTGGAATGCTATATCGCTGTGGATGGGATCCACAGCC
CTTCACCTTCGCTATCAACATCCGAGGAAGGTTGTAGATGAGCATGAGAGTGTGGAGCAGAGTGGGGAGC
GCAAGCCGAGCCCATCAACCTGGACAGCTGTCCGCTGCTTTACCCAGTGGAGAGAGCTAGGGGAAATGAG
ATGTACTACTGTTCCAGTGAAGACCCACTGCTTAGCAACAAGAGCTGGATCTCTGGAGGCTTCCACCCATC
CTGATTATCACCTTAAAGCGATTTCAATTTGTAATGGTGGTGGATAAATCAGAAAAATTTGCAAAATTCCTCG
GGAAGTTTGTATCCAAAGTCTTTTGTGACCAAGAGACCCGCTCTCTGCCAGCATAAACCACTCACACCCCA
GGGGATGAGCTCTGTGAGCCAGGATTTCTGCAAGGGAGGTGAAGAAAGTGGATGGCAGAGTTCGGGTGG
GGAAGAGAGCTGCTCTGTGAGCAAAAGCCCTCTCACTACGCTTAACATCATACAGCCCGGAAAGTTCTC
CTTCTTCATCAAGAAAGTGAACAGCTGCTCCAGCAAAACAGAGCCCTTAAGAGCCGACGAGCTT
TGGGAGGAGCAAGGAGGCTCCGCTGCCAGATTGGCAGCAAAATAAAGTGTCAAGTAGTAAAGAGAA
CTTGSATGCCAGCAAGAAATGGGGCTGGCAGATATGTGAGCTGGCTGACGCTTGAAGTGGGCAATGTG
CTGGGGGAGCCCAACAGAGTTTGGTCACTCTCAGGACATGAGGTAGCTTTGGCCAAATGGATTCCTTTATGA
GCATGAAGCATGTGGCAATGGCTACAGCAATGGTCACTTGGAAATCCAGAGTGAAGAGACAGCACTGATGACC
AAAGAAAGATACTCGTATTAAAGCCATTTAATCTATATGCAATTCGTCCTTACAGGAAATCTGGGTGGGG
CCATTTACGTCACTTATGCCAAAAACCCCAAACTGCAAGTGGTACTGTTACAAATGACAGCAGCTGTAGGAACCTCA

WO 02/00860

18/68

PCT/US01/20171

FIG. 1R

CCCGGATGAATTGACACCGACTCTGCCTACATCTTTCTTATGACGACGAGGGGATAGACTATGACAAATTTCT
GCCAAAGACTGATGGCAAAAAGATGGCAGACACAGCAGTATGGATGAAGACTTTGAGTCTGATTTACAAAAAGTA
CTGTGTGTACAGTAA

>SGPR431_SEQID_15
ATGGACAAGATCCCTGGAGGGCCCTTGAGTTCCTCGCATCCCTGCCCTCAAGCGGGTGATTTGCGGGAAGG
TGGTGAATCGGCGGACACTGGCTAGACGAGCGCAGTGGGCGCAGTGGGCGCTACGACCATGTTGACCTACGACCCGCTCAT
CCTGAGGGCCAGGACCCCTTCCAGCGCAGGTGGGCGCAGGCTGCTGAGGCTACGACGATACCCACCG
GCCAGAGTTCGAGTCCCTTCAACAAGACCTTCGCTGGGCGCTCTCATCAGGGCTACCACTCTCTGGACAG
GAAGGATGAGCCATCCTGGACTACATTACAAGCGGCTGAAGCTGATTTAGCTGTCCGTCGGTCTGGATC
TCCTTAGCTCCTCGAGGTAGAGGTGTACGGATGGTGTGAGAGCGGCGGCGCAGCTCTGTGCCGACT
GAGCGACCTTCTGACCGACTTTGTGCAATGCATCCCAAGGGGAAATTTGCCATCACGTTCTGCAACAGCTGGT
TCGAACGATAGGCCATTTCCAGTGCCTGCCACCGAGGAGAGCTGGGGAAATATGCTCCCGAGGTGACAA
AAGTGAGTAACCTTGTGCAAGCATCTGGAGGCGGCGCTGCCACACTAGCTGCCCTTCCCTGCAAGAGTTTGTG
CAAGCATCTCTCCACAGATGCATCAATTTGAAACCTTCTGAGCATTTGCAAGCTTTGCGAGCATATTCCTCTTCA
GATGATTACAGTTCTCATCAGGAGCCCTTACTACGGATCCAAATGTAAGATGCAAGTATGACCAAGCCCTTTG
CAGAATGATTGACTGGCTATCCTCGCCATTGGCTCAGCATGTGGATACATGGGTAATTTGCACTCCTGAAAGGACT
GGCAGCTGCCAGAGTTTACTATTTGATAGATGTTACTTTGCTGAAATAGAACTGGTTTTTAATCGACTTTGGT
TTCTCTTGTGAGACCTGGTCTTGGAGTCTTCTCAGATCTTCTCAGATCTGCTTAGCTTTTCAGCATCTCCAGAGCGGT
CCATTTGATTGTTCTCATGTGGTTAATTTGGTTCAATCTTTCAAAATGATGGTCTGCTTCAAGTACAGCCTTCT
TAGTACAATTAACAGAAATGATACACTGTATGATGATCTTCTGAGATTCCAGATCTCATGACCTATTTCTG
GAGCAATAAAGGATTTCTTAAGCCAGTGAAGAGAAATTAAGTTAATTTCTCAATCAAGTCCCTGGACTTCTC
AATCCAATCTTGGCGTCTGCTTGTAGACTTTCTGAAATCTGAACTGGGAAACTGGCTTATTAACCT
AGGAATACATGTTATGACAGTGTATACAGCTTTGTTATGGCCACAGATTTCAAGGACAAAGTATATCT
TTAATCTAAATGGGTGCAATTCATTAATGAAAAATACAGCATCTTTTGGCTTTTGGCCATACACAGAGG
AAGCATACGACCTCGGATATCTTTGAGCTTCCAGCTCCAGATGCTTACTCCAGATCAGCAAGACTGTT
CTGAATACCTCAGATTTCTCTTGACAGGCTCCATGAAGAAGAAAGATCTTGAAGTTCAGGCTTCACACAAGC
CTTCTGAAATCTGGAATGCAGTGAAGCTTCTTACAGAGTGTAGTAAAGCAGCAGTACTACAGAGACCC
CTGTCAGAGTGACGGTGAAGAGCTTTAATGAAAAATGTTTGGGAAAACTAGCACTACATACGTTGTTT
GAACTGCAGGAGTACCTACAAAAAGTGGAGCCTTTACAGATCTTTGCGTTGCTTTTGTCTCTCTTTG
GAAAAATGCTGTCCAAGATCCAGATCATCCAGATACAGATGAGTGGTCTAATGCAAGCTCTGTACCC
GGTCTCTCAGAAAGAACAGTAGTTTATATATCCACACACAGCTGCCCTTCTCTGTGACTCTCTGTGAATGAAAAA

CCATAGGCGAGTCTCCTAAATGAGTITTTACTGTTCTGAAACACATCTTGTCCTCTACGNAATCTCAACAAGATCTTGTGTT
TAATTAAGAATGTAACCTCTAGAAACAGGAGTGAACACACACACCTCTAGTAACTGCTTACTTAATATTTTITTTGGCT
CGAGAGATCTTACTGTTGATACCAAGATATTTATGAAACATGTGCTCTGCTCTGCTCAAAATCTGCTGAGAAACATCTG
AAATACGCGAGCACTGAATACCTTATCTACTCTCGAGATTTTCAATGATGAAACAGTAATCATCTGAGAACTG
GAAATTTTAGACAATCTTACTCGTCGCCACTGGTTTGGAGTGGCATGTAAAGAAATTACTCTTCTCTCATGTT
CGAAGAGTTGGTGTGTAGATGTGACTTCTACTGATCTTAGTGAGAACCCCTGCTAAAAATTTAAAGCCCTTCAGGAGCA
TGATGAAGCTTCTGTCGACAAAAATTTGGTGGCCCTATCTATTAACTGCCGTTTGTGTTCTACTCTGCTATATCTCTGAA
ATGGGGCATCTACTATCTTATGCGAANAATACACAAATCTTCAATCAGATGTACGACAGTCTGAGG
CTGTGGCATTAAGCATCTCCGAGAGCTATTACTTAGGAGAGATAGTCCGAGTCTTATTTTGAACAGGATTTGG
AAAAATGGAAGTGTCCAAAAGATGGTTTATTTAATGACAGTAGATGAGCATTTACTTCTTATTCAGTGACTCAGCA
AAAAATTCAGAGCAGTTTCCAAAGSACACAGCTATGTGCTTTGTATAAAAACACATAGTACTAATGGTTTAA
GTGGTATACCCCAACAGTGGACTCTGGATAAATGGAGACCCACTCTACACAGAAAGCAATGATGAGCTATATA
CAAAAGATAATAACTATCTATGATGAGCAACAGATTTGAATGTGATGACACAGCACACCGAGGAAGCTGTGGACCAAACTGGTGGAGGGGGT
CTGTGATTTGGGCCCAATGGATTTGATGATCAACAGCACTGATTTGA
GGAGGAGGATTTATACATTTGGGACAGCTGATTTGA

>SGP429_SEQID_16

ATGGCCCCGGGCTGCAGCTGGAGAAAGCGCGGCTGGCGCTGGCGGGAGACGGTGGCGGCCGAGGAGGTTGTC
GAGAGAAATCTGAGAGCCGCTTACCGCATCTGGCTGGAGCCGCTGCATTTGGCGGCGTGTCGAGGAGCAAGCTGCG
AAAGAAATCCGAATTCGTTGTGTTATGTTGAGCATATTCGTGAGAAATAGATGAAATAGTTTTCATTA
ACATGATGATCCCACTGTGAGAGGAGAAAAAGAACTCATTTGTGGCGCTGACTCACTTGTGAGCGACACTTGT
ATTATCAACATCTTTCTCAAGTGTGGTTCTCAACTTGGAGCTTCGGCAGGCATCTACTTATGTCGAAGCACTTGT
TAGTGACTACATGCTGGAGACGGCATCCAGAAAGAAATATGAGCCTCAACAAATTTGTGAGCATCTCCCA
GTACTTGTTCGCTGTTGCAAAACGATTAAGCCATAACNTATCCATCAGSATTTGTAAGCACTTGGGCGCTG
GACTGTGACACAGACGATGCTCAAGAAATTTCAAGAGCTTTATGCTCTTATGGAAGACTCTTGTGCTCAAC
AAAGAAATCAGATGTGCGCAATATTTTCAACAGCAGTCTGTGGAGAAATATGCTTATGAACTTGTGGCAACCA
TGTGGGCAAGAGCTCTAAGCTTTTGTCAAAATTTTATGAGCTGGATTAATATCCAAAGCCACCAACAACTTAACCA
GATTTATCTCGGAATTTTGAAGGAGAAAAAATTAAGAGSAGACAATCGCTATTTTGGAGAAAGTGTGCAAGAG
ACAGAAATGCAACAAGAAAGATCGACTTCTAGCCTTCTAGCCTTCTGCAAAATTTGGATATGAGAGCCCTTAT
TGACGGCAACTGACATAGAAAAGGTAATACCTACATTTGGCTCTCAGAAATTTGGATATGAGAGCCCTTAT
GTGAACATAAAGGTGGGTCTATGTGATTAAGCTACGCGAGCTCTCTACAGAGGAGTGAGTGTCTTATCT
GGCCATCTACATCGGCCGCTGAAAGATCCACATCTGGTGAATGTTAATGATGGAGCATAGAAAAG

WO 02/00860

20/68

PCT/US01/20171

FIG. 1T

ATGGAGGGGAGAAATTAACAACCTAGGGAATTGAGGAAGATCTAGAAACCTTCTAAGTCTCAGACACGTAACCCCAAG
TGTGGCAAGGAACCTCAATTGCTCTGAAATGCAATATATGTTGGGTAATTTCCAAATTTGAGGAGTGGTGTATTGA
CTACTGTTCAAGTTCAGGCTTCTTCAAGAGCTGGTAGATCGGGAATTTCCAAATTTGAGGAGTGGTGTATTGA
AATGCTGAGATCGGTAAGCAAGTGTGGATAGGAAGAAACGAAACACGAAAGAGTGAAGAGCTGTACCAAA
GGTTACCTGCTGAGCTGAGCCCTATGAGTTTGCTCTCTGGAATGGCTGCAAAAGTGTGGATGAATCAACAC
CTACCAACCTTATGATAATCACGCTTGCCTGTGTTCCATGACAACTTCAACCGGATCAAAATATCAATTATGAA
GAGGATATCTGAATGCACTGACATTTTCTATAGTAGATGAGGAGGTCCAAAGTAACCTGTAAGGCGCT
GTGAAGGAATGTAGTAGAACGTTGTCGCATATTGCGTCTGAAGAACCACTAAATGAAGATTATAAACTGTT
AATAATCTGCTGAAGCAGCAGTAAGGGCGATGGAATTTGGTGGGGAAGTCTCTTCGCGAGTTGGCGCA
GCTAGCTCTTGAACAGCTGGATGAGCAAGATGGTGTGATGCAAGAAACGAAACGGAAGATGAACGGTAGCACT
TAAATAAGATGAATCAAGGAAAGAAAGAAAGAGAGGAAATTAATTTTAAAGAGATATTCTGTGTCACCA
TGGTAGTTATGCATATCTGAAATGAAAGAGGCTTGTCTTAAAGGCTTGGAGCAACTCGACGACTACTT
TCCAAGGCTCCTGAGTTTCCAAGTTACAAGAGTGTCTTACAGTGCAAGATTTTAGAAAGAGAGGGGAAGA
AATGAAGCCTTACATAAGATGATTGCAACGACGAAAGACTTCTCTCCAAATTTGTTCCAGGATAAAACAGA
CCGTGCTCAGTAACCTGGCCGAGGATACGGATGCTCTACATGCTGTCTCAGTTCTCAGTTCTTGTAGGAAGTGGCG
GAAATTTGTTAGAAAGCCTACAGATGCAAGCCTGTGTCATCAGTTGGGAACAGTGTCTTTGTGCCCAACGG
GGCCTCATGTTTACATTTGCTTCCATGACCAAGAGATTTAAACTTATAGCTCTCATGCCCCAGTGAGTGG
CAATGATACAAAAGCTCTTTGTGGATCATGTAATTAATCACGAAATGAAGTGGGAGATGTAACCCCTT
CAGAAACAGTATATTTGAGCCCAACTCTGCCAATGCAAGAGGCTTATTGTGTCAGCAGAGGGG
ACCTGGTGAATACACTCAAGCCCACTCTATGTCATAAATTTGGGATAATAAAAGGTGATGAAGGATTCTGG
CTCGGAGCTGAATGTGAGTAGTTCTGAAACAGAGGAGCAAGAAAGGCTAAACCCAGATGGAGAAAGAT
CCAGATTTTAATCAAGCAATGGTGAAACAAAGCGCAAGATATCCCATCAAAATATATAGCCTATCAAAAGC
AAGTTATTCGCCGAATGCGACATAGAAAAGTTCGTGGTGAAGACACTCTCTGTTTCTGCTATCAGACGT
TAAAGAAATGAAAATTCAGATCATGCAATTTTCAATTTGCTCTTTTGAACCAAAATTTGTCATTTGATGGAAG
AATTAAGTGAATGACTGTGCCACCTAGCCCTTGGCGTCAATTCGATCTGCAATTTTATTAAGGCTGATG
AACCAATTCAGATTATGCTGCAATGGATGATGTCATGCAAGTTTGTATGCCAGAAAGGGTTTAAAGGTTACTG
GTCCTCTTGGACATTAA

>SGPR503_SEQID_17
ATGCTGAGCTCCCGGCCGAGCGCGATGACCGCGCCGACAGGGCCATCCAGCGCTTCTGCGGACCGGG
GCGCGCTCAGATATAAAGTCAATGAAAGACTGGGAGTTATAGGTGGAATTGCTGCTGCTTTCGAGCAGGAAAT
ATATGTTATTTGGGGTCCCATACAGAAAAAGAAAGCGTAGAAAAAGGGCTTGTGCGCTGGCCTGTGTTAATTTAGG

WO 02/00860

21/68

PCT/US01/20171

FIG. 1U

GAACACCTGCTTATGAACCTCCCTGCTACAAAGGCTGCTGCTGCTTTCATCAGGTGGGTGAAGAGTT
CACCTCCCACTCTCAGGATCAGAAGAGCCCTCCCTCAGCAGTATTTATCTTAACTTACACTCTTGCACCTTCT
GAAAGCTTGTCTCGCAAGAAGTTACTGATGATGAGTCTTACATGCAAGCTGCTTGGATGCTCTTAAGAAT
GTACAGATGGCAGATCTCATTTTGAAGAACAGGATGCTCACAAATATTCATGTCTTACCTCGTCTGGA
GATGAGGAGAGCCGCTCGGCTGACACATTTTGTGATGATTCCTGGAGCAGCAGTCAAGAAATAC
TCCCAACAATTTACCTGCCACAAGAGGCTACCTACCCACCTCACTCACTCACTCACTCACTCACTT
TCATGGAAGACTCAGTAGTATATGCTGTCGAACACTGTGAACACAGAGTCTGTTGGAATTTGATACCTTTGAT
AGCCTTTCACCTAAGTATTCAGCCGCCACATGGGTGACCAATTGACCTGACCACTGCTTACCACTTCATC
TCATCAGAACTCAGTGGGGGATGTTGTGTGACAACTGTACAAAGATTGAAGCCAAAGGAAGTTGAACGGGA
AAAGGTGGAACACCAAGAGGACCACTTTTGTAAACAGTTAAACTAGGGAAGCTCCCTCAGTGTCTCTGCAATCCA
CCTACAGCGGCTGAGCTGTGTCAGCCAGCGCAGCGCTCTGAAGCGGCAATGACAGCTGCAAGTTCAATGAGTTT
CTGATGATGGACATTTACAAGTACCACTCTCTTGGACATAAACCCTAGTCAACAACCCCTAACTGAACAAGAAC
CCAGGCTTACACTGAGCTGACAGATGGCCGGGAGGCCCCACACCACTTCTGATCAGCCAGGGGCCCC
AAACACAGATTTTATGAATGGCGCTGCTCCCAATCTTTATGCCAACGCTGTGAGCCGATGCCCTTCCCT
CTCCAGTTGTTCCGACTACAGCTCTCCACATACCTCTTCCGGCTGATGGCAGTTGTGCTCCACCAATGGAGA
CATGCACTCTGACACTTTGTCACTTACGAGCTGCCACCTTCTGCCAGAACCTCTCTCAACTAGCAATCA
GTGCTGTGGGTCTCCGATGACACTGTCCGCAAGGCCAGCTCAGGAGGTCTCTGCTCCAGCGCTACCTG
CTGTTCTACGAGCGGCTCTTCCAGGATGACAGCACCAGAGCCAGAGTGCAGATCTGAAGAATGA

>SGPR427_SEQID_18
ATGGAACCTGGGCCCCGGGACGCGCAGGAGGGGACGCTCGCGCCCGCCCGCCCGCGCGCTCCCT
GCGCCGCTGTTCAAGCGCTTCTGCTGGCGCTGGGCAGCGGCTCACSCCCCGGGGACTCACCCCGCGGCC
CCAGCGGGACACTGTGATGCGGACGGTGAGGGGGCTTCGCTGCGCCCGGGGCCAGTTCCAGCGGCCCC
GGGAGCCCCGGGAGAGAAAGCCCGCCGGACCCCAAGCTCCAGCTCCCGCCCGCGGATGGGGCG
GGCGCCGGGCTGAGGCTTTGAAGAACACGGCAACACTGTTTATGAACGCGGTGGTGCAGTGTCTCAG
CAACACCGACTCTGCGGAGTTCTGGCGCTGGCGGCTTACCGGCGCTCCGGCGCGCGCGAGGTAC
CGAGCAGCTGGCGGCTGGTGGCGGCTCTGGACTCGCGAATACAGCCCACTTTCGCGGAGTTCAAG
AATGCAGTTCCAACTACGGCTCTCAGTTCCAAAGCAATTCAGCAGCAGCCCTGGAATTCCTGCTGTTG
CTGGATCGTGTACATGAGGACCTGGAGGTTTCATCCCGAGGCGCGGTGTCGGAGAGCTTCGCGCTGAAGCCA
CTAAACCTCTGAGAACTGCCCTGTCACCATCAGCTTCTGCTAGGTCAAGCTTTGTGCAAGCCACTTC
AAGCAATATAGATCTTCTTGTGCTGCTCCACTGCTGAAACAGAGCAACACTTGTATCTTCTGCTGTGT
GTCCCTACTATCCCTTGGCGCAGAGAGGTTCTTGAAGTGTACCTTGGTCTTCCCTCTAAGAGCCAGCGGT

FIG. 1V

TCCCTGGGGTTGGCCCTGGCCGATGCGATCTCCCTCAGCAGACAGTGCGCAGCCCTGAGGAAGATGGTTGCGAGGAAGG
 AGCGCTCCCTGCAGATGAGTGATCTTGGTTGAAGTATCCAGTGAGTATCCAGCGGCTTTCTTTGATGAAGA
 GGACCTGATATACCATCGCAGAGGAGATAATGTGTATGCCCTTCAAGTTCCCTCCCTCACCAGCCAGGGGACTC
 TCTCAGCTCATCCACTGGGTCTGTGGCCCTCCCCACGCTGGCAGCCCTGAGGCCAGCGATCTCCCTCTCT
 CTCACAGTGAGAGCAAGTGCTAATCTCTTTCTGTAACTTGGGGGTCAGGCCAGCGCTAGCAGGTTTGG
 GCCACCTTCTCTGATAAGGAAGACAGAGCTTTCTCTGGCCCACTCCAGCATCTATCTCAGCAAGTCC
 GCCATCTATGAAGAGTGAGGCCCTGTACAGAACTGGGGTCTGTCTCCATCCCGTGTGGGACTCTCT
 GTGGCTGCAAGCTATTTGTCTCCGAGGACAGTGGCCCTCTGTCACTGGGAGTTGACAGGTTTGGCATCT
 CAGGAGCCAGGAGGCCCTCCACATGTCAAGCTGGCGTGGAGTGGGATAGCTCTGTCAAGGAGGCCCTGTTC
 GGGAGCCTCCAGGAGGAGCGCAGGATGCCGACAGTGTGGCAGCAGCAGCGCGCATCAGCAGCAG
 AGCTGTACCTTGGATGAATGTTTTAGTTTACACAGGAGGAGGAGCAGCTGCCCGAGGATGACGCTGGAAATG
 TCCCTCACTGCCAAGTCTCGCAGAGGGATGGTGAAGCTGAGTTTGTGGACGCTGCCCTGACATCTCATCATCC
 ACCTCAAAAGTTTCTGCCAGTGGCGAGAGAAACAAGCTCTCCACGCTGGTGAAGTTCCGCTCTCTGGA
 CTCACATGGCTCCCATGTGGCCAGAGAACAGCCCTGAGGCGAGGACTGGGCCCTGGCCCTTCTCTGGA
 AGCAGCCGGACTGGCTGCCACAGTTACCCGCTGGACTTCTCTAGCAGCTGTATGCCGCTGCAACCCACCAT
 GGCACTCGAAGTGCGCATTAACAGCTACTCCGGAACCTCTCTGAGTGGCAGTGGTACAGTTATGATGA
 CAGCACGTGGAAACCGCTTCGAGAAGATGAGGTCAACACAGAGGGGCTTATCTCTGTCTATCAGAAGCGGA
 ACAGATCCCTCCCTGGTCAGCCAGCAGCTCCATGAGAGGCTTACCAGCTCTCCCTCTCTGATCACTGGCTC
 TTACGGCTCGGAGCCACGCTGGCAGCAAGGGGAAGCCTGCTGTCTGTGAGGCTGGCCCTTGGCCCTCC
 TGCCCGAGTTCCCTGACTCTCCCATCTTACCAACAGCCTCTGCAATCAGGAAGGGAGGGTTGAGGCCAGG
 CGTTTGTACGGGCGTGAAGGCAAGCATTAGCATGAAGCAGCCACCTTCCGAGCCAAAGCAGGGAC
 CATTCAAGACCATGCTCTCGGTGGTCTTGGATCCAGGAGAAACCCAGGTCCTCCGCTCGAGTTGGTG
 GAGTACTTGAATCCAGACGAGACTCGGTCCAGAGCCAGTCCATTGTGCTGTTGACGGGCACTGGCG
 GTGAGGATGAGAAATCAGCATGGCGAGGTCCAGCTGGCCCTTCTGTCTAACGCGAAGATGGTGGGGGGG
 CATTGAAAGAGGTCCAGCGGGGTGCCCTGTCCCTGGGCTCAACCCCAACCATGTCTGGCCCTGGAACTCA
 GATGTTCCAAACACAGCAAGGAACTCAAGGAAATGCAAGGAGCATCAAGCTTCCGAGAAATTTGACCT
 GCCTCTACTGTGATGCCCTTCAAGTGAAGCATGAGAAACAGCTCGACCGAGGCCAAGGCCAATGAACCTG
 AAGGAGCTTCCAGATGGGAAGCAAAAGCAGCCACCTCCCTCTATATGGATCTCTGGAACAGCAAGA
 CAGTCCCGAGGCACTCTGAGCTAGACAGCCCTTGCAGGGGAGACTCACCTTCTGAGGTCCGTTTCCG
 AAGAAAGGAGACAGGAGGATGAGAGGGCAGGCTCTCCAGAGTGGCCCGCTCTCCCTGGTGGTGGCG
 GGCTGAGCCCTGCTGAGCGGAGGCTCCAGGCTCACCTCTGCTCCCTCAGATCCAGAGGCGCTCGGCCA
 GGGGCCCTGGGCGAGCGGGCTCGAGAGGGATGTCTGCTCAGCCCGCCAGCTCTCTCCGGCTCCCTCGTAAAGCCA

WO 02/00860

23/68

PCT/US01/20171

FIG. 1W

GCAGGCCCCGAGAGGCGAGTGCACTGGGCATGTGCAAAAGGAGCTGTTCCAGGGGAGCAGGCTTCTTATGGCAC
CTTTCAGAGAGTCAAAATATCACACTCTTCTTTAGGTCGAAAGAAAACCTTACCGGAGTCCAGCTTTTGA

