

(19) 日本国特許庁(JP)

## (12) 公表特許公報(A)

(11) 特許出願公表番号

特表2005-500010  
(P2005-500010A)

(43) 公表日 平成17年1月6日(2005.1.6)

(51) Int.C1.<sup>7</sup>

**C12N 15/09**  
**A01K 67/027**  
**A61K 31/7088**  
**A61K 38/00**  
**A61K 39/395**

F 1

C 12 N 15/00  
A O 1 K 67/027  
A 61 K 31/7088  
A 61 K 39/395  
A 61 K 39/395

Z N A A  
A 61 K 39/395  
A 61 K 39/395  
D  
N

テーマコード(参考)

2 G 04 5  
4 B 02 4  
4 B 02 9  
4 B 06 3  
4 B 06 5

審査請求 未請求 予備審査請求 未請求 (全 183 頁) 最終頁に続く

(21) 出願番号	特願2002-569876 (P2002-569876)	(71) 出願人	502334043 インファーマティカ リミテッド
(86) (22) 出願日	平成14年3月5日 (2002.3.5)		イギリス ロンドン ダブリュー1ティー 2エヌユー チャーロット ストリート 60
(85) 翻訳文提出日	平成15年9月5日 (2003.9.5)	(74) 代理人	100059959 弁理士 中村 稔
(86) 國際出願番号	PCT/GB2002/000937	(74) 代理人	100067013 弁理士 大塚 文昭
(87) 國際公開番号	W02002/070557	(74) 代理人	100082005 弁理士 熊倉 穎男
(87) 國際公開日	平成14年9月12日 (2002.9.12)	(74) 代理人	100065189 弁理士 宮戸 嘉一
(31) 優先権主張番号	0105402.2	(74) 代理人	100074228 弁理士 今城 俊夫
(32) 優先日	平成13年3月5日 (2001.3.5)		
(33) 優先権主張国	英國 (GB)		

最終頁に続く

(54) 【発明の名称】核内ホルモンレセプタリガンド結合ドメイン

## (57) 【要約】

本発明は、新規なタンパク質 (LBDG3と称され、本明細書では核内ホルモンレセプタリガンド結合ドメインと特定された) 並びに疾患の診断、予防および治療における前記タンパク質およびコード遺伝子由来の核酸配列の使用に関する。

**【特許請求の範囲】****【請求項 1】**

以下のポリペプチド：

(i) 配列番号：2 および配列番号：4 に記載のアミノ酸配列を含むポリペプチド；  
(i i) 核内ホルモンレセプターリガンド結合ドメインの活性を有するか、または(i)のポリペプチドと共に抗原決定基を有する、前記のフラグメントであるポリペプチド；または  
(i i i) (i) または(i i) の機能的等価物であるポリペプチド。

**【請求項 2】**

配列番号：2 または配列番号：4 に記載のアミノ酸配列からなる、請求項 1 記載のポリペ 10  
プチド。

**【請求項 3】**

LBDG3 ポリペプチドの核内ホルモンレセプターリガンド結合ドメイン領域を含む、請求項 1 の(i i) に記載のフラグメントであるポリペプチドであって、前記核内