>SGPR082_SEQID_19

ATGCACTCGTGCATCTTAAGAGTTACTATCTTCTTGCCCTGGTGTTCGCCCTTCAGTGCCTGCTGCGAGAC
CATAAAGGATGGACTTTGTTGAGGGCTATTCCATCAATTTTCCGACCAAGAGAGTCCGCACTCTTACC
CAGGAGACACAACACAGACTCTCGCAACAATTCATCGGAATGGGACAGACCTACTTTGACATGCAGATGCATGC
TCTGTACACACAGCCCCACTGTGGGTGCCTGATGGGTCCGACACCTCCATCTCGCCAGGAAGATGCAAGTGG
AATAAGCACACTCTAACTTACAGGATTATCAATTACCCACATGATATAGCCATCCGCAAGTGAAGACAGTATAT
ATAATGCAAGTTTCCATCTGGAGCAATGTGACCCCTTTGATATTCACAGCAAGTGGAGATGGAGATGCGAGACATCA
AGGTTTCTTCTGGCAGTGGGCCCATGAAGATGGTTGGCCCTTTGATGGGCCAGGTGGTATCTTAGGCCATGCC
TTTTACCAATCTGGAATCTGGAGTTGTCCATTTTGACAGAAATGAACACTGGTGCAGCTTCAGACACTGGAT
ATAATCTGTTCCCTGGTTCGACTCATGATGGGCATCTTTGGCCCTGCAGCACTTCGGAAATCAGAGCTCCA
TAATGTACCCACTTACTGGTATCAGGACCCCTAGAACCCTTCCAGCTCAGTGGCCGATGATATCCAAAGGATCCAGC
ATTTGTATGGAGAAAATGTTTCATCTGACATACCTTAA

>SGPR359_SEQID_20

ATGAAGGTGCTCCCTGCATCTGGCTTGCTGTCTCCTCATCATGGCTTTGAAGTTTCCACTGCAGCCCCCTCC
CTAGTTGCAGCCTCCCCAGGACCTGGAGGAACAACCTACCGCCTCGACAGCGGTATCTTGACAAATATTACAC
AAATAAGAGAGGACACCGAATTTGGTGAATGGTTGCAAGAGGAGCAATTCATGATAGGAGGATTAAGGAGCT
ACAAGCTTCTTTGGCTCCAGTCAACCGGAAGTTAGACGACACCAATGAAGCTGATCAAGAGCCCTCGCT
GTGGAGTCCCTGATGTGGCCAAATTATCGCCTCTTCCCTGGTGAACCCAAATGGAAGAAAATACTTTGACATACA
GAATATCTAAATACACACCTTCCATGAGTTCTGCGAGGTGACAAGCAAGTGGAGATGGCCTTGCAGGCTGG
AGTAGCCGCTCCCTCTGAGCTTTGTGAGAAATAACTCAGGAGAGCGGATATATGATATCTTTTGAATGGA
GATACGGGGAATTCCTATCCATTCTGATGGGCCCTCGGGGACTCTAGCCCATGCAATTTGCTCTGGAGAGGCT
GGGAGGAGATACATTTTCGCAATCTGAGAGTGGACTATGGGAACGAATGGTTTTAAATTTTACCGTTGC
TGCTCATGAAATTTGGCCATGCCCTGGCCCTGGCCATTCACAGACCCATCAGCACTGATGTACCAACTTAA
GTACAAGAAATCCCTATGATTCACCTCCCAAGATGATGTGAAGGGATCCAGGCATATACGACCTCGSAA
AGTATTCCTGGGGAGCCCACTCTGCCCATGCCCCCATCAAGCCCATCCCTGACCTCTGTGACCTCCA
GCTATCCTTTGACGCTGTGACAAATGCTGGGGAAGAGCTCTGCTCTTCAAGGACCGGATTTCTGGAGACGG
CAGGTTCACTTGGCGACAGGAATTCGGCCAGCACTATTACCACTCTTCCCAAGCTCATGTCCAATGTGGA
TGACGCTTACGAAGTGGCTGAGAGGGGCACTGCTTACTTCTTCAAGGTGCCCACTACTGGATAACAAGAGAT

WO 02/00860

24/68

PCT/US01/20171

FIG. 1X

TCCAAATGCAAGTCCCTCGGACTATTATGACTTTGGATTCCAAAGCAGCTGCAAGCAATAGATGCTGCTG
TCTACCTCAGGGAGCCACAGAAGACCTTTCTTTGTGGAGATGAATACTACAGCTACGCAAGGAAAGG
AAATGGAAAGAGCTATCCAAGAACTACTGAAGAAGATTTTCAGGAGTAAATGGCAATCGATGCTGCTGTA
GAATTAATGGCTACATTTACTTCTTTTCAGGACCAAAACATACAAAGTATGACACAGAGAAGGAAAGATGTGGTTA
GTGTGTGAATCTAGTTCCTTGGATTGGTTGCTAA

>SGPR104_1_SEQID_21
ATGAACGTGCGCTGCAGAGCTGGAGCTGGCAGCAACATGTTGGAGTACAAACGGGCCACGCTTCGGGGATG
AAGACGACCCGAGACCCCGTAGAGGGGGGGGCTCCCGGACGCCATGGAGGTGGGATTTCCAGAGGGGA
CAAGACAGCTTTAGGCTACGACAGCTGGAGCTGGTCTTAGCAGGTGCTCTCTACTGCTGGTGCACATG
CTTCTGGGCTGCTTTGTGGCCTAGGGTCCAGTACCACAGACGATCCACACGACCTGCCCTTACAGAGG
CCTGCAATCGAGTGGCTGGAAATCCTGGAGTCCCTGGAGCGGGGTGAGCCCTGTGAGGACTTTTACCA
GTTCTCTGTGGGGCTGGATTGAGAGAACCCCTGCCGATGGCGTCTCCTGGAACCTTCAACAGC
CTCTGGGACCAAAACCGGCATCTGAAGCACCTGCTTGAACACCACTTCACTCCAGCAGTGAAGCTGA
GCAGAGACACAGCCTTCTACCTATCTGCTACAGGTGGAGCGCATTTAGAGGCTGGGAGCCGACCTG
AGAGCCTCATTTGAGAAGATTGGTGGTTGGAACATTACGGGCCCTGGACAGGCAACTTTATGGAGTGT
GAAGCGATTAGCAGGACCTACAGGCCACCCATCTTCCCGTCTACATCAGTCTGCTGCTTAAGAGTTCCA
ACAGCAATGTTATCCAGGTGGACGATCTGGGCTTTCTGCCCTCTCGGGATTACTACTTAACAGAACTGCCA
ATGAGAAAGTCTCACTGCTATCTGGATTACATGGAGAACTGGGATGCTGCTGGGTGGGGGCCACCTC
CACGAGGAGCAGATGCAAGGTGCTGGAGTTGGAGATACAGCTGCCAACATACAGTGGCCAGGACCCAG
CGCGGACGAGAGAGATCTACCAAGATGAGCATTTCCGAGCTGCAAGCTCTGCGCCCTCCATGGACT
GGCTTGAGTTCTGCTTTCTGCTGTCACCATGGAGTTGAGTCTGAGCTCTGAGCTGTGGTGTGTGGGATG
ATTATTTGACAGAGTGTGAGCTCATCAACCGACGGAACCAAGCATCCTGAACAATTACCTGATCTGGAACC
TGGTGAAAGAACCTCAAGCCTGGACGAGCTTTGAGTCTGCACAGAGAGCTGCTGGAGACCTCTAT
GGCACTAAGAAGTCTGTGTCGAGGTGGCAGACCTGCACTCCACACGATGACGCCCTTGGCTTTGCTTT
GGGTCCCTCTTGTGAAGCCAGTTTACCGGCAAGCAAAATTCAGAGGGGATGATCAGCGAAATCC
GGACCGCATTTGAGGAGCCCTGGACAGCTGGTTTGGATGGATGAGAAGCCCGCCAGCCAGGAGAA
AGCAGATGCCATCTATGATGATTTGTTTCCAGACTTTATCTCGAGCCCAAGAGCTGGATGATGTTATGA
CGGGTACGAAATTTCTGAAGATTTCTTTTCCAAACATTTGAAATTTGTAACTTCTCTGCGCAAGGTTATGGCT
GACCACTCCGACGCTCCCGAGCAGATGAGGATGAGCAGTACCCCGCAGACAGTGAATGCTACTACTTTC
CAACTAAGATGAGATGCTTCCCGCTGGCATCTGCGGCCCTTCTATGCCGCAACCCCAAGGCC
CTGAACCTCGGTGGCATCGGTGTGGTCAATGGGCCATGAGTTGACGCAATGCCCTTGTATGACCAAGGGCGCGATG

TGACAAGAAGGGAAACCTCGCGCCCTGGTGCGAGAATTGAGTCCCTGGCAGCCTTCCTCGGAACACACACGGCCGTGG
ATGAGGAGCAACGATTAATCAATCCAGGTCATCGGGAGAGCTCAACGGCCGCGAGACGTCTCGGAAACACACACGGCCGTGG
TTGCTGAAACCGGGGGGTGAAGGCTCGCTTACAAATGCTTACAAGAAGTGGCTGAGAAAGCATGCTGGGGAGAGCA
GCAACTCGCAGCGTGGGGCTACCAAGGACACAGCTCTTGCTGGGAAATTCGCCAAGTGTGCTGCTGCTGGGC
CGCACACAGAGAGCTTCACAGACCGGCTGGTAGCGACCCACAGAGCCGCGCTTCGCGGCTGCTGGGC
ACTCTCTCCAACTCCCGTGACTTCCTGCGGGCACTTGCGCTGCCCTGCTGCGCTCCGCCATGAAACCCAGGCGACGT
GTGTGAGGCTGTGGTAG

>SPR303.1.5E01D_22
ATGCGGAGAGAAGGGCCCTTCGAGCGGCTGCTCGCGATGTCTCCCCATCAACTGCAGCGCTTTGCTCTCAAGC
CGCACTTGCTGGACTTCACTTCGAGGCGCAAGCTGAGGCGCGCCAGGTGAGCGAGCGCACTAAATCAGAT
TGATGTAATTGTGCTGATATGATATATACAGCTTCATATGCGACAGCAAGGAGATGAAGAATACATCGTACA
GGATTTAATCTCAGATGAAGATGAAGAAAGTCACTCTTGCTCCATGATCGCACAGCAAGTACAGCGAACCT
TAAGATAGATTTTGTGGAGTGAACATGAAGAAAGTTGTATAGAAGTATGATATACCCCTCTGTG
AGAGCGGCTATGTGCGCTGAACAGTTGTAGGCTGAGCTGCGCGAAGGCTTTCTCTCGTGGAATGAGC
TGCTATCAAGCAACTTTGATCTCATGGTGTCTTCAAGACAGATAGCTTTTACAAACATGAATGTAATT
GACCGGAACCAACCCCTGATGATTGAATTTAGTGAAGTGAGTTTGCCCGCACACCTGTATACATCATCAT
CTGGTGGCATTGTGTGGGTGAATATGACTTTGTAGAACAGGTCAAAAGATGGTGTGTGTGTGTTTAC
ACTCTGTGGCAAGCAGAACAGGAATAATGCATTAGAGTTGTGCTTAAACCTTGCCCTTTTATAACGACT
ACTCAATGCTCTTATCTCTACCTAAATATGATCTGCTATGCAAGACTTGCACTGCTGCTTATGAGAA
CTGGAGCTGTTTACTATTAGGAGAGCTCAATGTTATGATCCAGAGCTTGCTTCTTATCCCGGCTGAGTGG
GTGCTCTGGTTGTGGCAATGAACCTGGCCATCAATGTTTGGAAATCTTGTTACTATGSAATGTTGTSAGACTCAT
CTTGGTAAATGAAGTCTTGATCTCGGTGAATATCTGTTGTAGACCACTGCTTCCCAAGATGATGATTT
GGACTTAATTTGTTCTTGCTATACCCGTCGCCAGGATGCTGACCGCTTAGATGAAGGCTCATCTGTGAAG
TGAGTGTGGGCTCATCCACTGAGGTTGATGAGATATTTGATGCTATCATATGCAAAAGGTGCACTGTGTCATCC
GAATGCTGCATGACTAATTTGGGAATAAGTAAAAAAAACCTTTAAGTATT

>SPR402.1.5E01D_23
ATGCGGCGCGCCGCTGATTCGCTGTGGCTGCGCTGTGCTTTGGCCCTTGCCGCCCCCCGGGCTGTG
GGGTGGCCCCCGTCCGAGCCCCCATCTATGTACAGACGTGGGCGCTGCAGGTGTCCAGGGTAAACCGGGAG
TGAGCGGCTGTGGACAGCAATTTGGCTTGCTCACTCGGGGCGATCTTCCCTGACGGGCACTATTTCACTC
GGCAGCGGGGGTGGTGCAGACGTCTGACCCCGCACTGGGCGCACCATCTGTCACCTGAGAAAGAAACCC
ATGCGGCGCGCCGCTGATTCGCTGTGGCTGCGCTGTGCTTTGGCCCTTGCCGCCCCCCGGGCTGTG

[illegible]

FIG. 1AA

CCNAGTCCAGGGGAAACGAGTGGGCCCGGCTTGCGGTACTTCCCTAGCCTGTGGCCAGCGGACTTCAAG
GATACTGGAGAGTTCGGCGCTCCAGACGAGAAAGTGCCTGGCAGGTGAGCTCGACACGCAACAGACAGC
ATCGGAGCGCTCCCTTCTCCAGACATGGGTTTAAAGAGACACCAACCTACGATTTGGTGGAGCC
TGGTGGTGGAGAGTATGGAAATGACCGGCCATTTGGCTCTGGAGAGTGGCTCGGATGGAAATGAGCAGT
TTGTGGAGGGCGCCTTATGACCCGAGCTGGGTGGTACTGGCGCCCACTCAGCAAGGCACCAAGAGTA
CTAGTGGTGGTCTGGACATCCAACTCGACGCCATGAATTCAGCAGGCGCCTCTGGTCCCTGTGAGGGAC
ATCATTTGACCCCAAGTACTGCGCGCGGCGCTTCATCTCATGGGTGACCTTGGCTCTCCACTTCGAACACT
GTACACTTTGATGAGTACTGCGGCCCTGACGCCATTCGCTCGGGAGCCCAATTCACACTGAAGTTGGAGCGCAGTG
TTGGTGGAGCTGTGGAGCAGGTTAAGCAGCGCTTTGAGGCTCGGCATTCACACTGCATGCTGACCCAGCAGC
TGCAAGAGCGCTGAGGTGTTATTGTCACCAACAGAGTGTGACCGGCATTCACAAAGATCTCTTCTTCCGCCCA
GTTGCCCCCTTGTCTGGCGGACATGATCTGCGCCACCAATTATGGGGAACACTTGCTATGGGATCTGTG
AGGCGCATTGGCTGTGAAGTTGAGGCGCAGATGGATCTGGCTGGGGTTGTCTGGGGAACAAATGACCTGGTG
AAGCAGCAATACAGGTGTGTCAGGCCGATACCAACATGGATGAAGAAGAAATGACCAATGG
AGGCTTGTCAAGTCCCTGTGGCTCTGGCTCTCTTCTGTGCTGGCGCTGCAAGCCCGCAGATGGGCTGCT
GA

>SGPR446.1_SEQID_25

ATCTCACCACGATGTGTGGCGGAACCCCTCTGAGAACTGTGGGAGAGTGGACGCGGAGGAAGGGAGGTGG
CTGCGAGGTGAGGTGAGGACCAAGACGGGACACATCTCGCGGGCACATCTGCGCGCCAGTGGACCTGGGTGG
CTGACGCGGCGCACCTGCAATCCAGCGGCTTCCATACGTGTGCAAGATGGGAGTGTGGAGTGTCTATATGGA
ACACAGATGTGGTGTCTGACCCAAAGAGCTTTTGTCCACCTTAAGTCTCAACGATTCACACCATTCGAAAT
GACCTTGGCTTCTCCAGCTCTCAACTCTGTGAATTTCTCTCAAACTCCAGCTATCTGCACTCCCTCAGGAG
AATTTCCAGTGGGAGGTAGGACGAGTGTGTGACCGGATGGGGCAACACACCAAGCTGGGAGAGAAC
TTGCATCAGAAATCTTCAGGATGTGGACCAATGATGCTATGAGGAATGTAATACAGATATACAGAGGC
CTTGTGATCTAAGGATGTAATATAAGGGAATGGTCTGTGGCTTAAAGACGAGGAAGAGGATTTCTGTGCA
GGAGATCTGTGGGGCGCCTTGGCCTGTGGAATATAATGACACATGGGGTCCAGGTATGGAGTGTGGAGGCA
TCGGCTGTGGTTCG

>SGPR447_SEQID_26

ATGGCGGCGCGCGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCTGGACTCGGGGAAGCGCGGACCTG
CGCCACCTGGCGGAAATATCACGGCTGTGGCGGGCGGAGTGGAGTCCGCGCGCGGCGCTGGCCATGCGAG
GAGCTGTGGCTTAGGAGAGACGCGCAGTGTGGAGGAGCGGCTGTACGCGCGCGCTGTGGCTCTCGGCTG

WO 02/00860

28/68

PCT/US01/20171

FIG. 1BB

CGCACTGCTTCCAAAAGCACTACTATCCCTCCGAGTGGACGGTCCAGCTGGGGAGAGTGACTTCCAGGCCAACT
CCTTGAACCTCGGGCTACAGCAGTCGTTACAAGTGCAGACATCAATTGTAAACCCCTGACGCACTTGGGGT
TTACGCAATGACATTGCCCTGCTGAGACTGGCCTCTTGTACCTTACATGGGTACATCCAGCCCATTTGCAAT
CSAGTCTCCACCTTCAACTTGTGACACCGCCGAGCTGGGTGACCGGCTGGGGTTAATCAGCCCACT
GGACACCTCTGCCACCTCTTAAACCTCGGGAGACAGTACCACCTTAAACAAACACAGGTTAAATAC
CTGTTTGAACACCCCTTAGCCGTAGTATGATTCGGAATTCATTTTGTGCTGGTGGGATGGCACTGTA
GACACTGCAAGGTGACTCAGGTGGACCTTGGTCTGTGACAAGTATGACTGTGGTATCAGTTGGAATCGT
GAGCTGGGAATGGAATGCACTGCGGTCAACCAATCGGCCCTGGTGTACACCAACATCAGTGTGACTTCCACTGGA
TCCGGAGGGTGAATGTCCACAGTACACCCAGGCCAAACCCCTCCAGCTGTTGCTGCTCTTGGCCTGCTGTGG
GCTCCG

>SGPR432_1_SEQID_27
ATGGGCAACCTGGGGGAGCCCTGCTGGGTGGGCTCGCTCTTTGCTTGACGGGCTTAGTCTCGCTCT
ACGGCTGACAGCTGAAGCGGGCGCGCCCGGACCGGGATTACCGCGCGCTCTGCGACGTGGGCAACCGCC
ATCAGCTGTTCCGGCTCTTCTCTCCAGGTGGGCGAGGGTTTCGGGCTGGTGGAGCATTTGCTGGGACAGG
ACAGCATCTCAATCAATCCAAACAGCATATCGGTTGCATCTTACACACTACAGCTATTGTAGTCTTCAAGC
CGCTCAGCGTGCCTGTGGACAGCTGGCCCGCCCGCCCTCAGGAGGGCAACAGTCCCTGGCGA
GTGGCCTGGCAGGCCAGTGTGAGAGGCAAGAGCCCACTCTGCAGGGCTCCCTGTGGCAGACACCTG
GGTCTCCTGCTGCCACCTGCTTTGAAAAGGCAAGCAAGAACTGAATTCCTGGTCACTGGTCTGCGGCTT
CTTGCAGCGTGAAGGACTCAGCCCTGGGGCCGAAAGGTGGGGTGGCTGGCTGGCATTTGCCAGGGCT
ATAACCACTACAGCAGGGCTCAGACCTGGCCCTGCTGAGCTGGCCACCCACAGCCACACCCCTCTG
CCTGCCACAGCCGCCATCGCTTCCCTTTGGAGCCTCCTGCTGGGCCACTGGCTGGGATCAGGACCCAGT
GATGCTCTGGGACCTACGCAATCTGGCCCTGCTCTCATCATGTCGCCCAATGTAATCTATCTACACCA
GCTGACCAAGCAGACCTGTCCAAACCGCCCGCCCTGGGATGCTATGTGGGGGCCCGCCAGCCCTGGGGTGCA
GGGCCCTGTACAGGAGATTCCGGGGCCCTGTGCTGCTGAGCCCTGACGACATGGGTTTCAAGCTGG
CATATCAGCTTTGATCAAGCTGTGCCAGAGGAGCTCTGCTGCTGACCAACAGCTGCTCAGATT
CCTGGCTGCAAGGCTCGAGTTTCAAGGGGCAAGCTTTCCTGGCCAGAGCCAGACCCCGGAGATGAGTGATGA
GGACAGCTGTAGCCTGTGATCCTTGAGACAGCAGGTCCCGCAGCAGAGCAGCCCTCCCGATGGCCCTGG
GAGGCCAGGCTGATGCAACAGGAGAGCTGGCCCTGTGGCGGAGCCCTGGTGTGAGGAGGCGGTTGCTAACT
GCTGCCACTGCTTCAATTGGGGCCAGGCCCGCAGGAAATGAGAGCTGGGCTGGGACCGAGCGGAG
TGGGGCTGAAGCAGCTCATCTGATGGAGCTACCCCACTCAGCCCGCTAGGGGGGCTACGACATGGCCCTCTGC
TGCTGGCCCAAGCCTGTGACACTGGGAGCCAGCCTGGCGCCCTCTGCTGCTGCTATCTGACCAACACCTGCC

WO 02/00860

29/68

PCT/US01/20171

FIG. 1CC

TGATGGGGAGCGTGGGCTGGGTTCTGGGACGGGGCCGCCAGAGCATAGCTCCCTCCAGACAGTGCC
CGTGACCCCTCTGGGGCTAGGGCTCGACGGCTGCAGCTGCAGCTCTGGGGTGATGGAGCCCTATTCTG
CCGGGATGGTGTGTACAGTGTCTGGGTGAGCTGCCAGCTGTGAGGGCTGTCTGGGGCACCCACTGGTG
CATAGGTAGGGGCACATGTTCTTGGCCGGCTGCACAGCTTCGAGATGCTTGCCAAAGGCCGCCCAGG
CCGGGCTCTTACCGGCTCCCTGCTATGAGGACTGGGTCAAGCTTTGGACTGGCAGGCTCTACTTCGCCG
AGGAACAGAGCCGAGGCTGAGCCTGGAAGCTGCTGCGCAACATAAGCCAACCAACAGCTGCTGA

>SGPR529_SEQID_28

ATGAGAGCTCCGACCTCCACCTCTCCGCGCGCTCTGGCGCCCGGGCTCTGGCGAAGCTGCTGCCGCTGCTGA
TGGCGAACTCTGGGCGCGCAGAGCGCGCTGCTCCGCCAAGAGACACGGCTTGGACCCCGAAGCCTATG
GGCCCGCTGCGCGCGCTCGACCTCGCAGCTTGGAGTCTGCTCTTCAACGGCTCTGCTTCCACTGCGCGG
GTGCTGTGGACAGATTGGGTGCTGACGGCGGCTGCGGAAACAGCCACTGGGCTCGAGTAG
GGATACACACTGCTCTTCTTCAAGCGAGCAGCTCGCCGACGACTGCTGTGTGCTTCCATCCCAAGTAC
CACAGGGCTCAGGCGCCATCTGCCAAGCGAAGCATGAGACAGTCTCATGTGCTAAAGCTGGCCAGGC
CCGTAGTGGCGGCGCCCGCTCGGGCCCTGCAAGTTCCTACCGCTGTGCTAGCCCGGAGACCAAGTCC
AGGTTGCTGGCTGGGCGACACGGCGCGGAGTGAAGTACAAAGGCTGACCTGCTCCAGCATCAC
TATCTGAGCCCTAAAGAGTGTGAGGTCTTCTACCTGGCGTGGTACCAACAACATGATGTCTGGACTGG
ACCGGGCCAGGACCTTGCCAGAGTGAAGTGGAGGCCCTGCTGTGACGAGACCTTCCAGGCATCCT
CTGCTGGGGTGTTTACCCCTGTGGCTCTGCCGAGCATCCAGCTGTCTACACCGAGATCTGCAAAATACATGTCT
GGATCAATAAAGTCAACGCTCCACTGA

>SGPR428_1_SEQID_29

AAACGTACAATGTGGCCACAGGCTGCTTTCCAAACTCGTCATGTTACCATTTGCAACGGCTTCAAGTCC
AGAAATGGTGAGTGCCCGTGGCAAGTGAATCCAGATGTACGGAACACCTCTGTGGAGGCTCAATCTTACAT
TGGTGGTGGTCTTGACAGCCGCACTGCTCCGAAGAACCTATTAGACATGGCGGTGAATGTCACTGT
GGTATGGGAACGAAACATTACAACTCCAGTCCGAGAGAAAGTGCAAGGCTATTATTACAAATA
TTACAAACCGCCCGCTGACAGTGAACCTCTCTCTGCTTCTGCTTCCACACAGTSCAATTGAGCAATTCAA
AATGCTCTCTGCTGCGCTGCAAGGAGGAGGAGCTGGGAGTGGTGGATGGGACAGTGGGTAAAGCAAT
GGGTATGACCAATATGATGACTTAAACATGCACTGGAAAGCTGAGAGTGGTGCAATTAGCCGGAAGAAATG
TGCCAAAGGGTAAACCAAGCTGTCCAGGAACATGTTGTGCTTGGAAACCAACGACCAATGGGCGAGGC
CCAGGAGAGTGGGGGCTCTGTTTCCAGAAAAGAAACAGACATGGTACCAGCTGGGTATTATTCAG
CTGGGGTGTGGGCTGTGGCCAGAAACATGCTGGAGTGTACACCGAGTTGTCCAATTATCTGCTTTGGATCG

WO 02/00860

30/68

PCT/US01/20171

FIG. 1DD

AGAGGAAGACTGTGCTGGCAGGGAAAGCCGTATAAGTATAGCGCAGACTCTGTGTACGCTTTTGCTTCTCTCAACC
TGGGCCATCCTGTACTGTATTTTGTGATGCTTCTATTATCCT

>SGPR425_SEQID_30

ATGGAATATGCTGCTCTTGGTTGATATTTTCAACCCCTGGGTGGACCCCTCATTTGATGGATCTGAAATGAATGG
GATTTTATGTGGCACTTGAGAAAGGTACCCCGATTTGTCAAGTGAAGGACTTTCCATCTCACAGCCCGCATTT
GAGGCAGATGCTAAGATGATGTTAAATACAGTGTGGCATCGAATGCCAGAAAGAACTCCCACTCCCAAGCCT
TTCTGAATTGAGGATTATCTTCTCTATGAGACTGTCTTTGAGATGGCACCAGAACCTTAACCAAGGTGAAGTT
CAAGATTGGTTCTTGAGCCGACTCAAAATATCACCAAAAGGAGTATCTGTTAGGAAAGACAGAGGTGTAT
GGCACCGACAGCAGGTTTCAGCATCTTGGACAAAGGTTCTTAACCAATTTCCCTTTTCAGCACAGCTGTGAAGCTT
TCCAGGGCTGTAGTGGCTATCTCTTCCCTCAGCATGTTTAACTGAGCTGCCACTGTTCATGATGGAAG
GACTATGTCAAAGGGAGTAAAGCTAAGGTAGGTTGTTGAGATGAGGAATAAAGTGGAGGCAAGAAACG
TCGAGTTCTAAGAGGACAGGAGAGAGTGTGGTGGTGACCAAGAGAGGTTACCAGAGACATCTGCCG
GAGAGCGAAGGTTGGGAGAAAGAAATAATCTGGCCGGGTGAGAGATTGCCGAAGGGAGGCTTCTCT
TTAGTGGACCCGGGTCAAGATACCCACATTCGAAAGGCTGGGCGAGGAGCATGGGAGCGCTACCTT
GGACTATGACTATGCTCTCTGGAGTGAAGCGTCTCACAAAGAAATACATGGAATCGCCAAAC
GATCAAGAAATGCCGTGGTGAATGCCACTTCTCAGGATTTGATAACGATAGGCTGATCAGTTGCTCTATCG
GTTTTCAGTGTGCGACGAATCCAATGATCTCTTACCATACTGCGATGCTGAGTCGGGCTCCACCGGTTT
GGGGGTCTATCTGCGTCTGAAAGATCCAGACAAAGAAATGGAAAGCGCAAAATCATTCGGGTCTTACTCAGGCC
ACCAAGTGGGTGGATGTCCACGGGTTTCAGAGGACTACAACTGCTGTTTCGATCACTCCCTTAAATACGCC
CAGATTTGCCCTCTGGATTACCGGGAACGATGCCAAATTTGCTTACGGCTAA

>SGPF648_SEQID_31

ATGGGGACCCAGAAAGGACGACAGTGGGGTTGGGGAGAGGGGATACGGGTGTCAGAAAGAAATTTATTAA
CAGTGGATGGGATTAAGTCTGTGCTGGAGGATCCTGGTGGAGCGCAGAGGGTCTCGCTCACCTGGATTCTCT
CACTCCCTCCCGAGCTGACGCCGAACCTGTCTCTCTCCACATTTGGGTTCCTCTCTCTCTCTCTCTCTCT
CTGGCATCCACAGACGCCAGGATGGTGACAAGTTGCTGGAAGTACAGAGTGTGCAACCCACTCCACGCCAT
GGCAAGTGGCTCTCTACGAGCGTGGACGCTTTAACTGTGGCGCTTCCCTCATCTCCCAACACTGGGTGCTGTCT
GGCGCCACTGCCAAAGCCGCTTTCATGAGTGGCGCTGGGAGAGCAAACTGGCAAGCGCGATGGCCCA
GAGCACTACGACCGCTCTGGGTCTATTCACACCGCGCTACGAGCGCGAGCCACCCCAACGACATCA
TGTGCTGCGCTAGTCCAGCCCGCAGCTGAACCCCGGCTGCGCCCGCGGTGCTACCCACCGCTTGGCC
CCACCCGGGGAGGCTGTGTGGTGTCTGGCTGGGGCCCTGTGTGCCACACAGAGCTGGGACCGCTGGAG

FIG. 1EE

[illegible]

WO 02/00860

32/68

PCT/US01/20171

FIG. 1FF

ATGGCTGCTCAGCAGCTCAGCTGCTTTTGGAAAAATAAGACCCAACTCAATGGATTGCTACTTTTGGTGAAC
TATAACACCCCGCAGTGAAACGAAATGTGAGGAAATTTATCTTCATGAGAAATACCAAGAAACAAATGA
AATGACATTTGCTTTGGTTGAGCTCTCTACTGGAGTTGAGTTTCAATATATAGTCCAGAGTTTGGCTCCAGACT
CATCTATAAGTTGCCACCTAAACAAGTGTGTGTCACAGATTTGGATCCATTTGATAGATGTGACCTATACA
AAATACACTTCGGCAAGCCAGAGTGGAAACCATAGCACTGATGTGTGTAACAGAAAGGATGTGATGATGGCT
GATACTCCAGGAATGTTATGTGCTGGATTCATGGAAGGAAATAGATCATGTAAAGGAGATTTCTGGTGGACC
TCTGTTTATGATAATCATGACATCTGGTACATTTAGGTATAGTAAGTTGGGCAACATCATGTGCATTTCCCAA
AAACCTGGAGTCTACACCAGAGTAACCTAAGTATCGAGATTGGATTGGCTCAAGACTGGTATGTAG

>SGPR52_SEQID_34
AGAATAGCAGAGGGTCTGGATGCTGAGGAGGAGAAATGGCCCTGGCAAGCTAGCCCTCCACAGAAACAATGTCTA
CCGACCGGAGCCACATGGCTTAGTAAACAGCTGGCTTATCACTGCTCACTGCTTCAATAAGGGTCCATGATC
CCAAAGATGGAATGTTATTTAAGTAACCCACAAACACAGTCAATATCAAGAATGTTATAATCAAGAAACTAC
CATTAACCTTGCACACGATATGACATTTGCTTTGTGTCATCTCTTCCAGTGTATATACAAGCAACATCCAAA
AAGCATGCTTCCAGATGTTAATATATATTCCTATAACAATTCAGAAAGCAGTGTGTACTGATGGGATCATTTAA
CCTTTACGAACAACCTTTCTAATGTACTCCCAAGGGAATAGTGAAGATTATAGATAATAGGACCTGCAACAATGGG
GAGGCAGATGGCAGATCATCATCTGGAATGTTGTGCGGGTTCCTGGAGCCACGTTGGATGGCTGGC
AGGTTGACTCTGGTGGACCACTGGTTGGTACAGATTCTAAAGGCATCCTTGGTAAAGTTCCCTGCTGCTATTGA
AAGCTGGAGTAATGAACGTGCTCTTCCAAACAAGCCCTAGTGTCTACACTCAAGTGACATACTAT

>SGPR405_SEQID_35
ATGGTCAGCAAGGGGGAGTTGCTGCAGAGCCAGAGCCACACTATTGTGAGGACAGTGAAAGAGGCCCAACA
CCCTCAGAGTCCGGGSCAGCTTCTTAGAGAGGTGGCATTTAGGTGGGCATGGAGTTCCGGGATGCAGCGG
TGAAGGTGCGTGAAGCCCAATGAGAGGGCGCCCGGAGGGGGCGGCGAGGCAAGAGGGCTGTGCCGG
GACCAAGCCAGCGCAGCGGGTCAAGAGAGGGCTCGCGCGGGTGGACGCTGGAGCAGGAGCCAGG
GGAGATGCTTAGAGGATAAATAGCGGCGCATGAAGATCTCAGGCTGGCAACCAAGGAAAGCAGGC
TCCCAATAGACAGCAACACCTGAACCGGTATCAGCAGCTTCCCGTTAAGATCTCAGGAGCTGGCGAGGG
GGCTGAGCAGGACACTAAGAGGCAAAATGSCAGATTTAAGTGTCTATGGCCTTCAAGGACCTGCGGCT
GGTCATGAAGAGCGCTCAACTCTGTGTCAGAGGGGCAAGAGGCAATGGCTGGATGTGCTGCTGCTT
CTTCTGTTGGTTTGACCATTTCTAGACCTCCCGCCCGCTGAGGTCAGGAGGAGTTGAGGCTTGGC
GTTTAGAGTGCAGGTGGGGCAGCTGAGGCTCTATGAGGACACCGGCAAGGTGGTTGAGATCGTCCCGT
CACCCCAAGTACAACGAGAGCCTGTCTGCCCCAGGGGGTGGCGACATCGCCCTGCTGAAGCTGGAGGCCCG

GTGCCGCTGTCTGACATCATCCACCCGGTCTCGCTCCCGTCTCGCTTCCCTTGACGCTGCCCTCGGGGAAGACCT
GCTGGGTGACCGCGCTGGGTGTATTGGACGCTGAGAACTATGACCTTGGCCCTTGCCCTCTAGCTTGTGGAGGGGAG
GTGGAGCTGAGGAGGAGCGCTCTGTTCACACAGACTGTGCGCCCGGCTCTGCTCAACACACACTGAGCGT
TTGAGCGCTCATCAAGACGACATGCTGTGTGCGGGAGCGGAACCAACGGCTCTCGCCACGCGCAACAG
GGGCCCTCTCTGTACAGCGGAAATTGCACTTGGTCCAGGTGAGGTGTGTAGCTGGGCGCAACTCTGCG
GCCTTCCGGCTTCCGGCATGTACACCCGGTGACGATGCTGTCTGATCGCGAGCCATGCCCCCT
AGTACGACCCCTGGTGTGCGAAGATTGTCTCGCCCCAAATCAAGATTGTAAGCGCTTATGAGGACCT
TGATTGGATCAGGAGCGCGCTGCCCGCGCGGACCTGCAAGAGCGGAGTGGCCCTCATTCAGGAGCG
GAGCTTCGAGAGGATGTATCAAAAGGCCCACTGCCACGGCAGGTCAACATCATCAAGGCTGCCATGCC
GTGTGACGAGGAGGAGGCGGAGGGTTCTCGAGCGCGCTGTGAGAGCGGAGGACATGCCACCAACACAC
CAACACGAGGTGTCTCAAGCTGAGGCTGTGACGCGGACGCGAGGACATGTGCGCAAGTGCGGCTTG
GCCAGTTGTGTGGGACACACCATGCTGCTGTGCTTCTTGACTCTCTCCCTGGGAGCTGTAGC
ACTGGGAGTCCAGCTCCGTCGCCGAGATGACCTGGTGEGCATTTGTGGGCGGCGCAACACCCACGGGAA
GTGTGTGGACGTTGGTGTGACCGGCTGACTGATTTCCGAAAGAGCAACGCCCGCTCCACTCAACGG
ATTACACACAGGAGTATGTATCTGACGGGCGGGGCTGTGAATTGAGCCAGATGCTGTCCACCTCAA
CTACTGTCTTCTCGGGGAGACATCGCCCTGTGAAGCTGGCCAAAGATTCCCTGGAGTTGACTGACA
GTGACAACTGCTGGAAACACAGCTGGGGATGTGGCTTTGTGATATCGCGCGCTCTTACCGCCGCA
GCAGTGAAGTCTCTACACTGACATGCACTGTGACGCGACACTGACCTTTTCTGTGGTGGAG
ACAGAAATTATCTAGAGTACATGATCTGTCCGGCGGACGGGCGCGCATGTTGAAGGTGACTCAGG
CGGCCCTGTGTCTCAAGAAAGGTAAGTGGCTCAGCGGGAGTAGTAGCTGGGGATTTCAGCTGAT
CGGCCAGCTTTGGCTTACACACGCCAGACACACTGGCAGGGTGTCCAACATGACAGCGCCACAGAG
CAGCTGGCAGGTTGCCACCATTGCAGAGGCCACAGAGATGCGGCCAGGGCAGAGTGGTCTGCAGGCCCT
TACCCAGCATCTACTGCTACCGAGGCGCATCCCGAGGCCCTTACCCATGTCACTGTACCTGATGGCTGTC
CCAGACGACCTTACCCAGCATGCTACCGAGGGCTTCCCGAGCGCTTACCATGTACGACTGCTGACT
CTGTGGTCCCAAGACCTTACCCATCACTGCTACATGCTGAGTGGCGTCCGAGCGCTTATCCCGACAT
CATTGCTACCCAAATGGCTTCCCGACACCCCTTACCACGCTGACTGCTGCCACGAGGCGCTTCCCGAGCGCTTTC
ACGACATACCTGCTACAGATGGCCATCCCGAGGCCCTTAACACAGCGCACTGCTACACAGGCGGCATCCCG
AGAGCCCTTACCCACGTTACCTGTACACGATTGGCGGTCCCGAGGCCCATACCCCATGTCACTTGTCTACAGAT

FIG. 1HH

[illegible]

FIG. 111

>SGPR390 SEQID_38
ATGGAGCCGACTGTGGCTGAGCTACACCTGTGCCAGGACCAACAGGAAGTCCCCGCTGGATGCCGCGT
GCTGTGCGAGCGCCAGCATTTGGCGTGGTGGCCACCAGCCTTGTCTCTACCCCTGGGAGTCCCTTTTGGGAGG
AATGAACAACTCCAGACGCTGCTTAAAGCTGCAACACTCCTGGGAAGGTCTACAGCGTCACTCTCTGGAAG
CAAGCAAGCCACCAACCCACGAAGAAATTCGGAACATCCGAACATCAAGCAAACTCAAACTCCGGA
CACACCATTTAAGAAATGAACACTCAGCCCTTCTCTACAGCGCTTCCAGCTGGACCCACACGCGCGAG
CTGCGGGAATCCGCTGACAGCAGTTTGGCGGGGAGACCTCGAGCTATCACCGCACGCTGACGCCACCC
CTGGAGGCACTGCTGCACTTCTGCTCGACCCCTCCAGACGCTGAGCCTGGGCTGGAGGAGGAGCTATTGC
AGCGAGGATCCGGGCAAGGCTGGGAGGACGCGACTCTCCCTGGCTGGCTATGGCACAAATTTGTGGGCTGA
GCTCAGAGGAGACATAAGGAGCCCTTGGCAGAAAGAGACTTCAATCAGCCGCTGTCCAGGGAACCTCTTTT
CCTGGGGAACAGCAGTGTGACCAAGTGAACCCGAGTGTATCAGCAGGAGACTGTCCAGGATCGGCTC
CGACAGGCGCACTGCGAGTGTGGCTTGCAGCCTGCTGGAGGATGGCGGCAAGATCTGTGGCGGCGCATGGA
AGCATCCCGGGGAGTTTCCGTGGCAAGCCAGCCTTCGAGAGCAACAGGAGCACTTCTGTGGGCGGCCATC
ATCAACGCGAGGTGGCTGGTGTCTGCTCTCACTGCTTCAATGAGTTCCAGACCGGAGAGTGGGTGGGCTA
CGTGGTGGCACTTACCTCAGCGGCTCGAGGCGCAGCCGCTGCGGCGCAGGTGGTCCAGATCGTCAAGCA
CCCCGTGTACAACGCGGACACGCGGCTTACGCTGGCTGTGTCTGGAGCTGACAGCCCTGCTCTTCGCG
CGGCATCCAGCCGCTGTGCTCCGCGCTGCCACACATCTCCACCCAGCAAGATGCTGATCTCAG
GCTGGGCTACCTCAAGGAGACTTCCGTAAGCATCTTCTGGGCTGCAATGGTCAAGCCAGAGTGTGCA
GAAAGCACTGTGAGCTGTGGACGAGCACTGTGTCCAGCTTGTACGCGCAATCACTCACTGACAGGATG
GTGTGCGCTGGCTACCTGACGCGGAGGTGGCTCTCCGAGGTTGACTCAGGAGGACCCCTGTCTGCGAG
GAGCCCTGTGGCGGTTCTTCTGGCTGGCATGTGAGCTGGGATCGGATGTGCGGAAGCCCGGCTCCAG
GGGTCTATGCGCGAGTCAAGGCTACGTGATCCTGAGGCCACACCAAGCCAGCATGCTCTGGC
CCCCCAATGGCTCTGCCCCGCGCCAGCAGCCCTGGCCCGCCAGCTCTGAGGCCCTGTGTGAG
ACCCCAACCAATGATGACAGGCCCTCAGTACCGCTTCTGAGTGGTACCCCTTCTTAAGCTACAAAGATG
TGGGCGAGCCTGCAATGGAGAGCCCAACCGGCTGTGGCGGTTCCGAGCTGCTCGGGAGGTGC
CCTGCAAGGTGAGTGAAGAGGTCGCGGCACTTCTGCGGAGCACTGTGTGGGAGCCGCTGCTGCTGC
TGCTGCGCGCACTGCTTCAACCAACAGAGTGGAGCAGTTTGGGCGCCACCTGGGCACTGCTGCTGCT
GGGCTGGGCGGGAGCGGCTGAAGATCGGCTGCGGGGGTGTGTGCAACCCCTCTCAACCCCTGGCAT
CCTGGAATTCGACCTGGCTGTCTGAGCTGGCCAGCCCTGCGCTTCAACAAATACATCCAGCCTGTCTGCC
TGCCCTTGCCATCCAGAGTTCCCTGTGGCGGAGTGCATGATCTCCGATGGGAAATACGCAAGG
AAATGCCACCAAGCCGAGCTCTGCGAGAGGCGTCCGTTGGGCACTATAGACCAAGAAACCTGTAGTGTCTCT

ACAACCTTCTCCCTCAGACAGCGCATGCTTGGCAGAGCTTCTGTGAAAGGCAAAAGTGAATCTCTGCCAGGGTGA
CTGTGGGGCCCTCGCTCGAGAGAGGCCCTGGCGGTTTATCTGGCAGGGATCGTAGCTGGAGCTGGAGTATTG
CTGCTCGTCAGGTAAAGAGCGGGGCTGTACAGCGCTCAGCAGGCTAAAGCGCTGGATCTGGAGATCAT
GTCTCCAGCGCCCTTCCGATCTCCCTCCCTGACACAGAGTGTGCGCCATCGCCACAGCCGACGACGAC
CTGTGCTCAGATTCGGGGGGCCACACGACAGCACCCACTTGGGCTGCCAGCAGGTTGACGGGCCAA
CCTGCCAACTCAACCTTATCTCGCTGAGCAACCACTGCTAGGGACACAGCGCCATTTCCAGACGCCGCCGAGG
CCACACACACACAGCATACAGATGTGGCTGGCGCGCGCGGCTCCACAGGATTTGTGGCGGCGAGCG
CAGCGGGCCCTGGGGAGTGGCCCTGGCAGTGTGGCTTGGCGCGCGGGAACACCGTTGGGGGCG
GTGCTGTGGCAGAGAGTGCTGCTCTGGCGCGCCTACTGCTTCAGCTCAGCGGACCGCCCAAGCATGG
GGCGCTTCTAGGACAGCGCTTCTGACGGCGGAGGGGCGTGGAGCGCGTGGCGCGCATCTCAAG
CACCGCTTCTACAATCTCTACAGCGCTCGACTACGAGCTGGCGCTGTGSAAGCTGCGCGGGCGGTGCTGCGCA
CGCGCTGTGTCCCTCCATCTGCTGCGCCAGCCCGCGCGGACCTCATGGCGCGGCACTCGCAAGGCGCGCTCGCTCATCA
CGGCTGGGGCTGTGTGGCGAGAGGAGCTCATGGCGCGGCACTCGCAAGGCGCGCGCTGGCGCTCTC
AGCAGAGACAGCTCGCCCGCTTCTACCAAGTGCATACAGCCGAGTGTGTGGCGGTTCCCGGAG
GTGGCTGTGACAGCTGTCTCGGTTAGCTGGGAGCCCTCGGCTCGAGGGCGCTCTGACCGGTTGGGTGG
TAACTGGGCTCACTAGCTGGGCTATGGCTGTGGCGCGCCCACTTCCCAGGTTCTATACCCGGGTGGCAG
TGTGAGAGGCTGGATAGACAGACATCCAGAGTGA

>SGPR521_SEQ 38

ATGCAAGATCCCTCTTCTGCTGCCCTCGACATCTACTGCTATCTTTAGCTTGGAAACTGCAGGAAAGAAC
GCAAGTGTGAACGATTATTATGCGCGCCCATGTGCAAGAGGCTCCCAACCATTGGCAGGTGCCCTGCTCAT
GGCAATCCGCTCACTCGGAGGCGCTCTGGTCAATAGCGCTGGGTGCTCACTGCCGCCCACTGCACAGTGA
ATAGTACATCGCTGCACTGGGCACTGTACCTGGCGAGCAGAGCTCAGAGGCACTGCGACTCGAAGTC
ATTCCGCCACCGGGCTATCCACACAGCCGATTAAATGACCTTACTGTGTAAGCTCAATAAGCAGGCCA
GGCTGTTCATGTTGAAGAAAGTACAGCTGCCCTCGGCTCGGAACCGCTTGAAACCACTGTACTGCTCC
GGCTGGGCACTACCAACAGCCAGATGSACTTTCCTGSACTCTGCTGGCTGGATGTCAAGCTCATCTC
CCCCAGGAGCTACCAAGAGTTTACAAGSACTTCTCGAATCTGCGCTGGCATCCCGGACTCCA
AGAAAAGCGCTGCAATGTGTACTCAGGGGAGCCGTTGTGTGACAGAGTACCCCTGCAGAGTCTGGTGTCTC
GGGACATTTCCCTTGTGGCCCAACCGAATGACCCAGAGTCTACACTCAAGTGTGCAAGTTTCAACCAATGGAATA
ATGATCCATTTGAAGAAAGCTTCGTTAA

WO 02/00860

37/68

PCT/US01/20171

FIG. 1KK

>SGPR530_1_SEQID_40
GTCTCCACAGTGTGTGGGAAGCCTAAGGTGTGGGAAGATCTATGTTGGCGGGACGCGACGCTGGCCAGT
GGCCATGGACGGCCAGCCTGCTCTACTGGGGCTGCGACCTCTGTGGAGCTGTCCTCATCGACTGCTGCTGGCT
GGTATCACTACCCACTGCTTTCTCAAAATCCAGGCCCGGAGAACTATCAGGTTCTGTGGGAAACATCCA
ACTGTATCATCAACCCAGCACACAGAGATGTGTGTGACGAGATCATCCACATCCAGACTTTGAGAAAGCT
CCACCCCTTTGGAGTGACATTGCCATTGTGACGTGCACCTGCCCTATGAACCTTCACTTGTACATTGTCCTGT
CTGCCCTCCATCCCGGACATGCAGCTGCCCAAGTACGTGCTGTTGGATAACCGGCTGGGGATGCTCAAC
GAAGACATAAGAGGTGCAACTGTCAACCCCTTCTATCTCCAGGAGGCGAAGGTGGGCCCTCATTTGAGAACAC
ACTCTGTAATACCTTATATGGCAAGAACTGCAAGCGGAGCTTAAGCTTTTGACGAGAGATGCTGTGTGG
GGGGGTACTTCTGACAGAGAAAGTCCATCTGCAAGGCGATTCTGGGGGCTCTAGTCTGTCTACCTCCGAGT
GGCTGGGTCTGTGTGGGCTGGCCAGCTGGGGCTGGACTGCGGGCATCCTGCTACCCGAGCATCTTCAACCA
GGGTACCTACTTCACTGSAATTGACGAAATCATGAGGCTCATCTCTTTCTGACCCCGGCTGGCTCCTC
AC

>SGPR520_SEQID_41
ATGCTGCTGGCTGTGCTGCTGCTACCCCTCCCAAGCTCATGGTTTGCCACGGGACCCACTGTACACAG
CCTGCCCCAGCGCCCTGCAAGTCTTCACTCTCTCTTGGGGGAGAGACTGTGTTGGCGCGCAACCTAGACT
ACGTTTGTGAAGGGCGGTGGGGGAGAGCGTCCGAGCACTGCCAATGTGAGCGGGCCCGACGGCCGACATCG
TGGGGGACGCGGGCGCGCGGGGCTGCGCCTGTGGTGTGAGGCTGCAAGCTGCGCGGCGAGCCTCTG
TGGCGGGCGTCTGTGAGCGGCTCTGGGTGCTCACGGCAGCGCACTGCTTTAGGCTGCGGCTCGACCC
CGCAGCGCCCGAATGAGCTTCTGTGAGCTGTGACGCTGCGCAGAGGGTCCCGGGGGAGCAAGCGGAGGAG
GTGCCAGTGAAACGCATCTGCCCAACCCCAAGTTTGACCGGGGAGCTTCCACAGCACTGGCCCTGTGTC
AGCTGTGAGACCGCGGTGAGCCGGGGGATCGGGCGCCCGCTGTGCTGCCCCAGGAGCCCGAGAGCCC
CCTGCCGAAACCGCTGCGCCATCGGGGCTGGGGCGCCCTTCCGAAAGCGGGCTGAGGCTGAAGCAGTG
AGAGAGCCCGTGTTCCTCTGACGACCGACCTGCGGAAAGCGCTGGGGCGCGCTGCGCCCGCAGC
ACCATGCTCTGCGCCGGGTACCTGCGGGGGGCTGACTGCTGCCAGGTGACTCGGAGGCGCCCTGACC
TGTTCTGAGCCTGCCCCCGCCCTAGAGAGTCTGTTGAGAGTCACTCTGCGGGGACGGCTGCGGGGAG
CCAGGAGCCCGGGGTCTACACCGCGTGGGAGTGTTCAGGAGCTGGCTCCAGGAGCAGATGAGCGAGCC
TCCTCAGCCCGGAGCCAGCTGCAAGGAGCTTCTGGCCTGGGACCCCGGAGCTGAGGAGAGAGCC
GCCGCTCTGCGCTTCTATGCCGCTGTGCCGGGTCCCGAGGGGCGCTGTGCGCGCTGGCGCACAG
CAGTGCCTGCAGCGCGCGGCGATGCGAGCTGCGCTGCTGGCGCACAGCTGCTGGGCCCTGCTGGCGAAC

WO 02/00860

38/68

PCT/US01/20171

FIG. 1LL

GGCAGAGAGTGTCTGGGGCTCTCGGGAGTCTGGGGGCTGGCCCGCCCTGGCTCTCCCGGCTCCAGCG
CTCAGGAGTCTCTCTGACCCCGCGGGAGCTGGGCTTCACTCAGGATCGGGGCTGCAAGGCACTCGGT
TCCCGAAGCGAGCGCGGAGCGCGGAGAGCCAGCGCTGCCCTGGGCTGGAAGCCCTGCGACAGAAGT
TGCTGCCCTGCAGGGGGCCCATGCCCTGGATCCTGCAGGTCCCTCGGAGCACCTGGCCATGAACCTTTCATGA
GGTCTGGAGATCTGGGCTCCAGACACTGACCGGGCTTTTGAAGGCTGGGCTGGGCGAGGCTTGGGGGG
CGGCAATGTGGGCTTCAAGCGGCTGGGCTGGAGCGCCGACACTGCTCGCCGCTCCCGGCTGCT
GGTCAAGGCCCTGCAAGGCTTCCGCTGGCTGCGCTGCGAGGAGGAGCCGAGGACCCCTGGATGGATGT
AGGGCAGGGGCCCGGGCTGGAGAGGAGGGGACACCCACTCAACCTCAGGTACCCCGCGCCAGGCAACC
CTGA

>SGPR455_SEQID_42
ATGAGTCTGTATTTGCACTGCTGTATCTAAACACAAAGTCAAGTTTGGAAATGCTTTAGCCAAATCTGTCTTC
CTGACAGCGATGAAGTTGAACCAAGAAATCTTTGCTTATCCAGTGGATGGGCAAGATTTCCAAACATCAG
AATATTCAAATGCTCTACAAGAAATGGAACCTTCCCATGATGACAGAGCGTGTAACTGTGCTCAAGAGCAT
GAACCTCCCTCCCTGGGAAGGACCATGCTGTGTGCTGGCTCCCTGATTTGGGGAATGGAGCCCTGCCAGGG
GACTCTGGAGGACCATGCTTTGTAGAGAGGTGGTGAATCTGGATCTTCTGGGATAACTTCTCCAAAGTGTC
TGTTGTGGAGGTTTCACTCCCGTAAGAAACCACTGTGAAGGATCACTTGGCAATTTCTCCAAAGTGTC
TGAGTTGATGGATTTATCACTCAAACTGTTACAGGTTTGGATCGGGGCCAACCCCTCTCAAAGTGGGCTC
AAGGTATATACAAAGGCCCTGAGTTCTGTCGAAGAGTGAATGGAAGCCAGAGAGATAAATCTGTATAA
ATTTACAAGTTTAGACATGGAAAGCAAGTTGGATGTGATCATGACTATGTATCTTTACGATCAAAGCAGTGGAGTG
CTTTTAGTAAGGTCGTGGAAATATTTGCCCTCACCATTTGCTGGCAGAGACCGTGGGCCATGGTTCCATTT
GTTCTGATACAGAAGACAGTGGCAGTGGCTTTGAGCTTACCGTTACTGCTGTACAGAGTCAGAAAGTCAGAAAGCGGTC
AGGTTGGGAGTCTGCTATATTGGTAGAAGAGGACAAATCACTCTGCCAAGTATCTGATTTGTATCCAG
TAACACAAGGTGTCTGTTCAATTTGCTCCAGAGAGCACATTAATGAAGTTGACATTTGAGGACTTTGCTGTC
AAATTTAGTCCAACTGTATTTATGAGTGTGTTGATTTACGGTGAATCTGAAGAAAGCACAAAGTTAGCTAACT
TTGTGGAATTTGACCATCACTCAATATTCAGTTCTAGTAACATGACGGTGATATCTTTAAAGTGATGTTAA
ATCGTTTACAAGGCTTCAAGGCCAGATTTACCATTTGCCCTCAGAGTCTTTAAACAAATTTGAACCAAGTTACC
TCCCAAAACAACTCTGTATCTACCGTAAGCTATCTGCAATGATGCTGTGGCATCCCTCAATTTAGTCCCGAG
TGCTTTCCAGAGAAATCGAGAGGGGGAAGGCTGCCCGGCTGTTGGCCATGGGAGGCTGGGCTGGAGT
TTCTAGGGGATTTACCAATGTGGAGGTGCCATCATCAACCGAGTGGATTTCTGACCGGAGCCGCTGTGCA
TTGAAGAAATATCCACTCTCTGGACTATTTGCTGGGAGCCTGACGAAACCTGAAGGAATCAACAGAGCAG
GTGAGAAGGGGCCCAACACATAATAGTGCATGAAGACTTTAACACACTAAGTTATGACTCTGACATTTGCCCTAATAC

WO 02/00860

39/68

PCT/US01/20171

FIG. 1MM

AACTAAGCTCTCTCTGAGTACAACTCGGTGGTGGGCCAGTATGTCTCCACACAGCGCAGAGCCCTCTATTTT
CCTGGAGATCTGTGTGACCGGATGGGAAGCATCAGTGCAGAGCTCTCTCTGAATGTTCTTCTACATAGATG
GTGGCTAGCAAGTGGCTACAGCAGATTCAAGTGCATGTGTAGAAAGAGAGGTCTGTGAACACACACTTACTATT
CTGCCATCCAGAGGGATCACAGAGAAGATGATCTGTCTGGCTTTGGAGCATCTGGAGAGAAAGATTTCTGC
CAGGGAGCTCTGTGGGCCACTAGTATGTAGACATGAATAATGGTCCCTTTGTCTCTATGGCATTGTGAGCTG
GGAGCTGGCTGTGTCCAGCATGGAGCCGGGTATTTGGCAGAGTGATCTTCTTGGACTGGATTCGCAAT
CAAAATCAATGGTAAATTTTCAATGTTAATAACAATAAAGCTCTTCTTTAGAGTGGTITGGGAACAGTG
AGTTGTCTCTGAAGCAGAGCTAGAAAGCCTAGAGCCTTTTCCACACCACGGTATCTACTGGATTATAGA
GGAAGCTGGAATGTTCTTGGGTCTCAGAGTTTCAGAAAGCATGTAGCAAAATTTACCATTGAGTATCTGTCA
CTCCTGGGGTCTCCTGTGTGTCAGACTCAGTTCTAATTAATTAAGAAAGACACAGTGAAGAAAGACGGCA
GGTGGATTACATGGAAAGAGCTTTACTCAATGACTTTCATGATCTGTGACCGCTGGAGGTGACATTCCAT
GCCCTTGTACAGGTGCTTTGGTATAAGCTATTTGCTTGAAGTCTAGGTCCAAAGGACAGTAAATTAACG
AGACTTCCCAAAGTTCAACACAGAGACACTTGGTCCCTTGTGAGGATGTTCTTGCACCAAGCCAGAGGGATC
ATGAGATCCCAAGAAATTTCTCAGAACTACTATGGTGGTGGCAATGGAGATTAGTAGCCCTTTAAATCACATCA
TTCAGCTTAATATTAATTAATTTCCGATGAAGCCCAACACTTTTGTCTGTGATGGTATCTGCGTGTTCAGGAG
ATTTGGACGAGGAAATAATAGGTAGAATTTGATGAGCACTGAGCTTTCTTGGTTCCTTAAGCAATTGAGG
ACCAAGAGACCAAGCTCTTGTGGGAGACTGCAGTATCTATGAAATGATGATATCTTCTATCTTCTTCTTGGCC
TACAGAACAGCTGTACCATGCACTGCTCATGAGGTTGTTTGGAAATTAATAA

>SGPR507_2_SEQID_43
ATGAAATATGCTCTCTATTTTGGGTGTCTCTGCTGGGACATTTTCTTGTGCTGACTATCTGTTCAAGAAAGAGACC
CTGCTCCCTATTGTGTACCTCAAGTCTCACTTCAACCCCTGTGTGGGCTGCTCTCAACCCAGCTGGGTGC
TGGCCCAAGCTCACTGCTATTACCAATCTGAAAGTGTGCTGGGAAATTTCAAGACAGATCAGAGACGATA
CTGAACAGACAAATTAACCCATTCAGATCGTCCGCTACTGGAAGTACAGTACAGGCCCCACAGATGACCTCA
TGCTATCAAGCTGGCTAAAGCTGCCATGCTCAATCCAAAGTCCAGCCCTTACCTCGCCACCCACCAATGTCA
GGCCAGGCACTGTCTACTCTCAGGTTTGGACTGGAGCCAAACAGTGGGCTTTGGCAGCTGGAGCCA
CCAGGCCATCTGACTCTGCACAGAGGCCAGCCATTCTGATTGGCAGAGACACAATTCACATGAACAGGCCG
ACACCTGACTTGGGCGAGAACCTGGAGGCCCTGATGCTGATCGAGAATGCCAAAGAGGAGGAGGAGGAGG
AAAGCCACAGGAAATTCCTTATGTGTGAAATTTGTGAAAGTATTCAGCCGAAATTTTGGGAGGTTGGCGTGTGCTA
CTGTGATCTGCAAGACAAAGCTCCAGGAAATCGAGTGGGGCACTTTCATGGGAGGGGAGCGTGGCATCTACAC
CAATGTTTACAAATATGATCTCTGATTTGAGAACACTGCTAAGGACAAAGTGA

>SPR559_SEQID_44

ATGGCGGAAAATGATCCGCGTCTGTTGAAGGCCCCCTTCATTCCGATCGCTTTTGGCCTTGATGATTTGAAGA
ATAAGCTGTGTGACACAGCGAGATGCTGTGTCGACAGATCCTGTCACTGCTGCCAATGAAGTTTTTTCCTCA
ATCATGTCATATGGGATATTCGATATAGCAATGCGCATTTGCTGTGGCATCCACTTCGACTGCTCAGAGG
AAGTACAGATTCGCTCATCTTTAAGTGTATCGAGCTGATGCTGATGTGACGAGATCTCGGATTTGCAAAAGC
GGGAGGACGAGTACCGCTGTGTCGGGTGGGTGCAAAATGCCGTGCTCAGGTTCCACAGCTGCTTCGT
GGAAGACCATGTGCTCCGATGACTGGAAGGGTCACTACGCAAAATGTTGCCGTGTCGCCCACTGGGTTTCCCAAGC
TATGAGATTCAGATAACCTACAGTAGAGTCTGCTGGAGGGAGATTCGAGGAGAGATTTGTGTCATCTCGATCA
CTGTTTCCGAATGCAAGGTGTAAGCTATTACCACTCATGTTATGTAGGAGGAGGATGGCTCTGGGCCACG
TGTTATCTTCGAGTGCAAGACGCTGTGGTGTATAGAAGGGCTACAGCTACGCACTGAGTGGTGGGAAACATGTCC
TTGCTCTCGAGTGGCCCTGGCAGCGCGCTCAGTTCCAGGCGTACCACCTGTGCGGGGCGCTCTGTCAATCA
CGCCCTGTGGAATCACTACTGCTGACACACTGTTTATGACTTGTACCTCCCAAGTACGACCATTGACAGCTGG
GTCTAGTTTCCCTGTGGCAATGACGCCCATCCACTGGTGAGAGATGTGTACCAACGAAAGTACAAGC
CAAAGAGGCTGGGCAATGACATGCCCTTTAAGAGTGGCGGGCCACTCATCGTTCAATGAATATCCAGCTG
GTGTCCTGCCCATCTGTGAAGAAAGCTTTCCCGATGTGAAAGTGTGCTGAGCACTGAGTGGGGGCGACAG
AGGATGGAGCAAGTACGCGCTCCCTGTCTCCGTAACACGCGCGCTCCCTTTTCCCAACAGATCTGCACAT
CACAGAGCAAGTGTACGCTGTGGCTGAGCATCTCTCCCTCATGCTCTCGCGGGCTACCTGACGGTGGCGTGGACA
GCTTGCCAGGGGACACGCGGGGGCCCTGTGTGTGTCAAGAGAGAGGCTGTGGAAGTTAGTGGGAGCGGACCA
CGTTTGGCATCGGCTGGCGAGAGTGGCAAGCGTGGGTGTACACCCGTGTACCTCTCTCTTGGACTGGAT
CCACGCGAGATGGAGAGACCTTAAAAACCTGA

>SPR567_1_SEQID_45

ATTGAGAGGGACACGACACGCGGAATGCATCTCCAGACGAAGACACCTTCAGCTGGAGCATCTCCAGCCCGAGCGAT
CTCCAGCTGGACACCTCCAGCGCGGCATCTCCAGCCAGGCATCTCCAGCCGAGGCATCTCCAGCTGGAC
ACTCTCGGGCGGGGATCTCCAGCCGGGCATCTCCAGCTGTTACCTCCAGCGGGGATCTCCAGCGCGG
GGCATCTCCAGCCGAGGATCTCCAGCGAGGATCTCCAGCGAGGATCTCCAGCGGGATCTCCAGCGGGATCTCCGCT
GTGCACTATCTTCCAGTCTCCAGTCTCCAGCGAGGATCATCTCCGCAAGCTCAGCTCGGTTGCAACCTCCCGAAC
CAGAGTGTACCTTTGTAGACAACACACAGTGGGCTGTACCCATCCGATCTCTCGCAGGTGAGGACACCA
CAACAGGGCCACAGAGCCGACAGTCCAGTTCTGTGCAAGGCGCCAGGATCAGTGTACAAGAGGACGAG
GGAGAGAGCTGTCCCAACACAGCTGTTCGCTGTGACGGCGGGTGGAGTCCAACTGCAAGAGTGAAGAGCTGG
GTGCTGTGAGTTTGTACGTGGGACAGTCTGTCTTAAATCTACTTGGGCTCTCCCATCAGTGGCTTTCCCAT
GTGACGCAACATGGAATGACTCTACTGACAGAGCACTGCCAGACAGCTGGTTCGAGAGTGTCTCAGCGGAC
GTGACGAGCACTGGAATGACTCTACTGACAGAGCACTGCCAGACAGCTGGTTCGAGAGTGTCTCAGCGGAC

WO 02/00860

41/68

PCT/US01/20171

FIG. 100

ACCGAGTTGCCACAGGGATTTGCCAACAGCTTCTCAATCTTTGAGATACAACCTCCACCATTCCAGGAAGCCTC
CACAGTCTGAATGCCCTTCCAGCGGTATATCTCCCTCCAGTGTTCACACTGCGGAGTGAAGTCTGCACTTCGGCACCCAC
CGGATCGTGGAGGGCGCTGGCTCGGATAGCAAGTGGCCCTTGGCAAGTGAAGTCTGCACTTCGGCACCCAC
CCACATCTGTGGAGGACGCTCATTTGACGCCAGTGGGTGCTCACTGCCGCCCTGCTCTCTGCGTACCCCG
GAGAGTCTCTGGAGGGCTGGAAGGTGTACGGGGCCACGACCACTGACCACTTGGCTGAGGAGCCTTCC
ATTGCGAGATCATATCAACAGCAATTACCCGATGAGGAGGAGCTATGACATGCGCCTCATGCGGCTGTG
CAAGCCCTGACCTGCTCGCTCACATCCACCCTGCTTGGCTCCCATGACAGACCTTTAGCCTCAATG
AGACTGCTGGATCACAGGCTTTGGCAAGACGAGGAGACAGATGACAAGACATCCCTTCTCCCGGAGGT
GCAGGTCAATCTCATCGACTTCAAGAAATGCAATGACTTCTGCTGTATGACAGTTACCTTACCCCAAGGATGAT
GTGTCTGGGGACCTTCTGGGGGACAGACTCTCTGCCAGGAGAGCGGGGGGCTCTTGTCTGTGAGCA
GAACAACCGCTGTGTACCTGGCAGGTGTACCAAGCTGGGCAAGGCTGTGGCAGAGAAACACCTGTGTG
TACACCAAGTGACAGAAGTTCTCCCTGGAATTTACAGCAAGATGGAGAGCGAGGTGCGGATTCAGAAAAATCCTAA

>SGPR479_1_SEQID_46
ATGGCTGCCCTGCTTCCGTCATGGGCCCACTCGGGCCCTCTGCCCTGGGCTTCTGCTGCTCTCTGCTGCTG
TGGCCCTCCCGGTCGAGCATTTGGTCCACAGACAGCAAGGAAATCTCCCTAACTGGCAAGCGT
GGCTGTGTGCGGCGCAGATGGAGGGGAAATCTTGGGCGGCTCCCTGCGCCGAGAAAGTGGCGGTG
GCAGGTGAGGTGCACTACGAGGCTCCACGCTCTGGGCGGCTCCCTCAATGAGTACTGGGTGCTGTCA
GCTGGCACTGCTTTCAGGGGACAAATATCAAAATCTATGACATGACGTAGGCTCGTAAACCTCAGGGTG
GCCGGCAACCAACACCCAGTGGTATGAGGTGAACAGGGTGAATCTGACACCCACATATGAGATGACCAACCCAT
CGAGGTGAGGTGCGCTGCTGAGCTGAAGACCGCATTTGTTTCTGAGTCCGCTCCGCTTGGCTTGG
CAACTCCAGAAGTGAACCTTACCAGTGCCAAATTTGCTGGGCTACGGGATGGGACTAGTCTCAAAACAAGGTGAG
ACCTCAGAGAGCTGCAGAGGTGCAAGTCCCGCTGATCCTGGAGCCCTGTGCCACCTGCTCTACGGACACA
TGTCTACATCATGCCGACATGCTGTGTGCTGGGGACATCTGAATGCTAAGACCGTGTGTGAGGGGACTCC
GGGGGCCCACTTGTCTGTGAATTCACCCGAGCTGGTTGAGATTTGAGTGGGCGGAGGCTGCT
CCAACCTCTGTACCTGGAGTGTGCCAGTGTTCCTATTTCTCAAAATGATGTGAATACATAGAAATCAC
GCCACTCTGCTCAGCCAGCCCTGCTCTCTCTCCAGCTCTGGGGGCCCACTCTCAGCGTCTCTAATGGCCATGC
TGGCTGGCTGGTCAGTCTGTGA

>SGPR489_1_SEQID_47
ATGAGTCTCAAAATGCTTATAAGCAGGAAACAAGCTGATTTTACTAGGAATAGTCTTTTGAACGAGGTAAT
CTGCAACTCTTTTCGCTCCCAAGCTCCCAAGTTGTGGGCGAGTCTGGTTAAGGTACAGCCTTGGAAATATTTTA

WO 02/00860

42/68

PCT/US01/20171

FIG. 1PP

ACAATTTGAGTGGCATTCTTGAGGAAGCCAAAGTGGAGAAGGGTTCTATCCCTGGCAGGTATCTCTGAACAAA
GGCAGAAAGCATATTTGTGAGGAAGCATCTCTCACCAGAGTGGGTGATCAGCGCGCTCACTGCAATTCGCAAA
AGAACATTGTGCTACTTTGAATGTTACTGCTGGAGAGTATGACTTAAGCCACAGACACCCAGGAGCAAACT
CTCACTATTGAACCTGTCAATCATCATCATCCACATTTCTCCACCAAGAACCACTGGACTATGATATGGCCCTTTGAA
GATGGCTGGAGCCCTTCCAAATTTGGCCACTTTTGGGGCCCATATGTTCTCCAGAGCTGGGGGCAATTTTGAGG
CTGGTTTATTTGTACAACTCAGGCTGGGGCCCTTAACCTGAAGGTGGCGCTCTCTCAAACTCTTGCGAGAA
GTGAATCTGCCTATTTTGACCTGGGAAGAGTGTGGCAGCTCTGTTAACACTAAAGAGGCCCATCAGTGGGAA
GACCTTTCTTGCACAGGTTTCTGATGGAGGAGAGACGATGTACAGGAGATTACAGAGGTTCACTCATGT
GGCGGAATAAGAAAGGGCCCTGGGACTCTGGCTGGTCAATTTGGGAGGCTCAGGTGGGAGGATCGCTTGAGTC
CAGGAGTTCAAGACCAAGCCCTAGGCAACAAAGTGAGACTCTGTCTCACAAATAATTTCTTCAAAAAATTAGCCCG
GTGTGCACCTGTGTGACAGTGCAGTGCAGGATGTATAGTACGCGGCTGAGGGAGAGCTGCACTTCCAGAAAGC
CTCCACCTATATTAGAGCAAGCAACGGGTGTCTGGACCCCTGCTGTAACAGAGGAAATGCATGTGTTGCT
CAGTTTCCACCTAGATGTTGAGTCTTGTCCACACAGTTACCTGTCAATTTCTTTAGAAGACAGACCCATT
GGAAATTTGTGGAAGAGCCCTCCCTTCACTCTTATTTGGCTCAATTTCTTAAGGCTGAATTCGCTCTCTG
ATGCCACAGATTATGCACTGGGTTTAACTTACCTATAAAGCTCTTAACCAAACTACATTTCTGGTTGCAATTA
CTTAACCTGCTCTTTTGAAGAAAGGTCTCATACAGAGCTTAACTATCTGAAAGCTGCAAGCAAGGCTAACTGT
GACTGGAATTTTCAAGCCCTCCAAACATCACTAAATTAAGCTTTCAAGTCTGGAATAGAAAGAAAGTGGAG
ACTGCATTTCCGACTATGTGACAGTGCACAGCGATGAGAAGGAAGGAAATAGCTCGGCTGTGTGGCTAT
GATGCCCAACCCCTGTGCTGAGCCCTCCAGCATCATGCTCATCAGCTTCCATTCAGATGAAGAGGGACCTG
CAGGGGCTTTCAAGGCTATAGTCTCTCTTCAITTCCTAAAGCAGTATACCCAGATTTAAAGCATTCCTCATATCAGAGGAT
GAGTCAATGTTTCTGGAGACATGA

>SGPR465_1_SEQID_48
CGGTGGCCATGGGAGGCGCAGTCTCCTCTACCTAGCGGGGCACATCTGTGGAGCTGCCCTCATCGACAGCAACT
GGGTGGCCCTCTGCTGCTCAGTCTTCCAAAGATGCATCTCCCTCCAGGGGCCCGCTGTCCACTAACCCCATCT
GATTACCGGATCTGCTTGGGTATGACAGCAAGCCATCCACAGCATCCACAGCAGACCAAGCATGACAGTGAATAA
GATCATGTTGACGCTGACTATAACGAGTTGACCCGATGGGAGTGACATCACCTGTGCTGACGCTGCACCATC
ATGTGAATTCAGCTCCACATCCCTCCCGCCCTGCTTCCGGAACCAACACGCTGGCTGGCCCTGACAGCTCC
TGCTGGATATCTGGTTGGGGAATGGTCCCGAGGATGCTTCTGCTGAGCCCTTCCAACTTCCAGAGGCGAG
GGTGGGTGTCATGGACACACTGTCTGCGGATCTTTTCCAGCCCGAGTACCCCGGCGCAAGCAAGCAGTG
ACTACACCATCCAGGAGCATGCTGCGCTGGGACCTCATACAGGAAGGCCATTTGCCGAGTGAAGTCC
AGGGGTCCCTCGTCTGCCCATTAATGGCACCTGGTTCTGATGGGGCTGTCTAGTTGAGAGCCTCGACTGCTG

FIG. 1QQ

CTACCCGCTCGGTCCCAAGGCTTCCACCAAGGCTCCCTACTTCCACCACTGGATCATGCCAAGAAGAGGGAGGAG
GACCCCTCTCAGATCCCGGCTTGGCTCTCTCTCTCAGGAACACCCCAAGCTTGGACAGATGACCTTCTCTCAGGG
CATCTGTCACAAGCCCGGGCTCTGCGCAGCCCTCTGCTGCTCACAATGTTTCCCTCTGCTGCTGCTGCTCTCAGGG
GGAGCGCTG

>SPR524_1_SEQID_49

ATGCACAAAGAAACAGCGATGTTTTCAGCCGCACTGCTGACCTGAAATATCCAAATATCTCAGTCCAAAGTGGTCT
ATGTGCCCAAAAGACGTGCCAGTGTAGACAGACACCCGTTGCCAGGGAGACGACTACCATTCGACGGAAGACGAC
CACCACAAAGACCATTTCCGAAAGCCAAACCCAGAAAGTCCATCCAAAGAAAGTTCCTCTTGGAAATGTACAAA
ATAAATCTCTTCCAGATATTTTATCTCAGCATCATCAAGTGTGGACACTTCTGTGCTGCTATATACAT
AAACAGAAAGCAAGATGCTTTTACTTGTGGGAATTTTCGATCATCAACATAGTATTCCTCCGGAATACC
GACAAAGAGTCCAGGGAAATCTTTCAGTGTCCAGGACTGTGCGAGCAAGTGATTAACCTGGTTATACACAT
CTGCTTTCTCCAAATTTATGACAGCTGTGTGTGCAGATGTGACGACAAACAAAGCGGCTCTCTGTCT
TTGTTGGATTTGTTTTCATGCCACGTCCAAAGCCCTCTGTGTGAAGACTGTGTGCCGCATCTTGA
GGACTCATTCAGACAGCATCATAAACCGGACCTCTGTGGGAGCTTTCAGGAGCTGTGCTGGACATGTGAC
CTGTGGTCTTAATGGTGATTTTGGTCACTCTTAAACAAAGAAAGAAATGTTGCTGTCCACAG
ACAAAGGCTGCTCTGACTTCTATGCGAGCATCTGTCTCCACTCCCGCTGGAGATTTCTGCAAGCTCAG
GGAGCTGATGTGTCACTCAAGCTGTGGCCNAGTGGGCTACCTGATTCGTCTCTCAATCAAGTCCATCCAAA
TCGAAGCCGACACATGTGTCACTCACTGCTGACCATTTACGACTCCCTTTGGCCATCCGGGACAGCATCTGT
ACAGAATTTGTGAACCCCAAGAACATTAATGATCTTTGTTTCAACAATCTCCATTTGGTGACATTTAAGCT
CTCTATACAGAGGCTCTCAGGAATCCGGGCAATTTTGAAGCTATTCAGACAAAGATGTGAACAAACAGTG
TTGTCCAAAGACATCTGCTTGAAGGGAAATTTCAAGCCCATATTACGAGCTACTCTCTCCAAATGCTA
AGTGTACCTTGGAAATTCAGATCTCTATCACTGTGCGATAGCATGAAATCTATTAAGCTATTCATTAACCAG
AAGATATGAAGGCTGTGACATGGATGTGGGAATTAATGACGACATGTACTGTGGCTCTCATGATCAT
CAGACAAATTTTCGAGTCCCGCCGCTTGTTCACATTCAGCTCAGTGGATTCGAAGCTTTCAAGACAGCCA
GTCTTCCAGATATGBCAGTTTACAGATCAGTCAACCTCCGCTTGAAGCTTTTATAGTCTCTCCGGTTCT
GTGTCCGTCAGGCCCGCTGTGTGATGAGTAAGTCACTTTGATGAAGATGATGAACTGTTTGGTGAGCC
CTCAACTGCCGTGCATACCACTGCTTCAGGACGATTCGCCCTCTCATCTGTGATGGCTTCAGGCACTGTGAG
ATTGGCCGGGATGACCAACCTGCACTCAGATTTCCATGCACCAACAGCAATTTTAAGTGTGSCAATGATTT
TGCTTTAGGAACAAATGAANAATGTATGGGACAGTGGATTTGCCAGATGGAAATGATGAAGACAGCTGCAC
CTCAGACAGGATTTCTCCGGCTTTCCAGCTCATCTCAGCGGACAGACACCTGGAGGGGGTGTGGCCGTGG
CAGGTCAAGCTCCACTTTTGGATCTGTGCGCTCAGTCTCAGAGAGTGGCTTCTCTCTCTCTCTCTGCA

GCCACATGTTTTCATGGAACACAGGCTGTCAATCCACACACATGGACTGSCACACCTCGGGATGTATGTTCCAGGG
GAATGCGCAAGTTTGTCTCCCGGTGGAAGAAATGTTGGTCTCCAGAGTACTATTAACAGTCAGACATCTTGTGATATGA
TATTTGTTTGGTCACTCCAGTATTCGCTCGGCTGAGACCGTGAACACAGCTCATACAGCAATGCTTCACTGCTCC
CATCGTTGCAGAGATTCGCAAGTGGGGAAGTGTGGTAACTGCGTGGGGCGAAGACACGAACAGCAGATAT
AAAGGCTCCCTCGTTCTGCACGACGAGGAGTGAGACTCATTCATCAGACAGTCTGTGTTCCACTCAGCGGAT
CATCATCTCGGATGCTGTGCGAGGCTAATGTACGCCAAGAGAGTGCCTGCAAGGAGATTCGGGTGGAC
TTATTTCTGTGCGAAGAAAGTGGTGGAAATGGATTTGACTGGCAATGTAGCTGGGACATGGATGTGGAC
GACCAACATCTTCTCGGTGTTCACGAAGGGGTGTCAACATTTGTTCCCTGGATTCATAAATATGTGCCCTCTCTTTT
GTAA

>SPR422_SEQID_50
ATGACATGCAACAAAATTAAGAAGACCTTTTTCGACGGAAGAGACGTGGAGTTGCGCCCGAAGCAGCAAGTCTGT
AGCCATGTGGATGATGCGGCTTCTATGTTGTGCCGTGAGAGTGGTGCGAGTGCACCATAGGTCTCCTGGTTCAC
TTCTAGTATTTGACCAAAAAAAGAGTACTATCATGGCTCTCTTAAATTTTATGATCCAAACATCAATAACAATTT
CGGAATTAAGCAACACATACACTTAAGGAGTCTACGAGACGACCGAAAATTTGGTGTTATCTTTGAAATGTGAT
CTTTCTTTGTGTGTCACATGCTCCAGAGGATGTGTGGAAGTAGATGCTATATGTGTCTCCAGTCCGCCCTCTA
CTGAACAAGAGGGCAGTAAGAGAGAAGAAATCCAAAGCATCTTAACTCAGAAGATTAAGGAATTTAAGACGCTTG
CATAAATGCGCTCATCATGTTCAAGTTAATGTGGCCATGTTCCAGATGTCGAATGTCGGCTCGGCGCAGGTTCCGGAGCA
GGAAGAGTCCAGGCTGGGACAGGTCTCGCTGGTCCAGATGCTCAAGTACGAGCTCATCAACAGGGGAGTTAACTGTCC
AAGCAAGTTGTGGTAAACAGTGTTCATTTAAACGTGTCACAGATAGCATTTGTGAGTCTTGACCCAGAGGCGG
CTCGGCCCTTCCGAACTCTCCCTTCAGTATGATCATTCATTCAGTGTGGGCCACCTTGATTAGTGACACATGGC
TTGTACTGCGACACATGCTTCCAGAAGTAAATCCATCAATGCACTGTAGTTTGGAACACAAATATCA
CCCTCGTTAATGAAAGAAATGTCGAGAAGATTTATTCATGAGAAGTACCGCTGTCGACAGAGATGACGA
CATCTGCTTTGAAAGTCTTCCAGAGTCACTTTCCGATGAGATACCGCTGATCCGAGATTTGTTTCCGCAAGGCTC
TGATCTTCCGAACAAATTTGACTTCCACATCAGAGATTTGAGCACTTATCATAGTGGGAGATGCCAAAT
GATCTCCGAGAAGCCAGAGTGAATATCATATGACGATGTCGCAAGCAACACAGAGTGTATGGCAATGATATA
AAACCTTGAATGTTCTGCGCGATATGGAAGAAATTAATGCTCGAGGGTGATCTCGGGGACCTTTA
GTACAGAGGAGATCTGAAGAATACGTGGTATCTTCCGAAATTTGAGCTGGGAGATCACTGTGGTTCAAAAGGAC
AAGGCTGGAGTCTACACAGAGTGAATTTGAGGAACTGGATTTGCTTCAAAAGCAAGGCATTA

FIG. 1SS

>SPR538_SEQID_51

ATGACGCTGATGCTGCTGATGATGACCAACCCCTATGGAGGCCCAAGTATGACAGAGAGAGGCCCAAGCACTGGGATCT
TACAGACAGAGCTGGAGACACACAGCATCCCATTTCTCAGGGGCTGTGCTGGCGCTTCATTCGACGAGCTGGCTGT
TGCACTGTGGAGCGCTGGGCTGTGGCGCTGCAGGTGTGGCTCTAGCTGCTTCTGATGCTGTATCTCTGT
CTCTGCTCTACGCCAATTCGGGGAACTTGCAGGATGAGAGATAAATCTTGAGCTGCTCAGAGCAGAGCG
TGAGGAAGCTCTCTCCCTGCATCTCCCAAAACATGATCTTTCAGATAAAGCAGAAAGATCTTCTGTGGAAAG
CGAATGAGGAGTACACCAAGCTGGCTCTGCTTGCCTATGAGGCTGAGCGCCCGCTGGGCTCGAGAT
CTCTGGAGCCTTGGCATCTGACATCTACTACACAGAGGATAAACTCACTACATCAACATCAACAGTCT
CAGAGATTTGCTCAGCTCTCTCTAGACTGGAGGCTTCTTGGAGAGGGCTGAGCAGGCCACAGAACACTCG
ACTCTGTGCAAGTGTTTTCCCTCAGAGCTCTGAGTGTGAGAGGAGGCCCTGGCTTCCCGACATGTTGGTGG
GCAGTGTGGCTCTGGGCGCTGGGCTGAGCGCCACCTGGGCTGGGCTTCCGGACACAGTGTGGGG
CTCTGTCTAGCGCCAGCTGGGTGGTGGTGTGACTGTGCATATTGATGACAAATTGACGACTGGCCGCTCTGCCA
GCTGGCGGCTCATGGCGGCTGTGTAGCAGACAGTGGCTCAGGCCCAAGAGGGCTGTGGTGGAGAGA
TTATCCCAACCCCTCTCAGTGTSCCGAATCATGACTACGAGCTGCCCTTGAGGCTCCAGACCCGCTCTC
AACTCTCAGACACTGTGGCGCTGTCTCTGCGCCGACAAAGAACACTTTTCCAGAGGCTCGCGGCTGCT
GGGTGTGGCTGGGCGGACCCAGCCCTGACCACTACAGCTCGGATATCTCAGAGACAGCTGTGGCT
CTTTCTCAGCACTCAGCTCTGCAACAGCTTTGGGTGTACAGCGGAGGCTCAACCCGCGATCTTTGGGCTG
CTTACCTGTGACCGAAGAGGCTGTGATCGCAGCGAGATAGCGGGGCGCCCTAGTGTGCGAGATGGGAGCA
CATGGCGCTAGTGGGGTGTTCAGCTGGGCGCTGCTGCGCAGAGCCCAATCAACCCAGGTGCTACGCCCA
AGTAGTCTGAGTTCTTGAGCTGGATCCATGACACTGCTCAGGACTCCCTCTCTGA

>SPR527_1_SEQID_52

ATGGCGCGCACTCTCTCCCTTTGTGCTGTTCATCAGTCCCATCCAGGAGCCTTCCAGGACTCAGC
TCTCATGCTACCCAGGAACACTGAAGATCTGCACTGCGGGCGCCTGAGCCCTCGGCCGCGCATCTGGGG
GGCTCAACCGCCGCCGCGCACTGGCTTGGCAATGAGGCTGCACCATGAGGTGGCCACATCTGCGGG
GGCTCTCATCGCCGCGCTGGTCTCTCGGCTGCTCACTTTACAGCAATGGAGCTGGAAGCTGGAAGCGG
CGGCGAGGTGTGCTGACTGTGGGTGTGACTCCCAAGAGAGGCGCCCTGAGCGAGCTGTGAGCGGCA
GTGGCGGCACTGTGGTGGCGCAACTACAGCAAGTGGAGCTGGGCGCGACACTGGCCCTGCTGGCGCTG
GCCTCAACCGCAGCTGGGCGCCCGCTGTGGCTGTCTGCTTGGCCCGCGGCTTACACACCGCTTCGTGCA
GGCAACCGCTGCTGGCCACCGCTGGGAGACACTCCAGAGGAGATCTCTTGCTTCCCTGGGGTGCTA
CAGGAGATGGAAGTGAAGCTGTGGCGGAGCCACTGTCAATGTTCTACAGCGAGCGCGCTCCCTCTCAACC

[illegible]

FIG. 1UU

[illegible]

WO 02/00860

48/68

PCT/US01/20171

FIG. 1VV

CCAGGGCCCTCCAGCCGCAATGTTGGTGAGCTGTCTCCTCCGAGGGTGAGTGGCCATGGCAGGCCAGCCCTC
CAGGTTCCGGGTCGACATCTGTGGGGGGCCCTCATCGCTGACCGCTGGGTGATACAGCTGCCACTGCT
TCAGAGGAGACATGGCCCTCCACCGTGTGTGGACCGGTCTCTGGGCAAGGTGTGGGAGAACTCGCGCTG
GCTGGAGAGGTGCTTCAAGGTGAGCCCGCTCTCTGACCCGCTACCAAGAGAGAGAGCCCATGACTAC
GACCTGGCGCTGCTCAGCTCGACACCGGTGTGGCTGCGCCCGCTGCGCCCGCTCTGCTGCGCCCG
CGCTCCACTTCTCGAGCCCGCTGCACTGCTGATACGGGCTGGGGCGCTTGGCGAGGGCGGCCCC
ATCAGCAACGCTCTCGCAAAAGTGATGTGCACTTGTCCACAGAGAGCTGTGCGAGGGCTATCGGTACCA
GTGAGCCGACGCAATGCTGTGCGCGCTACCGCAAGGGCAAGAGGATGCCCTGTCAGGGTGACTCAGGTGGT
CGCTGGTGTGCAAGGCACTCAGTGGCGGTGTTCTGGCGGGCTGCTCAGCTGGGGCGCTGGCTGTGGC
CGCCTAACTACTTGGCGCTACACCGGCATCACAGGTGTGATCAGCTGGATCCAGCAAGTGTGACCTGA

>SGPR451_SEQID_55
GACCTGGCGCATCTTGTCTACCCAGCCCTCCAAATGCGGCTGGGGCTCCTGAGCGTGGCGCTGTTTGTGG
GGAGCTCTCACTTATCTCAGACCACTACTCGCCCTCTGAAAGGCACAGGCTCGCCCTCGCGGAAACCGGC
GGCTAGTTCACGAGGCTGAGCGCTCGCAAGAGGCTCGCGCGGAGGGGAGGGGCGCATGCAA
AGSATTTGGAACAGCACCGCTTAAGGATGTTGCAAGGGTCTCGGATTAGGGGCAACCGAACACAGCT
GGCGCATGGCGGTGGGTGAGCCCTGCAGATTAATATGCGCGTGTCTTGTTCATGTATGGGGAAACCT
AGTGAGAGAGGGTGGGTCTCAGAGCTGCCACTGCACATAAGACACTAGCGATCCTTTAATGTGGACAGCTG
TGAITGGAACTAATAATACATGAGCGCTATCTCATACCAAGATAAAATTAAGCAATCATTAATTCATCCA
AACTTCAATTTGGAACTTATGTAATGATATGCACTTTTCACTTAAAGAGGAGTATGACTATATT
CAGCCTATTGCTAGCTTTGATGTTTCCAAATCCTGACGGAACACAAAGTGTTTATAGTGGCTGGGAA
GAACAAAGAGAGGTACGCTACAATATTTACAAGATGCAGAAAGTGCATTAATTTCTCGAGAGATGTGTAA
TTCTGAGAGGAGTTATGGGGATAATTTCTAACACTTCAATTTGTGAGGTGATGAAGATGGAGCTTTTGATAT
TGCAGGGGTGACAGTGGGGGACCAATTAATGTCTACTACCAATATAAAGATTTTGTAAATGGAAATACC
AGTTACGGACATGGCTGTGTCGAAGAGGTTTCTGGTGTCTATTTGGGCCATCCTTCTACCAAAAGTGGCTG
ACAGAGCATTTCTTCCATGCAAGCACTCAAGGCATACTACTATAAATATTTACGTGGCCAGATCCTCATAGCTT
TATGTTTTTGTCTTACTAGCAACACATAA

>SGPR452_1_SEQID_56
AGCCCCCGGAGCCGACGAGCCCTGACTGTAGGCTCCAGGCTCCCTGGAAGCCCTGGCCACGCTCGGCCCG
CAGCCCTCAGACTGGCTGTGCTCGCGGATCTTGGCTGGTTCGAGGCTGATGGAGCTGCCCACTCCATGGGCC
TGGGCAGCAGCTTGAAGTGGGGGTGGGCCAAGCCCTCTGGGATGCCCGTCCGAGAGATGACCTGGTGGGCAT

WO 02/00860

49/68

PCT/US01/20171

FIG. 1WW

TGTGGGGGCCAATGCCCCCCGGGGAAGTGCCCGTGGCAGGTGAGCCTGAGGGTGTACAGCTACCACTG
GGCTCTGTGGCGCACATCTGTGGGGGCTCCCTCATCCACCCAGTGGGTGCTGCTGCTGCTGCCACTGCATT
TTCTGGAAGGACACCGACCGTCCATCTACCGGATCCACGCTGGGACGTGTATCTCTACGGGGGCGGGGGC
TGCTGAACGTCAGCCGATCATCTCCACCCCACTATGCTACCTCGCGGCTGGGTGGGATGTGGCCCTGCT
CCAGCTGCGGGGTCACCTCTCTCCACAGTCTGCTGCCCGCCCTACCGCTTCGAGCAGGGAGTGTGCA
GGTCTGGAACCGCTCTGTGAGCAGCCCTACCGACGCTCTGAGGACACTGGCGCACCTGGCGACGCTCATC
CTGGATGACATGCTGTGTGCGGAGCGAGGCGGAGACTCTCTGCTACGGTACTCCGGCGGCTCTGCTCT
GCAGGCTGCGGGGCTCTGGCGCTGTGTGGGGTGTGAGCTGGGGCTACGGCTGTACCTGCGGACTTTC
CCGGCGTCTACACCCACGTCAGATCTACGTCTGATCTCTGATCTCTGCAAGTCCGGGAGTTGCCCTGA

>SGPR504_SEQID_57
ATCGGGGGCCACGAGGTGACCCGCCACTCCAGGCGCTACATGGCATCGGTGGGCTTCGGGGGCCAACATCACT
GCGGAGGCTTCCTGCTGCGAGCCCGCTGGGTGCTCTCGGCCGCCAGTCTTCAGCCACAGG

>SGPR469_SEQID_58
GGAGATTCTGGGGGGCCCTTGCTGTGTGAATTAATGGCACATGGGTCCAGGTGGGGATTGTGAGCTGGGGCA
TTGGCTGCGGTGCAAAAGGATACCTCGAGTTTACACAGAGTTAGTTTCTACAAGAAATGGAAT

>SGPR400_SEQID_59
ATGGCAGGAGACAGGTACCGCCAAATGTACGAGATACCCCTGGACAGAAACGATGCTCTTCTGAAAAAC
ATTTCTCTTTTATAGAGCATCACTCTTGTGTTGTAACACACAGATCCAAATATAGTCTGGCGAGCTTT
GAGAGTCAGGTTCTCCCGATTGACAGCAGGGGACGCCGACACCCCAATTCCTCCCTTGGGTCTCTGGAG
CTGCATCTCTGAAGTCTGCTCCTTCAGGATCTGGCAGGGGTGAAGACCAAGAGAGGGGACACAGAGA
CACGGCACTGCTGGCTATGCATTCAGCTGCTCTTCTGCTGGGATTTCGGGTGAGCCCCCAGAAATGGGTCT
GTGGCGGCCACAGTCTCATCTGGTATTGCCCTCAGGCTTGGGGCTAGTGTGGGCGAGTGGCCCTGGCAGGT
CAGCATCCCGCAGGCTTGATTACGCTGCTCAGATACCCCTCATCTCAGAGAGTGGGTGCTGACAGTGGCGA
TCTGCTTCCCATATCCCGCCACCCTGATTCCAAAGCAACACATCTAGTGCCATGCTGTGTGAACTGCCCT
CCCCAGTTCTGTAGCCCTGTGTCTGCTCATCTGCCCTCCCTCATCTGAAGTCTACCTGAAGAATAACAA
CTCCTGCTGGGTGACTGGATGGGGCTACTGGAAATATCCAAATATCAAGCGTTCTTATACACTGAAGAGCT
GAAAGTGGCCCTCATTTGATCTCCAGACATGCGGTGACCACTATCAAAATGAATCTTCTGACAGGAGTTGAGCT
CATCATGAGTGAAGCTATGATCTGCTCCAAGCTCCCAAGTGGGGCAGATGGATCAGTGTACTGTAGAAATCCAGC
CCTCAGGCACCTTTCACAGGCGCTTGGCTTCCCCAGTGA

WO 02/00860

50/68

PCT/US01/20171

FIG. 2A

>SGP437_SEQID_80
MKCLGKRRGQAAAFPLCWLFLKIQPHSHLYNNRYAGDKVIRFIPKTEEEAYALKSYQLKVDLWQPPSSISYVSEG
TVTDVHIPQNGSRALLAFQEAQNIQYKVLIEDLQKLEKSSSLHTQRNRSLSGYNVEVYHSLLEIQNMWHLNKHHS
GLHMFSGRSEGRCLFILKLGRSRSLKRAVWIDCGHAREWIGPACQWYFKEALLTYKSDPAMRKMLNHLVEYIMP
VFNVDGYHFWTNDREWRKTRSRNSRFRRCRGVDANRNWKVKCGKFGTNWDPPKVSAGFTLQNMSPEDSHGR
LMFFCM

>SGP413_SEQID_61
MKPLLETLYLLGMLVPGGLGYDRSLAQRQENVKSVPSWSLETYSYNIYHPMGEIYEMREISEKYKEWTQHFLGV
TYETHPMYYLKISQSPGNPKKIIMDCGHIAREWIAFPACQWFWKEILQNHKDNSSIRKLLRNLDYVLPVLNIDGYIT
WTTDLRLWRKSRSPHNGTCFGLDLNRNFNASWCISGRNQDQTCGCTGPVSEPETKAVAFIESKDDILCFLTM
HSYGQLILTPYGYTKNKSSNHPMIQVQKAAALKAQYGTNYRVGSSADILYASSGSRDWARDIGIPFSYTFELRD
SGTYGFVLPEAQIPTCEETMEAVLSVLDDVYAKHWHSDSAGRVTSATMILLGLLVSCMSLL

>SGP404_SEQID_62
MVSNDSTWYTVTKNGSGDMIFEGNSEKEIPVLNELPYPWVARYIRINPQSWFDNGSICMRMEILGCLPDPNNYYHR
RNEMTTDDLDLDFKHNYKEMRQLMKVYVNEPCNITRIYNIQKSHQGLKYAVEISDHPGEHEVGEPEFHYAGAHGNE
VLGRELLLLLVQVCOEYLARNARIVHLVEETRIHVLPSLNPDGYEKAYEGSELGGWSLGRWTHDGINNNFPDLNT
LLWEAEDRONVPRKVPNHVIAPEWFLSENATVAAETRAVWMEKIPVLGGNLQGGELVYAYPYDLVRSFWKTQEH
TPTPDHVRWLAYSASTHRLMTDARRVCHTEDFQKEEGTVNGASWHTVAGSLNDFSHTNCFELSIYVGCDK
YPHESQLPEEWENNRESLIVFMEQVHRGIKGLVRDSHGKIPNAISVGINHDIRTANDGDYWRLLNPGEYVYAKAE
GFTASTKNQMVGYDMGATRCDFTLSKTNMARIEMEKEFGKQPVSLPARRLKLRGRKRQRG

>SGP636_1_SEQID_63
MMWRCPGLGLLLPLAGHLALGAQQGRGRRELAPGLHLRGIRDAGGRYCOEQDLCCRGRADDCALPYLGAICYDLF
CNRTVSDCCPDFWDFCLGVPPFPPIQGMHGGRIYPVLGTYWDNCNRCTCQENRQWQDQEPCLVDPDMIKAIN
QGNYGWQAGNHSFVGMTLDEGIYRLGTIRPSSVNMNMHEVTVLNPGEVLPFAFEASEKVPNLHEPILDQGNCA
GSWAFSTAASDRVSHSLQHMTPVLSQNLSCDTHQQCGCRGGRLDGAWWFLRRGVVSDHCYCFPSGRERDE
AGPAPPCMMHSHRAMGRGRQATAHCPNSYVNNNDIYQVTPYRLGSDNKEIMKELMENGVPQALMEVHEDFFLYK
GGIYSHTPVSLGRPERYRRHGTSHVKITGWGEETLPDGRTLKYWTAANSWGPWAWGERGHFRVIRGVNECDIESFVL
GVWGRVGMEDMGHH

WO 02/00860

51/68

PCT/US01/20171

FIG. 2B

>SGP414_SEQID_64
 MCENCADLVEVLNEISDVGGDLQLRKEHTLKIFTYINSWTQRQCLCGFKEYKHLEIFNQVWALINLVIAQVQVLRD
 QLCKHCTTINIDS TWQDESNOAEEPI NIDRECNEGSTGEROKSIEKKSSTRICNL TEESSKSSDPSLWSTDEKELL
 LCVAKFIQIFPLYTAYKHNTHPTIEDISTQESNILGAFCDMNDVEVPLHLLRYVCLFGCKNGLSLMKDCFEYGTPTPLP
 FLIAHAFTVYVSNIRWLHIPAVMQHIIPRTYVIRYLCKLSQDELQSGAARNAADLMWSTVREPLDITLTCDFKESLDLAF
 KYFMSPTLTMRLAGLSQITNQLHTFNDVNCNIESVSDTETSIKELADWILISNNVVEHIFGNLHIEIKQCOVILNFLAAE
 GRLSTQHIDCIWAAALQKHCERYHDLFPLKILDPVLRLLNLVSALEPSVHTQTLVLSMILKALWNAALAAQAQ
 LSKQSSFASLLNTNIPICNKEEEELRETAPSPWSPASPOSSDSNDTHQSGGSDIEMDEQLNFTKHVQORLSDTEE
 SMQSSDETANSGEDSGSGGSGHSDGSSNEVNSHASQAGSPGSEVQSEDIADIEALKEDEDDDDHGHNPP
 KSSCGTDLNRNRLKESQAGICLGDQGTSEINGTSGTGKDLVFNTESLPSVDNRMRLDACSHSEDPEDHISGEMN
 ATHIAQSGESCITRTGDFLGETIGNELNCRQFQHPQH HHHHHHHHHDHGMVDMLSADDSVSCSSQVSAKSEK
 NMADFDGEESGCEEELVQINSHAEI TSHLQOHLNLA SIYHEHL SOGPVYVHKHOFNSNAVTDINLDNVCKKGNLTLW
 DIVQDEDAVNLSEGLINEAEKLLCSLVQWFTDQIRMRHIEGLENLGNIRSVVISLRLLPKLFGTFQFGSSYDTHWIT
 MWAEKELNMKLFNDNLVYITQTVREGQKHALYSHSAEVQVRLGFLTCVFTSLGSPDHFRLSLEQVDILWHCLVED
 SECYDDALHWFLNQVRSKDOHAMGMETTYKHLFLEKMPQLKPTISMTGLNLFQHLNCLARLATSAYDGCNSSELCG
 MDQFWGIALRAQSGDYVRAAIQYINSYINGKTGLEKEQEFISKCMESLMIASSLSLEQESHSLMVERGLMLKTHLE
 AFRRRFAYHLROWQIEGTGISHLKALSDKQSLPLRVQOPAGLPDKMTIEMYPQDVADLRAEVTHWYENLOKEOI
 NQQAQLQEFQGSNRKGEFPGGLMGPVRMISSGHELTTIDYDEKALHELGFKDMQMVVYSLGAPRERKGEVQLPA
 SOLPPQKDNIPMLLLQEPHLLTLFLLLEMLASFPPSGK/AVDDSESLRCEELHLHAENLRRVWELLMLLPTCPN
 MLMAFQNIIDEQSKAQSDHRSRHEVSHYSMMWLLVSWAHCCSLVKSADSDHLQDWLKKLLTLLIPETAVRHESCG
 LYKLSLGLDGGDSINRSFLLAASTLLKFLPDAQALKPIRIDDYEEEPILKPGCKEYFWLLCKLVNHIKIDASQTTLLDL
 DALARHLADCIRESREILDHODGNVEDDGLTGILLRLATS/VV/KHPPEKFSREGQEFRLQIFNLLFLLPSLKORQOPKCKS
 HSSRAAYDLLVEMVKGSVENYRLHNWMAQHMQSHAPYKWDYWPHEDEVRAECRFVGLTNLGATCYLASTIQQLY
 MIPEARQAVFTAKYSEDMKHKHTLLELQKMFYTLMESECKAYNPRPFCCTYTMQKQPLNTGEQKDMTEFFTDLTIKIE
 EMSPELKNYKSLFGGVITNNVYSLDCEHVSTAEFTYVRCQVADMKNIYESLDEVTKDTLEGDMNYTCSCQKKY
 RAEKRAFCFKLPRIXSFTNTMYTFNMVTMMKEVNTHFSPRLDMTPYEDFLMGKSERKEGFEVSDHKSQSESY
 EYDLIGVYHTGTADGGHYYSFIRDVNPWAYKNKWYLFNDAEVKPFDSAQLASECFGEGMTTKTYDSVTDKFMDFS
 FEKTHSAYMLPYKRMEPEEENGREYKFDVSSLELWVHDNMQFLQDKNIFEHTYFGFMWQLSCSPSTLPDPKAVS
 LMTAKLSTSEVLETFIHSKEKPTMLQWIELLTKOFNNSQAACEWFLDRMADDDWWPMQILKCPNQVIRQMFQRLCIH
 VIQRLRPVHAHLYLQPMEDGSDMDTSDVEDIGRSCVTRFVRTLLIMEHGVKPHSKHLYEYFALYEFAMGEEES
 QFLLSLQAISTMWYFMGTGKGPENQVEVLSEEEGEEEEEDLSLAEKYPALAEKMLVALLVEQSRSEHRLTL
 SOTDMAALTGGKGFPLFQHIRDGINIRQTCNLIFSLCRYNNRLAEHVSMLFTSAKLTPEAANPFKLLTNLMIEFAGG

WO 02/00860

52/68

PCT/US01/20171

FIG. 2C

PPGMPFFASYILQRWEIEYNPSQCLDWLAVQTPRNKLAHWWLQNMENWVERFLAHNYPVRRTSAAYLLVSLIPS
NSFRQWFRSTRSLHIPTRDPLSPDTTTLVHQVYVNLGLLSRAKLYYDAAVHGTTLKVPYFSTYCLISKTEKLMFS
TYEMDLWNLFOPKLSEPAATNHNKQALLSFYVNVACDCPENLIVQNPVYTKNAFNYLADHDDQDVVLFNRGMPL
AYYGILRCCQSPAFTRQLASHQIQWAFKNLTPHASQYPGAVEELFNLMLQFIAQRPMDEEELDKQFKTTISC
YLRCLDGRSCWTTLSAFRILLESDEDRLLVFNFRGLIMTESFTLHMYHEATACHVYHDLVELLSIFLSIKSTRPY
LQRKDVQALIQWQERIEFAHKLTLINSYSPPELRNACIDVLKELVLLSPHDFLHTLVPELQHNHCTYHHSNIPMSLGP
YFPCRENKIGKSNIRPPPELNMCLPTMTVETSGKGDVYDRMLDYFYSYHQFHLCRVAINCEKFTETLVKLSV
LVAYEGLPLHALFPKLWTELCOQASAMSKNCIKLLCEDPFAEYKCILMDERTFLNNIVYTFMTHFLKVQSOVFE
ANCANLSTLITNLISQYQNLQSDFSNRVEISKASASLNGDLRALALLSVHTPKQLNPALPTQLLELSCORTQQRNS
LOEQEAKERTKDDGATPIKRRRVSSDEEHTVDSCIDMKTREVLTPTSTSDNETRDSIIDPQTEQGLPSPENS
VKEYRMEVPSPFSFSDMSNIRSQAEQNNRYDDCKEFKDLHCKSDSTLAESESEFSTTSISAVLSDLALRSCDG
QALPSQDPEVALSLSCGHSRGLFSHMVQQHDLDTLCRTIESTHVVTRISGKGNQAAAS

>SGP430_SEQID_65
MSPLKIHGPIRSMQTGITKWKEGSFEIVEKENKSVLVHYNTGGIPRIFQLSHNIKWVLRPSGAKQSRMLTLTQDNS
FLSIDKVPKDAEEMRLFLDAVHQNRLPAAAMPQSGSGFAGILGSRSTQKTSRQLSYSDNQASAKRGSLETKDDIP
FRKVLGNPGRGSIKTVAGSGIARTIPSLTSTPLRSGLLENRTKRMISTGSELNEDYPKENDSSNNKAMTDPSPR
KYLTSREKQLSLKQSEENRTSGGLPLQSSSFYGRAGSKEHSGGTNLDRTNVSSQTPSAKRSLGFLPQVPVLSV
KKLRQNDYTGWNKPRVPLSSHQQQLQGFNLGNTCYMNAILOSFLSQSFANDLLKQIPWKKIPLNALIRFAHL
LVKKDIGNSETTKDILLKVKNAISATAERFSGYMQNDHAHEFLSQCLDQKEDMEKLNKTKWTEPVSGEENSPIASATR
AYTCPVITNLEFEVQHSICKACGEIIPKREQFNDLIDLPRKKPLPPRSIQSLDLFFRAEELEYSCCKGKCALVRH
KFNRLPRVILJHLKRYSENVVALSLNNKIGQVPIRYLTLSHCTENTKPPFTLGSAAHMAASRPLKASQMVNSCITSPS
TPSKFTFKSSSLALCLDSDEDELKRSVALSQRCEMLGNEQQQDELEKOSKLCPIEPDKSELENSGFDMSSEEL
LAAVLEISKRDASPSLSHEDDDKPTSSPDTGFAEDDIQEMPENPDTMETKPKTITELDPASFTEITKDCDENKENKTP
EGSQGEVDWLQYDMEREREQELQALQSLQEQAWEQKEDDLKRATELSLQEFNNSFVDALGSDSDSGNE
DVFDMEYTEAEAEELKRNAETGNLPHSYRLISVSHIGSTSSSGHYSDVYDIKKQAWFTYNDLEVSQIEAAVQSDRD
RSGYIFFYMHKEIFDELLETEKNQSLSSTEVGKTTQAS

>SGP496_1_SEQID_66
MTLLAPWYTGPMIPMDVNEPSSVTTAPTLLSSSLQHISSFLATGKKLSLHFGHPRECEVTRIDDKNRRGLEDESPGAKL
FNNDGVCCCLQKRGPNVITSVCSPRTLQISVFLSEKYGIVKFSDELPGVIGSNIGDAHFQEFRAGISWKPWDP
DDPIQFPDCCSSSSSRIPSVSVLVAVPLVAGHKGOAFIERMLGCFKELQELTOEGPGGGHPRSAWPPRRHAQWP

WO 02/00860

53/68

PCT/US01/20171

FIG. 2D

PEPCQGEPPPPVEAEVEEAETAEAKERYEAEAKVEGAFAAGKAVDATEKIVETAGKVDAAGKVETAEQ
 PGRAELKLEPEPEPEAEQEPKQELLEDENPARSGGNSDEVPPPTLPSPDPDPDPSPRSPRPPRPQ
 TRLRTPQPRPPRPPRPPRGFGGCLDVFAVGPFGCSHVNSFKVGENWFWQELRVYQCFWGCVPETRKSKA
 KSCICHVCGTHLNRHSCLSVFFGCTEKHEHAETKQHNLAVDLYYGGYCFMCKDYVYDKIEQAKEEQGEALK
 LOASTSEVSHQQCSVPLGKEKFTWETTKPELELLGHNRRRTTSSFTIGRLGJNLGNTCFMNCIOVALTHTPILRD
 FFLSDRHRCEMPPELCLVCEMSFLRELYSGNPSHPYKLLHLVMIHARHLAGYQQDAHEFLAALDVLHRRCHKG
 DDVGKAANNPNHNCIIDIQFTGGLOSDVTCQACHGVSFTTIDPCWDISLDLPGSCTSWPMPSPRESSVNGESHIPGI
 TTLTDCLRRFTTRPEHLGSSAKIKGSCQSYQESTKOLTMNKLPAVACFHFKREHSAKORRKITTYISFPLELDMTPFM
 ASSKESRMINGQLQLPTNSGNNENKYSLFAVNHQGTLESQHYTSFRIRHHKQWFKCDDAVITKASIKDVLDSGYLLF
 YHKQVLEHESEKVKEMNTQAY

>SGP495_SEQID_67

MRVKDPTKALPEKAKRSKRPTVPHDESDSDIAVGLTCQHYSHASVNHVKRAENLWSV/CSECLERREFYDGLVL
 TSDIWLCLKCGFCGCKNSESQHSLSKHFKSSRTEPHCIINLSTWIIWCYECDEKLSHONKKVLAQVDFLQKHASKT
 QTSAFSRIMKLCCEKCEIDEIQGKGKRNLSVRGITNLGNTCFNAVMQNLAQTYTLTDLMEIKESSTKLKIFPSSDS
 QLDPLVVELSRPGLTSALFLHSMKETEGPLSPKVLFNQLCQKAPRFQDQDQSDQLHYLLDAVTEETKRIQ
 ASILKAFNPTTKTADDETRKVKISTVKDPIDISLPIIEERYSGKPLWGRMKNYRSLRETDHRYSGNVNIENIHOPRA
 AKKHSSKDKSOLJHDKIRKLSGGETVYQKNENLEMNGDSLMAFASLMNSESRLNESPTDSEKEASHSESNDVA
 DSEPSSESASKQTGLFRSSSGSGVQDGPLYPLSAGKLLYTKETSDGDKMAEASBELRLSTVYTGQDQDFDRENQP
 LNIISNLCFLEGHKHLRSYSPQNAFQTLSSQSYITTSKECSIQSLYQFTSMELLGNKNKLLCENCTKNKQKYQEETSFAE
 KKVEGYTTNARKQLLISAVPAVLILHLRFRHQAGLSLRKYNRHYDFPLMLDLAPFCSATCKNASVGDKVLGYLGIVEH
 SGSMREGHYTA YKVRTPSRKLSEHNTKKKNVPLKKAADSESAGQWVHVSDTYLQVVPESRALSQAAYLLFYERV

>SGP407_SEQID_68

MEYVPYFRSPNRTLIPERIWSNPLLVLIAYKTVSWPRQLLAKQANKWMPFVPSKTLPWDPLELKYQQNRYPYS
 PDPNFTFLRLCLNAFSAAVFYLPQPSWHKPEGLKPAGYPRVPDIPYSGSYTLKSTTEAAGLHQSLPMVQLPLHPTKG
 SALLKESELNDADWANLMMWKRYLEEQEDSKMVDLFVGOMKSYLKCQAGYHSMTFKVFQDGLSLTPKKGFGAGK
 VSLRDCLSLFTKEELELENAAGTLPTKSEVLSTSCVPFGTTQAASTVATTQPCASARLVGTFTMTLVSPNLTRDTE
 GIELTYMKALVLIDLFRKASTDILLFNHDSGSKWKRLPEPGGLEKKHEELRLRPLKEEYHWLVPLVPLKLTGSPHRWRP
 RKRALASCSWCLQRYTMMRYMGVQDKAGNRNQMLLGGPRVIGDTSVNSQTTTRDKACRPPSHSVTSQSFVACL
 DPDFLYGHQSYWMKAHLNDLILREGPYTMAQSFYWGFPAGGNLSALEMLPDGPAPRTFLQKKSCLFPLFSYILLH
 KAGKLFQPDAGHGLVKKHVAPTRGIVIMEPRQLGGKGSLSKLQACALGGMNSGMEPQKSAPFAAGKGLAPPLPVC

WO 02/00860

54/68

PCT/US01/20171

FIG. 2E

NLRFKLRVYKFEELWSRAGLGKSDNHSSRQMPWGAAGVACQHPCKLPRIVAEI TPPKLSFGFLNTVQSSVLP TSL
 SQFFLNDSPPEAIPQSLPGSPRTNSFPKDFPKDKLVLSLLTMYELDRLF

>SGP453_SEQID_69
 MLAMDTCKHVGVQLQLAQDHSNLPQKVHCVDCNTTESIIWACLSQSHVACGRVIEEHALKHFQESSHPVALEVNEMY
 VFCYLCDDYVLNDNATGDKLLRRTLSAISKQNYHCTTRSGRFLRSMGTGDDSYFLHDGAQSLQSEDQLYTALWHR
 RRILMGKIFRTWFEQSPIGRKQEEPFQKVVVKREVKKRQELVQVKAELSMPPRKSLRLQGLAQSTIEIVSVQVP
 ACTPASPAKDKVLTSENEISQKVSQSVKRRPVTGVTGLRNLGNTCYMNSVLOVLSHLLFRQCFKLQDLNQWLA
 MTASEKTRSCKHPPVTDVYQNMNECQEKDTGFVCSRQSSLSGSLSGGASKGRKMEIQKPTSQYISLCHLHTL
 FQVMWSGKWALVSPFAMLSHWRIPAFRGYAQQDAQEFCELLDKIQRELETTGTSLPALIPTSORKLIKQVNLNVN
 NIFHGQLLSQVTCACDNKSNITIEPFWLDSLEPPERYQCSGKDIAQPCLVTEMLAKTETETEALEGKIVYCDQCNSKRR
 RFSSKPVVLTEAQQLMICHLPQVLRHLKRRFRWSGRNNREKIGVHYGFEEILNMEPYCQRETILSRPECFYDL SAV
 VMHHGKGFSGSHYTAICYNSEGGFWHCNDSKLSMCTMDEVCKAQAAYLFYTORVYENGHSHKLLPPELLLSQHPN
 EDADTSSNEILS

>SGP445_SEQID_70
 MRVKOPTKALPEKAKRSKRTVPDHEDSSDDIAVGLTCQHYSHAISVNHVKRAIENLWSVCSECLEERRFYDGQLVL
 TSDWILCLKGFGQCGKNSQSHSLKHKFSRTEPHCIINLSTMIWCYECDEKLSHCHNKVLAQIVDFLQKHASKT
 QTSFASRMKLCCEKCEETDEIQKGGKORNLVSRGTNLGNTCFHNAVMOQLAQTYTLTDLMEIKESSTKLKIPFSSDS
 QLDPLVVELSRPGPLTSAJFLFLHSMKETEKGPLSPKVLFNQLCQKRVHLHL

>SGP401_1_SEQID_71
 MTVRNIAICNMGTNVALEKIDIGPEQFPINEHYFGLVNFNGNTCYNSVLQALYFCRPFRENVLAYKAQKKKENLLTC
 LADLFHSIATQKKVGVIPPKKFSIRLKENDLFDNYMQQDAHEPLNYLLNTIADLQEEKKQKQNGKXNGNMNEPA
 ENNKPELTWWHEIFQGYLTNETRCLNCEYTSKDEDFDLSDVDEQNTSITHCLRFSNTETLCSEKYYCETCCSKQ
 EAQKRMVKKLPWVLAHLKRFKYMELRRTYTKLSYRVVFPLELRLFNSTSDAVNLD RMYDLVAVVWCHGSGPNRGH
 YTTIVKSHGFWLLFDDDIVKIDAQAEIEFYGLTSDISKNSSEGYILFYQSRE

>SGP408_SEQID_72
 MYPGEENQLVPKEAPLDHTSDKSLLDANFEPGKNFLHLTDKQGEQPKILLEDSSAGEDSVHDFRIGPLPREGSVGST
 SDYVQSYSYSSILNKSETGYVGLVNOAMTCVYLSLLQLTLMTFERNALYKVEFESEEDPVTYQLRFLVLLQT
 SKKRAIETTDVTRSFQWDSSEAWQQHDVQELCRVMFDALQKWKQTEQADLINELYQGLKLDYVRCLECGYEGWRI

FIG. 2F

DTYLDIPLVIRPYGSSQAFASVEEALHAFIQPEILDGPNQYFCERQKKKCDARKGLRFLHFPYLLTLQLKRFDFDYTTMH
RIKLNDRMTPEELDMSTFIDYDEKSPQTESCTDSGAENEGSCHSDQMSNDFSNDDGVDEGICLETNSGTETKISKS
GLEKSLIYELFSVMAHSGSAGGHYYACIKSFSDQWYSFDDQHVSRITQEDIKKTHGGSSGSRGYSSAFASSTNA
YMLIYRLKOPARNAKFLVGEYPEHIKILVQKERELEEQEKQREIERINTCKIKLFC LHPTKQVMMENKLEVHKDKTLK
EAVEMAYKMDLLEEVIPLDCORLVKYDYDEFHDYLERSEYEGEEDTPMGLLGGVKSITYMFDLLLETRKPDQVFQSYKPG
FVEMKVHVVDLKAESAAPITVRAYLNQTVTEFKQLSKAIHLPAETMRKILVERCYNDRLLSVSSKTLKAEGRFRSNKY
FVESSETLDYQMAFADSHLWKLDRHANTIRL FVLLPEQSPVYSKRTAYQKAGDSGNVDDDCERVKGPVGSLSKV
EALIEESTELKLSLSLQOQDGDNGDSKSTETSDFENIESPLNERDSSASVDNRELEQHIQTSDPENFQSEERSDS
VNDRSTSSVSDILSSSHSDTLGNADNAQIPLANGLDHSITSSRRITKANEGKKTWDTAEEDSGTDSYDESGKS
RGEMQYMYKAEPYAADEGSGEGHKWLMVHVDRITLAAFKHLEPFGVLSHFVKVFRVYASNQEFESVRLNETL
SSFSDDNKITIRLGRALKKGEYRVKYVQLLVNEQEPCKFLLDVAFKGMTVRQSKKEELIPQLREQCGLELSIDRFLRK
KTKWNPVTVELDYHIYFEEDINISSNWEVLEVLGVGEKMSQSLAVLSRRWKPSMKLDPFQEWLESSSVDELREK
LSEISGIPLDIDIEFAGRGTPPCDISVLDIHQDLDWNPVSTLNVWPL YICDDGAVIFYRDKTEELMELTDEQRNELMKK
ESSRLQKTGHRVITYSPRKEKALKIYLDGAPNKDLTQQ

>SGP480 SEQID_73
MGAKESRTIGFLSYEEALRRVTDVVELKRLDAFKRTGCLSYIMGQHCIFREVLGDGVPPKVAEVIYCSFGGTSKGLHFN
NLUVGLVLLTRGKDEEKAKYIFLSFSESGNYVIREEMERMLHVVDGKVPDTLRKCFSEGEKYNYEKFRWVLELNKDAF
TFSRWLLSGGVYVLTDDSDTPTFYQTLGAVTHLEESDIIDLEKRYWLLKAOSRTGRFDELTGFLVSPPIRPSLEGL
FNAFDENRDNHIDFKEISCGLSACCRGPLAERQKFCYFVDVDRDGYLSRVELRDMVVALLEVWKNRDTDDIPELHM
DLSDYVEGILNAHDTTKMGHLTLEDYQIWSVKVNLANEFLNLLFQVGHVILGLRPATPEEGQGIIRGWLERESRYGLQA
GHNWFIISQWQWQWKEVYKYDAMPVIEPSSVINGKYSFGTAAHPMEQVEDRIGSSLSYVNTTEEFSDNISTAS
EASSETAGSGFLYSATPGADVCFARQHTSDNNQCLLGANGNILLHLPKPGADNQPLVTQEPVKATSLTLEGGRL
KRTPLQIHGRDYEMVPEVWRALYHMYGANKALPRVVKNSKTDIPELELFPYLLFRQQPATRTQOSNIWVMGNV
PSPNAPLKRVLAYTGCFSRMQTKIEIHEYLSQRLRIKEEDMRLWLYNSENYLLTDDDEHKLEYLKIQDEQHLVIEVRNK
DMSWPEEMSFANSKIDRHKVPTKEGATGLSNLNTCFMNSSIQCSQELLAFELDLGLHEDLNRYHEKPYVELKDSQGRP
AKCYGDILVQELWSGTOKNVAPLKLWTTAKYAPRENGFQQQDSQELLAFELDLGLHEDLNRYHEKPYVELKDSQGRP
DWEVAAEAWDNHLRNRNSWDLFHGQLRSQVKCKTCGHSVRFDFNFNLSLPLMDSYMHLEITVVKLDGTTVPVRYG
LRLNMDEKTYTGLKQSLDGLNSEQILLAEVHGSNIKNIFPDQDNQKVRLSVSGHLCFAEIPVVPVSPISASPTQTDSS
SPSTNEMFTLTNGDLPRPIFINGMPTNVPCGTEKNTNGMVGNGHMPSLPDSPTFTGYIAVHRKMMRTLEYFLSSQ
KNRPSLFGMPLIVPCTVHTRKDLYDAVWQVSRLASPLPQEAASHAQDCDSMGYQYPTFLRVQKDGNSCAWC
PWYRFORGCKIDCGEDRAFIGNAYIADVWDPTALHLRYQTSQERVVDEHESVEQSRRQAEPINLDSCLRAFTSEEL

WO 02/00860

56/68

PCT/US01/20171

FIG. 2G

GENEMYCSKCKTHCLATKKLDLWRLPILIHILKRFQFVNGRWIKSQKIVKPPRESFDFSAFLVPRDPALCOHKPLTP
QGDELSEPRILAREVKVDAQSSAGEEDVLLSKSPSSLSANIISPKGSPSSSRKSGTSCSPKSSNPNSSPRTLLGRS
KGRLRLPOIGSKNKLSSKENLDASKENGAGQICELADALSRGHVLGGSOPELVTPQDHEVALANGFLYEHEACGNG
YSGQLGNIHSEEDSTDDQREDTRKPIYNYAISCHSILGGGHYTYAKPNCKWYCYNDSSCKELHPDEIDTDSAY
ILFYEQGIDYAOFLPKTDGKKMADTSSMDDEDSEYKCYVLQ

>SGP431_SEQID_74

MDKILEGLVSSHPPLKRVIVRVKVESAEHMLDEAQCEAMFDLITRILEGQDPFORQVGHQVLEAYARYHRPEFES
FFNKTFVLGLLHQGYHSLDRKVDVALDYIHNLKJMSCPVLDLFLSLQVEVLRMWCERPEPQLCARSLDLTDFVQCI
PKGKLSITFCQQLVRTIGHFCVSTQERELREYVSVQTKVSNLLQNIWKAEPATLLPSLQEVFASISSTDAFSPSVALA
SLVQHPLQMITVLIIRSLTTDPNWKDASMTQALCRMIDWLSWPLAQHVDVTWVIALKGLAAVQKFTLIDVTLKIELVFN
RLWFPLVRPGALAVLSHMLSFQHSPEAFHLVPHVWNLVHFKNDGLPSSTAFVLVQLTELHMMYHYSGFDPLEYPI
LEAKDFPKPSEEEKILINQSAWTSQSNLSASCLSRLSGKSETGKTGLNLGNTCYMNSVQIALFMATDFRRQVLSNL
NGNSLMWKLQHLFAFLAHTQREAYAPRIFFEAASRPWFTPRSQQDCSEYLRLLDRLHEEEKILKVAQSHKPSSELEC
SETSLQEVASKAAVLTEPTRTSDGEKTLIEKMFGGKLRTHIRCLNCRSTSQKVEAFDLSLAFPCPSSLENMSVQDPAS
SPSIQDGLMQASVPGPSEEPVYNYPTTAAFCDSLVNEKTIQSPNEFYCESENTSVPNESNKILVNKQVPQKPGGET
TPSVTDLLNYFLAPEILTGDNOYCENCASLQNAEKTMOITEEPEYLITLLRFSDQKYHVRKILDNVSLPLVLELPVK
RITSFSSLSSESWSVDFTDLSENLAKKLKPSTIDEASCTKLVYLLSSVVVHSGISSESGHYYSYARNITSDSSYOM
YHQSALALASSQSHELLGRDSPSAVFQDLLENKEMSKEWFLNDSRVFTFTSFQSVQKITSRFPKDTAYVLLYKQHS
NGLSGNNPTSLGWINGDPPLQKELMDAITKDNKLYLQEQELNARARALQAASASCSFRPNQDNDNDPPGSCGPTGG
GGGGGFNTVGRLVF

>SGP429_SEQID_75

MAPRLQLEKAAWRWAETVRPEEVSQEHETAYRIWLEPCIRGVORRNCKGNPNCLVGIGEHWLGIDENSFHNDDP
NCERRKKNISFVGLTNLGATCYNYNTFLQWVFLNLELRQALYLCPSSTCYMLGDQIEEKDYEPQITCEHLQYLFAALLQ
NSNRRYIDPSGFVKGALGDTGQQDQAQEFKSLFMSLLEDTLKSGKNPDVRNINQQQFCGEYAYVTCVQCGRSKLL
SKFYELNIOGHKQLTDCISEFLKEEKLEGNRYFCENCQSKQNAKIRLLSLPCTLNQLMRVFDQRQGHKKLN
TYIGFSELDMEPYVEHKGGSYVELSAVLIHRGVSAVGHYAHKDPQSGEWYKFNDEDEKMEGKKLQGLIEEDLE
PSKSQTRPKCGKGTCHSRNAYMLVYRLQTEKPNITTVQPAFLQELVDRDMSKFEWECIENAMERKQSDVKGAK
HEEVKELYORLPAGAEPEYFVSLWQKWLDESTPTKPIDNHACLCSHDKLHPDKISIMKRISYAADIFYRYGGGPR
LTVKALCKECVVERCIRLRKNQLEDYKTVANNLLKAAVKGDFWVSGKSLRWRQLALEQLDEQDGAEQSGNGKM
NGSTLNKDESKEERKEEELNFNEDILCPHGELCISENERRLVSKEAWSKLQYFPKAFEPFYSKECCSQCKILERE

WO 02/00860

57/68

PCT/US01/20171

FIG. 2H

EENEALHKMIANEOKTSLPNLFODKNRPCLSNWPEDTDVLVYSOFFVEEWRKVRKPTRCSPVSSVGNALLCPHG
 GLMTFASMTKEDSKULMPSEWOMIOKLEVDHVIKTRIEVGDVNPSETQYISEPKLCPECREGLCQQQRDLREY
 TOATYVHKVVDINKVMKDSAPELNVSSSETEEDKEAKPDGKDPDNQSGNGGTRKQKSHQNTIAYQKQVIRRM
 RHRYRGEKALLVSNQTLKELKIQIMHAFVAPFDQNLSDGKILSDDCATGLTGLVIPESVILLKADEPIADYAAAMDDV
 MQVCMPEEGFKGTGLLGH

>SGP503_SEQID_76
 MLSSRAEAAMTAADRAIQRFRTGAARYKYMKNWGVGGIAAALAGYVWGPITERKKRRKGLVPGLVNLGNTCF
 MNSLLQGLSACPAFIRWLEEFTSQYSRDQKEPPSHQYLSLTLLHLLKALSCQEVTDDEVLDSCLLDVLRMYRWQISS
 FEEQDAHELFWMTSSLEDERDQPRVTHLFDVHSLQEQSEITPKQITRTRGSPHPTSNHWKSHQPHFGRLTSNMV
 CKHCEHQSPVRFDTFDSLSPAPATWGHPLTLDHCLHFEISSESRDVVCDNCTKIEAKGTLNGEKVEHQRTTFVKQ
 LKGLPQCLCHLQRLSWSSHGTPKREHEVQFNEELMDYKYHLLGHKPSQHNPKNKNGPPTLELDQDGPAPT
 PVLNQPGAPKTQIFMNGACSPSLLPTLSAPMPFPLPVWPDYSSSTYLFRLMAVWHHGMHSHGHFVTVYRRSPPSARN
 PLSTSNQWLWVSDTVRKASLQEVLSSSAYLLFYERVLRMQHQSQECKSEE

>SGP427_SEQID_77
 MDLGFQDAAGGGFLAPRRRRRSLRRLFSRFLALGSRSRPGDSPRPQPGHCDGDEGGEFACAPGVPAPQSP
 GEERPPGPQQLPAGDGPAPGAQGLKNHGNTCFMNAVVOCLNNTDLAEFLAGRYRAAPGRAEYEQLAALV
 RALWTRREYTPQLSAEFKNAVSKYGSQFQNSQHDALFLLWLLDRVHLEGSSRGFVSEKLPPEATKTSNCLSPS
 AOLPLGQSFVQSHFOAQYRSSLTCPHCLKQSNFTDFLCVSLPIPLRQTRFLSVTLVFPKSKQRFRLVGLAVPILSTVA
 ALRKMVAEEGGVPADEVILVELYPSGFQSFDEEDLNTIAEGDNVYAFQVPPSPSQGTLSAHPLGLSAPRLAAREG
 QRFSLSHSESKVLLFCNLVSGGQQASRFQPPFLIREDRAVSWAQLQQSILSKYRHLMKSEAPVQNLGSLFSIRVVGL
 SVACSYLSPKDSRPLCHWAVDRVHLRRPGPPHVKLAVEWDSVSKERLFGSLQEEERAQDADSVWQQQAHQOH
 SCTLDECQFYTYKEEQLAQDDAWKCPQVQQGMVKLSLWTLPIILHILKRFQVGERNNKSLTVKFPGLSLNMA
 PHVAQRSTSPAGLGPWSPWKQPDCLPTSYPLDFLYDLVAVCNHGNLQGGHTAYCRNSLDGQWYSYDDSTVEP
 LREDEVNTRGAYILFYOKRNSIPWSSASSMRGSTSSLSHDHMLRLGSHAGSTRGSLSWSSAPCPSLPCQVDPSPIF
 TNSLCNKEKGGLPRRLVRGVKGRSISMKAPTTTTRAKQGPFTKIMPLRWSEGSKEKPPGASVELVEYLESRRRPRST
 SQSIVSLTGTAGEDEKASPRSNVALPANSEDDGGRAGIERGAPAGVCPASQPNHCLAPGNSDGPNTARKLKENAGQD
 IKLPRKFDLPLTMPSVEHEKPARPEGQKAMNWKESFQMGKSSPSPYMGFGSGNSKDSRRGTSELDRLQGLTTL
 LRSVFRKKENRRNERAEVSPQVPPVSLVSGGLSPAMIDGQAPGSPPALRIPEGLARGLGSRRLERDVWSAPSSLRLPR
 KASRAPRGSA LGMISQRTVPGEQASYGTFRQVKYHTLSLGRKKTLPESSE

WO 02/00860

58/68

PCT/US01/20171

FIG. 21

>SGPr082_SEQID_78
MQLVILRVITFLPWCFVPPPAADHKGWDFVEGYFHQFELTKKESPLLQTQTQLLQQFHRNGTDLDDMQMHALL
HQPCHGVDPGSDTSISPGCKWNKHTLTYRINYPHDMKPSAVKDSYNVSWNTPLFQCVQNGDADIKVSRWQ
WAHEDGWPFDPGFGILGHAFLPNSGNPGVWHFDKNEHWSASDTGYNLELVATHEIGHSLGLQHSGNQSIMYPTIYW
YHDPRTFQLSADDIQRIQLHYGEKCSDDIP

>SGPr359_SEQID_79
MKVLPASGLAVFLIMALKFSTAAPSLVAASPRTWNNRYLAQAYLDKYYTNKEGHOIGEMVARGSNMIRKIKELQAF
FGLQVTGKLDQTTMMVKKPRCGVPDVANYRLFPGEPKWKKNTLTYRISKYTPSMSSVEVDKAVEMALQAVSSAVPL
SFVRINSGEADIMISFENGHDGSDPGDPRGTLAHAFAPGEGGLGDDHFNPEKWTMGNTGNFLFTVAHFEFGHAL
GLAHSPTPSALMYPTTYKYNPYGFHLPKDDVKQIALYGPVKFLGKPTLPHAPHKPSIDLCDSSSSFDVATMLGK
ELLFKDRIFWRQVHLRTGIRPSTTSSFPQLMSNVDAAYEVAERGTYAFKGPQHYWTRGFMQGPRTIYDFGFP
PHVQQIDAAYVIREPQKTLFRVGDYYSYDERKRMKEDYPKNTTEEEFSGVNGQIDAAVELNGYIYFESGPKTYKYDT
EKEDVSVVKSSSWIGC

>SGPr104_1_SEQID_80
MNVALQELGAGSNMVEYKRATLRDEDAPETPVEGGASPDAMEVGFQKGTROLLGSRTOLELVLAGASLLAALLGC
LVALGVQYHRDPSHSTCLTEACIRVAGKILESLDRGVSPCEDFYQFSCGGWRRNPLPDGRSRWNTFNSLWDQNGAI
LKHLLENITFNSSEAEQKTORFYLSQLVERIEELGAGPLRDLEIKGGWNITGPWDQDNFMIEVLKAVAGTYRATPFF
TVYISADSKSSNNVQVDSGLFLPSRDYLRNRTANEKVLTAYLIDYMEELGMLGGRTSTREQMQQVLELEQLANI
TVPDQDRDEEKIYHKMSISELQALAPSDMWLEFLSPLSELDSEPVVYGYMDYLQQVSELINRTEPSILNNYLIW
NLVOKTSSLDREESAEQKLETLYGTKKSCVPRWQTCISNTDDALGFALGSLFVKATFDRQSKIEAGMISEIRTA
EEALGQLVWMDKTRQAAKEKADAIYDMIGFFDLEPKELDDVYDGYEISEDSFFQNMILNLVNFSAKIMADQLRKPP
SRDQWSMTPTQTVNAYYLLPTKNEIVFPAGILOAPFYARNHPKALNFGGIVWVGHLELTHAFDDQDQGREYDKEGNLRPW
WQNESLAFRNHTACMEEQYNQYQVNGERLNGRQTLGENADNGGLKAAYNAYKAWLRKHGEEQQLPAYGLTNHQ
LFFVGFAGVWCVSRTPESSHEGLVTDPHSPARFVRLGTLNSNRDLRHFHFGFVGPSPMNPQQLCEVV

>SGPr303_SEQID_81
MPEKRPFERLPADYSPINCSSLCKPLDLLDFTFEGKLEAAQVRQATNQVMNCADIDITIASYAPEGDEEHATGFNYQN
EDEKVTLSFPTLTGTGTGLKIDFVGLNLMKMGFYRSKYTTTPSGEVRYAAVTQFEATDARRAFCQWDERAIKATFDS
LVVPKDRVALSNMNVDRKPYPDENLVEVKFARTPVSTYLVAFVGEYDFVETRSKDGVCVCTVTPVGAEQGKF
ALEVAAKTLPPYNDYFNVPYPLPKIDLIAADFAAGAMENWDLVTYRETALLDPKNSCSSSRQWVALVGHLELAHQWF

WO 02/00860

59/68

PCT/US01/20171

FIG. 2J

GNLVTMEWTHLWLNIEGFASWIEYLCVDHCFPEYDIWTFVSADYTRAQELDALDNHPLEIVSVGHPSEVDEIFDAIS
YSGGASVIRMLHDYIGDKVKKKTLIS

>SGP402_1_SEQID_82

MRPAPIALWLRLLALALVRPRAV/GWAPVRAPYY/SSWAVQV/SQGNREVERLARKEFGFVNLGPFFDQGYFHLRHRG
WVQSLTPHWGHHLLKNPKYQWFWQQTQRRYKRSVWVPTDPWFQWQWMMSEAPDLSILOAWSQGLSGQGI
VSVLDDGIEKDHPLWANYDPLASYDFNDYDPDPQPRYTPSKENRHGTRCAGEVAAMANNFGCGVGVAFNARIGG
VRMLDGTITDVEAQSLSQPOHHYASWGPEDDGRITVDGPGILTREAFFRRGVTKRGGGLGLFIWASGNGGLHYD
NCCDGYTNSHTLSVGSSTTQQGRVPWYSEACASTLTITSSSGVATDQIVTDLHHGCTDQHTGTSASAPLAAGMI
ALALEANPFLTWDMOHLVWRASKPAHLOAEDWRITNGVGRQVSHHYGYLLDAGLLVDARTWLPTQPRKCAVR
VQSRTPIPLIYIRENVASACAGLHNSIRSLHQAQLTLYSRRGDLEISLTSPMGTRSTLVAIRPLDVTGEGYNNWVF
MSTHEWDENPQGVWTLGLENKGYFNTGLYRYTLLYGTAEQMTARPTGPQVTSSACVQORDTEGLCOACDGPAYI
LGQLCLAYCPRPFNHTRLVTAGPGHTAAPALRYCSCSHASCYTCRGGSPRCDTSCPPSSLTLDQQQSGSCMGPTTPD
SRPRLRAACPHHRCFASAMVLSLLAVTLGGPVLGMSMDLPLYAWLSRARATPTKQVWLPAGT

>SGP434_SEQID_83

GFGRQGGCAGRSTALPLRAPLRARRPGRSERMGAATCRGSRIPSGPPVOGERSAPREFGYTSLSLWPADFKDNW
RIAGSRQEVALAGEPADQQQTHLRPLPYRQTLGYKEDTTNPVCGEPWWSEDEMTNRHWPWEVSLRMENEHVCGS
ALIDPSWVVTAAHCSQGTKEYSVVLGTSKLPMMNFSRALWVPVDMIMHPKYWGRFIMGDVALVHLQTPVTFSEYV
QPICLPEPNFNLKVGTCQWWTGWSQVKQRFSGSTANSMLTPELQEAEEVIMDNKRCDRHYKKSFFPPVPLVLDGMI
CATNYGENLCYGDGGGFLACEVEGRWILAGVLSWEKACVKAQNPVYTRITKYTKWIKKQMSNGAFSGPCASACLLF
LCWPLQPQMGS

>SGP446_1_SEQID_84

ILTPVCGRTPLRIVGGVDAEEGRWPWQVSVRTKGRHICGGTLVTATWLTAGHCISRRFHYVKMGDRSVVYNNETSV
WVSQRAFVHPKFSTVTIRNDLALLQLHPVNFNTSNIQICIQENFQVEGRTRCWWTGWGKTPERGEKLAISEILODV
DOYIMCYEECNKIQKALSTKDVIIKGMVCGYKEQKDSGQDSSGRLACEYNDTWVQVGVSWGIGCGR

>SGP447_SEQID_85

MGARGALLALLARAGLKGPEACGHREIHALVAGGVESARGRWPWQASLRLRRRHRCGGSLLSRRRWLSAAHCF
QKHYYPSEWTVQLGELTSRPTPWNLRAYSSRYKVQDIIVNPDALGVLRNDIALLRASSVTYNAYQPICIESTFNFVH

FIG. 2K

RPDCWVTGWLISPGTGLPPPYNLREAOVTILNTRCNLYFEQPSRSMIWDMSMFCAGAEDGSDVTDCKGDSGGPL
VCDKDGLWYQVGIVSWGMDCCQPNRPVYTNISVYFHWIRVMSHSTPRNPSQLLLLLLWAP

>SGP432_1_SEQID_86
MGSTWGSFGWVRLALCLTGLVLSLYALHVKAAARDRDYRICALDVGTAISGRVFSRRWGRGFLVHVLQDQSL
NOSNIFGCIYFTLQLLLGLQAAQACGORGPGPKQEGNTVPGEPWQOASVRRQGAHICSGSLVADTWLTAAH
CFEKAATELNSWSVYLGSLQREGLSGAEVGVAAQLPRAYNHYSQSDALLQLAHTTHTPLCLPQPAHREPF
GASCWATGWDODTSDAPGTILNRLRUSRPTCNCTYNQLHQRHLNPARPGMLCGGPQPGVQCGQSGGGPVL
CLEPDGHWVQAIIISFASSCAQEDAPVLLTNTAAHSSWLOARVQGAAPFAQSPETPEMSDEDCVACGSLRTAGPQ
AGAPSPWPEARLMHQGLACGGALVSEAVLTAACHFGRQAPFEEVWSGLGTRPEEWGLKQLILHGAYTHPEGG
YDMAILLAOPVTLGASLRPLCLPYPDHLPDGERGWVIGRARGAGISSLQTVPTLLGPRACSRLLHAAPGGDGSPI
LPGMVCTSAVGELPSCGSLGAPLVHEVRGTWFLAGLSHFGDACQGPAPVFTALPAYEDWVSSLDWQVYFAEEP
EPEAEPGSLANISQPTSC

>SGP429_SEQID_87
MRAPHILSAAAGARALAKLLPILMAQLWAAEAALLPQNDTRLDPEAYGAPCARGSQPWQVSLFNGLSFHCAGVLV
DQSWLTAACHCGNKPLWARYGDDHLLLLQGEQLRRTTRSVVHPKYHQSGGSPILPRTDEHDLMLLKLRPVYPGPR
VRALQLPYRCAGPGDCCQVAGWGTTAARRVKYNGLTCSSITLSPKECEVFYPGVVTNNMCAGLDRGQDPCQSD
SGGPLVCDDETLLQILSWGVPYPCGSAQHPAVYTQICKYMSWINKVIRSN

>SGP428_1_SEQID_88
NVQCGHRPAPFPNSSWLPFHERLQVQNGECPWQVSIQMSRKHLCGSILHWWVLTAAHCFERRTLTLDMAVNVTVV
MGTRTFSNIHSEKQVQVVIHKYKPPQLDSDLSSLLLATPYQFQSNFKMPVCLQEEERTWDDWCWMAQWVTNGYD
QYDDLNMHLEKLRVYQISKECAKRVNQLSRNMICAWNPEGTNGQGPGEVGGPLVCQKKNKSTWYQLGIISWGYGC
GQKNMPGVYTELSNYLLWIERKTVLAGKPYKPEPDSVYALLSPWAILLYFVMLLS

>SGP425_SEQID_89
MENMLLWIFFTPGWTLLIDGSEMEWDFMWHLRKVRIVSERTFHLTSPAFAEADAKMMVNTVCGIECOKELPTPSLSEL
EDYLSYETVFENGTRTLTRYKVQDLVLEPTQNTITTKGVSVRRKQYVGTDSRFSILDKRFLTNPFTSTAVKLSGSGIL
ISPOHVLTAACHVHDGKYKGSKKLRVGLLKMRNKGKRRGSKRRSREASGGDQREGTRELPERAKGGRRR
KKSGRGQRAEGRPSFQWTRVKNTHIPKGWARGMGMDATLDYDVALLELKRAHKKYMELGISPTIKKMPGGMHIFS

WO 02/00860

61/68

PCT/US01/20171

FIG. 2L

GFDNDRADQLVYRFCVSDESNDLLYQYCDAESGSTGSGVYLRDKDPDKNWKRIIAVYSGHQWVDVHGVQKDYN
VAVRITPLKYAQICLWIHNDANCAYG

>SGP#48_SEQID_90

MGDPGSAEWGKGKIPVRRNLLTVDSISLCLEGSWWRRQKGPSPGFSHSLRPLQPNPSPSTMWLLTSLFLA
STAAODGDKLLEGDECAHISQPWQVALYERGRFNCASLISPHWLSAAHCQSRFMRVRLGEHNLKRDRGPEQLR
TTSRV/PHPRYEARSHRNDIMILLRLVQPARLNPOVPRPAVLPTRCPPHGEACVYSGWGLVSHNEPCTAGSPRSQVSLP
DTLHCANISISDTSCKSYPPGRLTNTMVCAGAEGRGAESCEGSGGFLVCGGILGGIVSWGVDVPCDNTTKPGVYTKV
CHYLEWIRETMKRN

>SGP#396_SEQID_91

MGPAGCAFTLLLLLSISVCGQPVYSSRWGGQDAAAAGRPWPQVSLHFDHNFICGGSLVSERLTLTAHCCIQTWTTFS
YTVWLGSITVGDSSRRKRVYYYSKIVHPKYQDTTADVALLKSSQVTFTSAILPCLPVSYTKQLAIPFCWWTGWKKE
SSDRDYHSAQEAEPIDRQACEQLYNPIGIFLPALEPVKEDKICAGDTQNMKDSGGLSCHIDGVMQITGV
VSWGLECGKSLPGVYTNVIYQKWINATISRANNLDFSDFLPVILLSLALLRPSCAFGPNTHIRVGTVAEAVACIQGWE
ENAWRFSPRGR

>SGP#426_SEQID_92

MMYAPVEFSEAEFSRAEYQKQFQWDSVRLALFTLAIAIIGIAGIVTHFVEDDKSFYYLASFKVTNIKYKENYGISS
REFIERSHQIERMMSRIFRHSVGGRFKSHVKLSPDEQGVILVILFRYPSTDSAEQIKKIEKALYQSLTKQLSLTIN
KPSERLTRCGIRMTSSNMPLPASSSTORIVQGRETAMEGEWPWQASLQLJSGHOCGASLISNTWLLTAHCFWKNK
DPTQWIATGATITTPAVKRNVRKILHENYHRETENDIALVQLSTGVEFSNIVQRVCLPDSSIKLPPKTSVYVTFGFSI
VDDGPIQNTLRQARVETISDVCNRKDVYDGLITPGMLCAGFMEGKIDACKGDSGGPLVYDNHDIWYVYVGVSWGQSC
ALPKKPGVYTRVTKYRDWIAASKTGM

>SGP#552_SEQID_93

RIAEGLDAEEGEWPWQASLPQNNVYRRGATWLSNSWLITAAHCIFRVHDPKEWNVLNPNQTSQSNKKNVIQENHYHP
AHDNDIAV/VHLSPLVYTSNIQACLPDNYIFLYNSEAVVTAWGSFKPLRTTSNVLHKGLVKIIDNRTCNNGEADGRVI
TSGMLCAGFLEPRVDACQDSDGGFLVGTDSKILAKGSLVLKAGVNERALPNKPSVYTVQVTTY

WO 02/00860

62/68

PCT/US01/20171

FIG. 2M

>SGP405_SEQID_94
 MYSKGGVAAEPPEPHYCEDSERGPNTLTGPGSLPRGGGIEVGMFEPPGSGEGCVKPHHEAAREGAGRGKRAVPGPK
 RRCQSGAEGPAAAGVTLEQETRGDVLKEDKNERADEEILRLAPGKGLPIDSXHLKXPVSSFPVRSQELGEGAGAGTLR
 GKMAEFNWSMAFKGPAAGHEERLNSVSSRAKKGIGWDVAAASLRGVDHFDLPPPLGVRELEACAFRVQVGGQLRL
 YEDDORTKWEIVRHPQYNESLSAQGGADIALKLEAPVPLSELHPVSLPSASLDVPSGKTQWVTGWGVGRGELL
 WPLSLWEATYKVRSNVLCNQTCRRRFPNSHTERFERLIKDDMLCAGDGNHGSWFGDNGSGPLLCRNCTWVQVVEV
 SWGKLCGLRGYPGMYTRVTSVSWIRQPCPSAQTPAVRRFVLPNPDIWEALTPSVMGSGAPLPAPDLOEAEVPI
 MRTRACERMYHKGPTAHGQVTHIAKAMPAGRGQGGSCQAALRTEDLTPTNTEVSPRADPRLSQPEDIWPEWAW
 PVWGTMLLLLLFLAVSSLGSCSTGSPAPVPENDLVGIVGHTNPGEVAVAGADRRSLHFPEGHRPVHLPSHQG
 CUSVRGPAAECQPDPRPNYSVFLGADIALKATSSLEFTDNCVMTGWGMVGLLMLPPYRPOQVKVLT
 SNADDERQTYDAPFGAGDRKFQDDMICAGRTGRRTWKGDGSGGPLVCKKGTWLQAGVVSWSGFSYDRPSIGVYTR
 PETSWOGANHADAGRPAVRPTMQPRDMGQGEWCRPFTHTVTCYPTAIRPFTHTVTCYLMVAVPSTLTHVTCYP
 TAVPRPFTHTVTCYLMVAVPSTLTHVTCYMMVAVPRPFTHTVTCYPMVAVPSTLTHVTCYPTAIRPFTHTVTCYMAIPRPSTTP
 PATRRPSPAPSPTSPATRWPSPGSPMSPATR

>SGP485_1_SEQID_95
 MLIFSLLLLSLVTRTQLGPRTPLEAGVAILGRARGAHRPQPPHPPSPVSECGDRSIFEGRTRYSRITGGMEAEVGE
 FPWQVSIQVRSEPFQGGSIINWILTAACHLYSEELFPEELSVLTGTNDLTSPSMEIKEVASIILHKDFKRANMDNDIA
 LLLASPIKLDLKVPICLPTOPGPATWRECVAGWGQTNAAADKNVKTDLMKAPMVMDEECCKMFPLTKNMLC
 AGYKNESYDACKGDSGGLVCTPEPGEKWKYQVGHISWGKSCGEKNTPGIYTSLVNYNLWIEKVTQLEGRPFNAEKRR
 TSVKQKPMGSPVSGVPEPSPRSWLLLCPLSHVLFRAILY

>SGP634_SEQID_96
 MASLWLLSCFSLVGAAGFCGVPAHPVLSGLSRVNGEDAVPGSWPQVSLQDKTGFHFHCGGSLISEDWVWVTAHC
 GVRTSDVYVAGEEDQGSDEENIQVLKIAKVFKNPKFSILTYNNDITLLKLATPARFSQTVSAVCLPSADDDFPAGTLCA
 TGWGTKYNNANKTPDKLQQAALPLLSNAECKKSWGRRITDVMICAGSGVSSCMGDSGGPLVCKQKGAWTLVGIVS
 WGSRTCSITTPAVYARVTKLIPWVQKILAA

>SGP630_SEQID_97
 MEPTVADVHLVPRTTKEVPALDAACCRASIGVATSLVLTGLVLLGGMNNSRHAALRAATLPGKYVSVTPPEASKTT
 NPPEGRNSEHRTSARTNSGHTIFKCKNTQPFSTQGFHVDHTAELRGIRWTSLSRRETSYHRTLTPTEALLHLLR

WO 02/00860

63/68

PCT/US01/20171

FIG. 2N

PLQTLISLIEELLQGRIRLREHGISLAAYGTIVSAELTGRHKPLAERDFKSGRCPSGNSQCQVTKVNPEC
DDQEDCSGSDSEAHCECGLQPAWRMAGRIVGMEASPGEPWQASLRNKEHFCGAAMINARWLVSAAHCFNEFQ
DPTKWVAYGATYLSGSEASTYRAQVQIMKHPLYNADTADFVALELTSPLFRGHIQPVCLPAATHIFFPSKKCLIS
GWGYLKEDFRKHLPRPAMKPEVLQKATVLDLQALCASLYGHSLTDRMVCAGYLDGKVDSCDGGGLVCEEPS
GRFLAGIVSWGIGCAEARPGVYARVTRLRDWILEATTAKSMPLATPMAPAPASTAWPTSPESVSTPTKSMQ
ALSTVPLDMWTVPKLQECGARPEMEKPTRVVGFGAASGEVPWQVSLKEGSRHFCGATVAGDRWLLSAAHCFNHT
KVEQVRAHLGTASLLGGSPVKIGLRRVVLHPLYNPGLDFDLAVLELAPLAFNKYIQPVCLPLAIQKFPVGRKCMIS
GWGNTQEGNATKPELLOKASVGIIDQKTCVLYNFSLTDRMICAGHLEGKQDQDGGSGPLACEEAPGVFYLQVVS
WGIGCAQVKKPGVYTRIRLKGWILEIMSSQPLMSPSTTRMLATTSPRTTAGLTVPGATPSRPTPGAASRVTGQPA
NSTLSAVSTTARGQTPFPDAPEATTHTQLPDCGLAPALTRIVGGSAAGGEWQVSLWLRREHRCGAVLVAER
WILLSAAHCFDNYGDPKQWAAFLGTFPLSGAEGOLERVARIYKHPFYNYLTDYDVALLELAGPVRRSRLVRPCLPEP
APRPDGTGTRCVITGWGSVREGGSMARQLQKAARLLSEQTCRRFYPVQISSRMILCAGFPQGGVDSCSDAGGGLA
CREPSGRWMLTGVTISWGYGCGRPHFPVYTRVAAVRGWIGQHIQE

>SGP#521_SEQID_98
MARSLLPLQILLISLALETAGEEAQGDKIIDGAPCARGSHPWQVALLSGNQLHCGGVLVNERWVLTAAHCKMNEYTV
HLGSDTLGDRRAQRKASKSFRHPGYSTQTHVNDLMLVKLNSQARLSNMVKVRLPSRCEPPGTCTVSGWGTITS
PDVTFPSDLMCVDVKLISPQDCTKVKYKDILLENMILCAGIPDSKKNACNGDGGGLVCGRTLQGLVSWGTFPCGGQPN
DPGVYTVQVCKFTKWINDTMKKHR

>SGP#530_1_SEQID_99
VSTVCGKPKVVGKTYGGRDAAAGQWPWQASLLYWGSHLGCAGVLDSCWLVSSTHCFLNKSKQAPKNYQVLLGNQLY
HQTQHTQKMSVHRITHPDFEKLHPFGSDIAMLQLHLMNIFTSYVVPVCLPSRDMQLPSNVSCWITGWMLTEDHKRV
QLSPFPYLQEGKVGLIENTLNTLYGORTAKARPKLCTRRCCVGGYFSTGKSICKGSDGGPLVCYLPASAWLVGLAS
WGLD CRHPAYPSIFTRVITYFINWIDEIMRLTPLSDPALAPH

>SGP#520_SEQID_100
MLLAVLLLLPLPSSWFAHGHPLYTRLPPSALQVFTLLGAETVLGRNLDYVCEGPCGERRRPTANVTRAHGRVGGSA
APPGAWPWLRLQLGGQLGGVLVAASWMLTAAHCFVGCSTRSAPNELLLWTVTLAEGSRGEQAEVVPVNRILPH
PKFDPRTFNDLALVQLWTPVSPGGSARPVCLPQEPPEPAGTACAGWAGALFEDGPEAEVREARVPLSTDTCR
RALGFLRPSTMLCAGYLAGVDSCDGGGLTCSEPPRPREVLFQVTSWGDGCGEPKPGVYTRVAVFKDWL
QEQMSAASSSSREPSRELLAWDPPQELQADAARLCAFYARLCPGSGQACARLAHQOQLQRRRRCLESLAHTLLGL

[illegible]

FIG. 2P

[illegible]

WO 02/00860

66/68

PCT/US01/20171

FIG. 2Q

>SGP#24_1_SEQID_108
MDKENSADVSAAPADLKSINISVQVWSAQKKLPVRRPPLPGRRLPLPGRPPQRRIGKAKPKQSKKKKVPFWVWQNKIIL
FTVFLFIADIAWTLVLWYISKTESDAFYFAGMERITNIEFLPEYRQKESREFLSVSRITVQOVNLVYTTSAFSKFEQS
VWADVSSNNKGLLVHFMVMPRAKHFCEDCVAILKDSIQTSINRTSVGLQGLAVMDMSVGLNGDCWFLK
KKRKENGAVSTDGCSQYFAEHLISLHYPLEISAAGRLMCHFKLVAIVGYLURLSIKIOIEADNCVTDLSLTYDLSLPIR
SSILYRICEPTRTLMSFVSTNNMLLVTFKSPHRRLSGIRAYFEVPEQKCENTVLVKDITGFEKGKSSPYPSYPKCK
CTWKFTSLSTLGLALKFYNSITKSNMKGCEHGWINEHMYCGSYMDHQITFRVPSPLVHQLQCSRLSDKPLLA
EYGSYNISQPCVGSFCSSGLCPQAQCDGVNDCDEDELFCYSPQACNTSSFRHGPLICDGFRCDCENGRD
EONCTQSPICNNRTFKCGNDICFRQNAKCDGTVDPCDGSDEEGCTCSRSSALHRIIGTDTLLEGGWPMQVSLHF
VGSAYCGASVISREWLLSAAHCFHGNRLSDPTPTAHLGMYVQGNKAFVSPVRRIWHEYVNSQTFDYDIALQLSIA
WPETLKLQICPIPTGGRVRSGEKCVWTGWGRRHEADNKGSVLQQALEVELDQTLVSTYGIITSRMLCAGIMSG
KRDACKGDSGGPLSCRRKSDGKWLTGIVSWGHCGRNFPFVYTRVSNFVPMWHKYVPSSL

>SGP#22_SEQID_109
MTLNKIKDLFAGKGQWDLAPEAEMLKPMWIAVLVLSLTVAVTIGLLVHFLVFDQKKEYTHGSFKLLDPOINNFGOSN
TYQLKDLRETTENLYSLKMYLSFVCHSPDEEGVKKVDVIMVFCQFPSTEQRAVREKKIQSILNQKIRNLRALPINASSVQV
NVAMVKNVGPVGSAGEAGPLGAGPAWSPMSSTGELTVQASCGRKRVPLNVRNIAAGVIAKAAWPAQASLQY
DNHQCAGATLISNTWLVTAAHCFQKYNPHQWTVSFGTKINPPLMKRNVRFIHEKYRSAAREYDIAVVQVSSRVTS
DDIRQICLPEASAFQPLNLTVHITGFGALYYGGESQNDLREARVKIISDDVCKQPOVYVYNDIKPGMFCAGYMEGIYDAC
RGDSGGPLVTRDLKDTWYLGIVSWGDNCGQKDKPGVYVYQVYTYRNWIAKSTGI

>SGP#38_SEQID_110
MSLMLDDQPPMEAQYAEEGPGPIFFRAEPDQDQHPISQAVCWRSWRRGCAVLGALLAGAGVGSWLLVLYLCPA
ASQPISTGLQDEEITLSCSEASAEALLPALPKTVSFRINSEDFLLAQVRDQPRWLLVCHEGWSPALGLQICWLSGHL
RLTHHKGVNLTDIKLNSQDEFAQLSPRLGGFLEEAWQPRNNCTSGQVSLRCSECGARPLASGOSVAPGRWP
WQASVALGRFRHTCGGSLAPRWVVTAAHCHMSFRLARLSSWRVHAGLSVSHSAVRPHQAGALVERIHPHLYSAQNHD
YDVALRLQALNFDITVGA/CLPAKEQHPKSRGWSGWGHTPHSYSSDMLODVTVPFLFTQLCNSSCYSYG
ALTFRMLCAGYLDGRADACQDGGPLVCPDGTWRLVGVVSWGRACAEFPNHPGVAKVAEFLDWHIDTAQDSLL

WO 02/00860

67/68

PCT/US01/20171

FIG. 2R

>SGP#527_1_SEQID_111
MARHLILPLVIMLVISPIGAFQDSALSPTQEEPEDLDCGRPEPSARIVGGSNAQPGTWPVQVSLHHGGHICGGSLLA
PSWVLSAACHFMNTGLEPAEWVSLLGVHSDGDLGDGHTRAVAIWWPANYSQVELGADLALLRASPASLGPAY
WPVCLPRASHRFVHGTAWMATGWGDVQEADPLPWLQEVLELRLLEATCCCLYQSPGPFNLTLQILPGMLCAGY
PEGRRDTCQDGGGGLVCEEGRWFQAGITSFGCGRRNRPGVFTAVTYEAWIREQWVGSEPGFAFTQPKT
QSDPQPREENCTIALPECCAKPRPGAWPWEAQWMPGSRPCHGALVSESWVLAPASCFLDPNSSDSPRDLDAW
RVLLPSRPRAEVRLVQHENASWDNASDLALLQLRTPVNLASAARVCLPHEHYLPGRSCLRLRWGRGEPALG
PGALLEALLGGWCHCLYGRQGAAPLPDGPPhALCPAYQEEKEEVGSCWTHGPWISHVTRGAYLEDDQLAWDWG
PDGEETETQTCPPHTEHGACGLRLEAAPVGLWPMLAEVHAGDVCTGILLAPGWVLAATHCVLRPGSTTVPYIEV
YLRAGASSLPQGHQVSRVLSIRLPQHLLGRPLALLELSRRVEPSPSALPICLHPAGIPPGASCWILGWKEPDQDRVP
VAAAVSILTQRICDCLYQGLPPGTCLVLYAEGGECNRCMTSAPPLLCQMTGSGWILVGMVQSGSRELFAAIGPEEAWI
SQTVEANFLPPSGSPHWPPTGGSNLCFPELAKASGSPHAYVYFLLLTLLIQS

>SGP#542_SEQID_112
AMGLGLRGWGRPLLTVA TALMLPVKPPAGSWGAIIGGHEVTPHSRPPYMASVRFGGQHGCGGELLRARWVYSAAH
CFSHRDLRTGLVVLGAHVLSAEPTQVFGIDALTTHPDYHMTHANDICLLQLNGSAVILGPAVGLRLPGRRARPPT
AGTRCRVAGWGFVSDFEELPPGLMEAKRVLPDVCNSSWKGHLTLTMLCTRSGDSHRRGFCSDSGGGLVORNR
AHGLVSFSGLWCGDPKTPDVYTVQVSFAVAMWMDVRRSSPQGPLPGTTRPPGEAA

>SGP#551_SEQID_113
MPVAEPQVAGGGGDDGDEEAEPEGMFKACEDSKRKARGYLRVPLFVALLVLSAGVLLWYFLGYKAEVWVS
QYVSGSLVNLNRHFSQDLTRRESSAFRSETAKAKMLKELITSRLGTYNSSSVYSF-GEGLTCFFWFIQIPEHRRRL
MLSPVQALLVEELLSTVNSSAAVYRAEYEDVDEGLVILEASVYKDIAINSLTLCYRYSYVYGGQVLRKGPDLHAS
SCLWLHQPKDMLKRLWTLAECRDLAMVDVAGFLEKRLITSVYGCSSRQEPVWEVLAGSAMA VVWKGLHSY
DPFVLSQPVVFOACEVNLTLNRLDSQGLVSTIFYPSYSPQTHCSWHLTVPSLDYGLALWEDAYALRROKYDLPC
TQGGWTQNRRLCGLRLQPYAERIPVATAGTINFTSQISLTGPGVGVHYGLYNQSDPCPEEELCSVNGLCVPACDG
VKDCPNGLDERNCVCRATFQCKEDSTCISLPKVDGDPCLNGSDEEQCEGVPCGTFTFQCEDRSCVKKPNQC
DGRPCDRGSDDEHCDGCLQGPSRIRVGGAVSSEGEWPMQASLQGRHICGGGALLADRWVITAAHCFQEDSMAS
TVLWTVFLGKVVQNSRWPGVSKVSRLLHPYHEEDSHDYDVALQLDHPVVRSAARVPCVLPARSHFFEPGLHC
WITGWALREGGPISNALOKVDVQLIPQDLCEAYRYQVTPRMLCAGYRKGKDACQCGSDSGGLVCKALSGRWFLA
GLVSWGLGCGRPNTYGVYTRITGVISWQQVWT

FIG. 2S

>SGP#451_SEQID_114
DILPSCDVSFASPMIRLGLLSVALLFVGSSHLVSDHYSPSGRHRLLGSPPEPAASSQQAEEAVRKRRLRRRRGGGAHAKDCG
TAPLAKDVLGSRIGTAEAGAAGWPWVSLQIKRYVHLVHVGCGTLVLRERWMLTAAHCTKDTSDPLMMVTAVGTNNHH
GVHPYHTTKKIKAILHNPFILESYNNDIALFLHKAVRYNDYIQPICLPFDVFOILDGNTKCFISGWGRTEEAGNATNILLQD
AEVHYSIREMCMENRSERYGGIIPNTSCFAGDEDAFDTCRGDSGGPLMCYLPEYKRFVFMGITSYGHGCGRRGFRGVY
IGSPSYQKWMLTEHFFHMASQGLTNILRGQIALLCFVILLATT

>SGP#452_1_SEQID_115
SPPOQPTPCRLOQASLEALATLAPQSPDWLFCADLGMFEADGAAHSMGLGSSLKWAWAKSPGMPVPENDLVGVG
GHVNPAPQPVWQVSLVSYHHWASWHAHCQSLHPQWLTAACHICFWKDTDSYIRHAGDVLYYGGRLGLNVSRIG
VHINHPYVTAGLDGAVALLQPSGLSPESPPYRLQASGVLENAVCEQPYRNASGHTGRQLILDMLCAGSEG
RDSCYGDGGPLVCLRLGRSWRLVGSWSVGCTLRDPFGVYTHVQIYVILWLOQQVGELP

>SGP#504_SEQID_116
IIGGHEVYTHSRPYMASVRFGGSHHCGGFLRLRWVWSAAQCFSH

>SGP#469_SEQID_117
GDSGGPLVCELNGTWQVGVISWGICGCRKGYPGVYTEVSFYKKWII

>SGP#400_SEQID_118
MAGEQVTVANVSRYPGQKTMSPFEKTFLLSYRASLLAVVTHRSNNRSGRAFESQVLPDLTAGDAADPPPIPLGFGAAL
LKSGSPFRWQGVKTKGEEGRDRTGTAGYAFTLLLLGISGEPEWVCGRPTVSSGSIAGLGA(SV/GQWPWQVQIRQGL
TQVQKIRSYLKLKVLVLTQTCGDGTHQNEILLHVEIJEAMICSKLPVGMQMDQCTVRIHPSGTTHRPLCPQ

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 January 2002 (03.01.2002)

PCT

(10) International Publication Number
WO 02/000860 A3

(51) International Patent Classification¹: C12N 15/55, 9/64, C07K 16/40, C12N 5/10, G01N 33/50, 33/53, C12Q 1/68	MANNING, Gerard [US/US]; 444 Fremont Street #4, Menlo Park, CA 94025 (US); CAENPEEL, Sean [US/US]; 447 Orange Street, #22, Oakland, CA 94610 (US); CHARVYDCZAK, Glen [US/US]; 936 Sir Francis Drake Boulevard, Kentfield, CA 94904 (US)
(21) International Application Number: PCT/US01/20171	
(22) International Filing Date: 26 June 2001 (26.06.2001)	(74) Agents: BURROUS, Beth, A. et al.; Foley & Lardner; Washington Harbour, Suite 500, 3000 K Street, N.W., Washington, DC 20007-5109 (US).
(23) Filing Language: English	
(26) Publication Language: English	(81) Designated States (national): AI, AG, AT, AM, AU, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GR, HK, IR, IT, IJ, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LI, LU, LS, LT, LV, MA, MD, MG, MN, MW, MX, MY, MZ, NZ, NI, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
(30) Priority Data: 60/214,047 26 June 2000 (26.06.2000) US	
(7) Applicant (for all designated States except US): SUGEN, INC. [US/US]; 230 East Grand Avenue, South San Francisco, CA 94080-4811 (US).	
(72) Inventors; and	(84) Designated States (regional): AR, AU, AT, BE, BR, BF, BG, CA, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IL, IN, JP, KE, KG, KP, KR, KZ, LC, LI, LU, LT, LV, MA, MD, MG, MN, MW, MX, MY, MZ, NZ, NI, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
(75) Inventors/Applicants (for US only): PLOWMAN, Gregory [US/US]; Winding Way, San Carlos, CA 94070 (US); WHYTE, David [US/US]; 2023 Barclay Way, Belmont, CA 94002 (US); SUDARSANAM, Sucha [US/US]; 20 Corte Palencia, Greenbrae, CA 94904 (US).	

[Continued on next page]

(54) **Title:** NOVEL PROTEASES[illegible][illegible]

(57) Abstract: The present invention relates to protease polypeptides, nucleotide sequences encoding the protease polypeptides, as well as various products and methods useful for the diagnosis and treatment of various protease-related diseases and conditions.

WO 02/000860 A3

WO 02/000860 A3



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.

(88) Date of publication of the international search report:
3 April 2003

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 01/20171															
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/55 C12N9/64 C07K16/40 C12N5/10 G01N33/50 G01N33/53 C12Q1/68																	
According to International Patent Classification (IPC) or to both national classification and IPC																	
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K G01N C12Q																	
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched																	
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EP0-Internal, EMBL, BIOSIS, SEQUENCE SEARCH, WPI Data, PAJ																	
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category *</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>DATABASE EMBL SEQUENCE LIBRARY [Online] 22 November 1995 (1995-11-22) HILLIER, L., ET AL.: "the WachU-Merck EST Project - yt08c10.r1 Soares retina N2b5HR Homo sapiens cDNA clone IMAGE:223698 5", mRNA sequence" XP002193829 accession no. H86718 ---</td> <td>1,3-8, 24-28</td> </tr> <tr> <td>A</td> <td>WO 98 35988 A (ZENECA LTD ;EDGE MICHAEL DEREK (GB)) 20 August 1998 (1998-08-20) the whole document ---</td> <td></td> </tr> <tr> <td>A</td> <td>EP 0 887 416 A (SMITHKLINE BEECHAM PLC ;SMITHKLINE BEECHAM CORP (US)) 30 December 1998 (1998-12-30) the whole document ---</td> <td></td> </tr> <tr> <td></td> <td>-/--</td> <td></td> </tr> </tbody> </table>			Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	DATABASE EMBL SEQUENCE LIBRARY [Online] 22 November 1995 (1995-11-22) HILLIER, L., ET AL.: "the WachU-Merck EST Project - yt08c10.r1 Soares retina N2b5HR Homo sapiens cDNA clone IMAGE:223698 5", mRNA sequence" XP002193829 accession no. H86718 ---	1,3-8, 24-28	A	WO 98 35988 A (ZENECA LTD ;EDGE MICHAEL DEREK (GB)) 20 August 1998 (1998-08-20) the whole document ---		A	EP 0 887 416 A (SMITHKLINE BEECHAM PLC ;SMITHKLINE BEECHAM CORP (US)) 30 December 1998 (1998-12-30) the whole document ---			-/--	
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.															
X	DATABASE EMBL SEQUENCE LIBRARY [Online] 22 November 1995 (1995-11-22) HILLIER, L., ET AL.: "the WachU-Merck EST Project - yt08c10.r1 Soares retina N2b5HR Homo sapiens cDNA clone IMAGE:223698 5", mRNA sequence" XP002193829 accession no. H86718 ---	1,3-8, 24-28															
A	WO 98 35988 A (ZENECA LTD ;EDGE MICHAEL DEREK (GB)) 20 August 1998 (1998-08-20) the whole document ---																
A	EP 0 887 416 A (SMITHKLINE BEECHAM PLC ;SMITHKLINE BEECHAM CORP (US)) 30 December 1998 (1998-12-30) the whole document ---																
	-/--																
☒ Further documents are listed in the continuation of box C. ☒ Patent family members are listed in annex.																	
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family																	
Date of the actual completion of the international search 21 March 2002		Date of mailing of the international search report 12.07.02															
Name and mailing address of the ISA European Patent Office, P.O. Box 5518 Patentstrasse 2 NL - 2280 HV Rijswijk Tel (+31-70) 340-2000, Tx 31 651 epo nl Fax (+31-70) 340-3016		Authorized officer Holtorf, S															

Form PCT/ISA210 (second sheet) (July 1999)

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 01/20171

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 245 051 A (UNIV LELAND STANFORD JUNIOR) 11 November 1987 (1987-11-11) the whole document ---	
A	WO 99 23106 A (HUMAN GENOME SCIENCES INC ;NI JIAN (US); RUBEN STEVEN M (US)) 14 May 1999 (1999-05-14) the whole document ---	
P,X	DATABASE EMBL SEQUENCE LIBRARY [Online] 2 August 2000 (2000-08-02) GU J., ET AL.: "Novel genes expressed in hematopoietic stem/progenitor cells from Myelodysplastic Syndromes patient - Homo sapiens carboxypeptidase B precursor (CPAH) mRNA, complete cds." XP002162176 accession no. AF221594 ---	1,3-8, 24-28
P,X	WO 01 23588 A (LEXICON GENETICS INC) 5 April 2001 (2001-04-05) the whole document -----	1-30

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

INTERNATIONAL SEARCH REPORT	International application No. PCT/US 01/20171
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.: 14-19 because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically: see FURTHER INFORMATION sheet PCT/ISA/210	
3. ☐ Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).	
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)	
This International Searching Authority found multiple inventions in this International application, as follows:	
see additional sheet	
1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.	
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.	
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:	
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-30 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 01/20171

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 14-19

Present claims 14-19 relate to a method of treatment by administering a substance; said substance being defined by reference to a desirable characteristic or property, namely the modulation of the activity of a protease as characterized by SEQID60.

The claims cover all substances having this characteristic or property, whereas the application provides no support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for such substances.

In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the substance by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, no search has been carried out for claims 14-19.

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

International Application No. PCT/US 01/20171

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

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

1. Claims: 1-30 partially

invention 1: claims 1-30 partially

Human nucleic acid molecule encoding a polypeptide as characterized by SEQID60, further comprising a vector or operatively linked promoter, a hybridizing nucleic acid molecule, a human protease polypeptide as characterized by SEQID60, an antibody binding to said polypeptide, a hybridoma producing and a kit comprising said antibody, a method for identifying a substance modulating the activity of said polypeptide by contacting a polypeptide as characterized by SEQID60 with a test substance, method for identifying a substance modulating the activity of said polypeptide by recombinantly expressing said polypeptide in a cell and adding a test substance to said cell; method for treating a disease by administering such a substance in vivo or in vitro, said method wherein said substance is a protease inhibitor, method of detection of a protease polypeptide in a sample by contacting said sample with a nucleic acid probe that hybridizes to a nucleic acid target region, an isolated nucleic acid molecule that encodes a biological domain of a protease or an amino acid sequence that is 90% identical to a protease as characterized by SEQID60, a nucleic acid molecule that is characterized by SEQID1, nucleic acid molecule consisting of about 10-30 nucleotide bases of a nucleic acid sequence of SEQID1 or of a nucleic acid sequence encoding a polypeptide of SEQID60, a recombinant cell comprising such nucleic acid sequence, a polypeptide of SEQID60, a method for detecting the presence or amount of a protease polypeptide in a sample by forming a protease-antibody immunocomplex.

invention 2-59: claims 1-30 partially

as invention 1 but limited to SEQIDs 61-118 and SEQIDs 2-59, respectively.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 01/20171

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9835988	A	20-08-1998	AU WO 9835988 A1	08-09-1998 20-08-1998
EP 0887416	A	30-12-1998	CA EP 0887416 A2 JP 11113579 A US 6060277 A	24-12-1998 30-12-1998 27-04-1999 09-05-2000
EP 0245051	A	11-11-1987	US DE 3781017 D1 DE 3781017 T2 EP 0245051 A1 JP 62285783 A US 5213977 A	27-11-1980 17-09-1992 18-03-1993 11-11-1987 11-12-1987 25-05-1993
WO 9923106	A	14-05-1999	AU 9673898 A CA 2308112 A1 EP 1036086 A1 JP 2001521730 T WO 9923106 A1	24-05-1999 14-05-1999 20-09-2000 13-11-2001 14-05-1999
WO 0123588	A	05-04-2001	AU 7989000 A EP 1220932 A1 WO 0123588 A1	30-04-2001 10-07-2002 05-04-2001

フロントページの続き

(51)Int.Cl.⁷

F I

テーマコード (参考)

4H045

A 6 1 P 3/10
 A 6 1 P 7/02
 A 6 1 P 9/00
 A 6 1 P 9/02
 A 6 1 P 9/10
 A 6 1 P 9/12
 A 6 1 P 11/02
 A 6 1 P 11/06
 A 6 1 P 15/10
 A 6 1 P 17/06
 A 6 1 P 19/00
 A 6 1 P 19/02
 A 6 1 P 21/00
 A 6 1 P 25/00
 A 6 1 P 25/02
 A 6 1 P 25/04
 A 6 1 P 25/06
 A 6 1 P 25/14
 A 6 1 P 25/16
 A 6 1 P 25/18
 A 6 1 P 25/22
 A 6 1 P 25/24
 A 6 1 P 25/28
 A 6 1 P 27/02
 A 6 1 P 27/06
 A 6 1 P 29/00
 A 6 1 P 31/04
 A 6 1 P 31/10
 A 6 1 P 31/12
 A 6 1 P 31/18
 A 6 1 P 35/00
 A 6 1 P 35/02
 A 6 1 P 37/00
 A 6 1 P 37/02
 A 6 1 P 43/00
 C 0 7 K 16/40
 C 1 2 N 1/15
 C 1 2 N 1/19
 C 1 2 N 1/21
 C 1 2 N 5/10
 C 1 2 N 9/64
 C 1 2 Q 1/02
 C 1 2 Q 1/37
 C 1 2 Q 1/68
 G 0 1 N 33/53
 G 0 1 N 33/566

A 6 1 P 3/10
 A 6 1 P 7/02
 A 6 1 P 9/00
 A 6 1 P 9/02
 A 6 1 P 9/10
 A 6 1 P 9/12
 A 6 1 P 11/02
 A 6 1 P 11/06
 A 6 1 P 15/10
 A 6 1 P 17/06
 A 6 1 P 19/00
 A 6 1 P 19/02
 A 6 1 P 21/00
 A 6 1 P 25/00
 A 6 1 P 25/02
 A 6 1 P 25/04
 A 6 1 P 25/06
 A 6 1 P 25/14
 A 6 1 P 25/16
 A 6 1 P 25/18
 A 6 1 P 25/22
 A 6 1 P 25/24
 A 6 1 P 25/28
 A 6 1 P 27/02
 A 6 1 P 27/06
 A 6 1 P 29/00
 A 6 1 P 29/00
 A 6 1 P 31/04
 A 6 1 P 31/10
 A 6 1 P 31/12
 A 6 1 P 31/18
 A 6 1 P 31/18
 A 6 1 P 35/00
 A 6 1 P 35/02
 A 6 1 P 37/00
 A 6 1 P 37/02
 A 6 1 P 37/02
 A 6 1 P 43/00
 A 6 1 P 43/00
 C 0 7 K 16/40
 C 1 2 N 1/15
 C 1 2 N 1/19
 C 1 2 N 1/21
 C 1 2 N 9/64
 C 1 2 Q 1/02
 C 1 2 Q 1/37
 C 1 2 Q 1/68
 G 0 1 N 33/53
 G 0 1 N 33/53

1 0 1

1 0 5

1 1 1

Z

A

D

M

G O 1 N 33/53 N
 G O 1 N 33/566
 C 1 2 N 5/00 B
 C 1 2 N 5/00 A

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

(74)代理人 100105991

弁理士 田中 玲子

(72)発明者 プラウマン, グレゴリー

アメリカ合衆国 94070 カリフォルニア州 サン カルロス, ワインディング ウェイ 35

(72)発明者 ホワイト, デイビッド

アメリカ合衆国 94002 カリフォルニア州 ベルモント, パークレイ ウェイ 2623

(72)発明者 シュダーサナム, スーチャ

アメリカ合衆国 94904 カリフォルニア州 グリーンプレー, コルテ パレンシオ 20

(72)発明者 マニング, ジェラルド

アメリカ合衆国 94025 カリフォルニア州 メンロ パーク, フレモント ストリート 844, #4

(72)発明者 ケーネピール, ショーン

アメリカ合衆国 94610 カリフォルニア州 オークランド, オレンジ ストリート 447, #22

(72)発明者 チャリィドクザック, グレン

アメリカ合衆国 94904 カリフォルニア州 ケントフィールド, サー フランシス ドレイク ブールバード 936

F ターム(参考) 4B024 AA01 AA11 BA14 CA04 CA06 DA02 EA04 GA11 HA01

4B050 CC03 DD11 LL01

4B063 QA18 QQ36 QR32 QR55 QS02 QS31 QS34

4B065 AA93X AA99Y AB01 AC14 BA02 CA33 CA44

4C084 AA17 NA14 ZA011 ZA021 ZA081 ZA121 ZA151 ZA161 ZA181 ZA201

ZA221 ZA331 ZA341 ZA361 ZA421 ZA431 ZA451 ZA541 ZA591 ZA681

ZA701 ZA811 ZA891 ZA941 ZA961 ZB021 ZB071 ZB111 ZB151 ZB211

ZB261 ZB271 ZB331 ZB351 ZC192 ZC202 ZC211 ZC351 ZC551

4H045 AA11 CA40 DA76 EA28 EA50 FA74

专利名称(译)	新型蛋白酶		
公开(公告)号	JP2004501637A	公开(公告)日	2004-01-22
申请号	JP2002506176	申请日	2001-06-26
[标]申请(专利权)人(译)	苏根公司		
申请(专利权)人(译)	Sujen公司		
[标]发明人	プラウマングレゴリー ホワイトデイビッド シュダーサナムスーチャ マニングジェラルド ケーネピールショーン チャリイドクザックグレン		
发明人	プラウマン,グレゴリー ホワイト,デイビッド シュダーサナム,スーチャ マニング,ジェラルド ケーネピール,ショーン チャリイドクザック,グレン		
IPC分类号	G01N33/53 A61K38/00 A61K45/00 A61P1/04 A61P3/00 A61P3/04 A61P3/10 A61P7/02 A61P9/00 A61P9/02 A61P9/10 A61P9/12 A61P11/02 A61P11/06 A61P15/10 A61P17/06 A61P19/00 A61P19/02 A61P21/00 A61P25/00 A61P25/02 A61P25/04 A61P25/06 A61P25/14 A61P25/16 A61P25/18 A61P25 /22 A61P25/24 A61P25/28 A61P27/02 A61P27/06 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P31/18 A61P35/00 A61P35/02 A61P37/00 A61P37/02 A61P43/00 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N9/64 C12N15/09 C12N15/55 C12Q1/02 C12Q1/37 C12Q1/68 G01N33/566		
CPC分类号	A61K38/00 A61P1/04 A61P3/00 A61P3/04 A61P11/02 A61P11/06 A61P15/10 A61P17/06 A61P19/00 A61P19/02 A61P21/00 A61P25/00 A61P25/02 A61P25/04 A61P25/06 A61P25/14 A61P25/16 A61P25 /18 A61P25/22 A61P25/24 A61P25/28 A61P27/02 A61P27/06 A61P29/00 A61P31/04 A61P31/10 A61P31/12 A61P31/18 A61P35/00 A61P35/02 C07K2319/00 C12N9/6421		
FI分类号	C12N15/00.ZNA.A A61K45/00 A61P1/04 A61P3/00 A61P3/04 A61P3/10 A61P7/02 A61P9/00 A61P9 /02 A61P9/10 A61P9/12 A61P11/02 A61P11/06 A61P15/10 A61P17/06 A61P19/00 A61P19/02 A61P21 /00 A61P25/00 A61P25/02 A61P25/04 A61P25/06 A61P25/14 A61P25/16 A61P25/18 A61P25/22 A61P25/24 A61P25/28 A61P27/02 A61P27/06 A61P29/00 A61P29/00.101 A61P31/04 A61P31/10 A61P31/12 A61P31/18 A61P35/00 A61P35/02 A61P37/00 A61P37/02 A61P43/00.105 A61P43/00.111 C07K16/40 C12N1/15 C12N1/19 C12N1/21 C12N9/64.Z C12Q1/02 C12Q1/37 C12Q1/68.A G01N33/53. D G01N33/53.M G01N33/53.N G01N33/566 C12N5/00.B C12N5/00.A		
F-TERM分类号	4B024/AA01 4B024/AA11 4B024/BA14 4B024/CA04 4B024/CA06 4B024/DA02 4B024/EA04 4B024 /GA11 4B024/HA01 4B050/CC03 4B050/DD11 4B050/LL01 4B063/QA18 4B063/QQ36 4B063/QR32 4B063/QR55 4B063/QS02 4B063/QS31 4B063/QS34 4B065/AA93X 4B065/AA99Y 4B065/AB01 4B065 /AC14 4B065/BA02 4B065/CA33 4B065/CA44 4C084/AA17 4C084/NA14 4C084/ZA011 4C084/ZA021 4C084/ZA081 4C084/ZA121 4C084/ZA151 4C084/ZA161 4C084/ZA181 4C084/ZA201 4C084/ZA221 4C084/ZA331 4C084/ZA341 4C084/ZA361 4C084/ZA421 4C084/ZA431 4C084/ZA451 4C084/ZA541 4C084/ZA591 4C084/ZA681 4C084/ZA701 4C084/ZA811 4C084/ZA891 4C084/ZA941 4C084/ZA961 4C084/ZB021 4C084/ZB071 4C084/ZB111 4C084/ZB151 4C084/ZB211 4C084/ZB261 4C084/ZB271 4C084/ZB331 4C084/ZB351 4C084/ZC192 4C084/ZC202 4C084/ZC211 4C084/ZC351 4C084/ZC551 4H045/AA11 4H045/CA40 4H045/DA76 4H045/EA28 4H045/EA50 4H045/FA74		
代理人(译)	森田浩二		

田中玲子

优先权60/214047 2000-06-26 US

外部链接[Espacenet](#)

摘要(译)

本发明涉及蛋白酶多肽，编码该蛋白酶多肽的核苷酸序列，以及用于诊断和治疗各种蛋白酶相关疾病和病症的各种产品和方法。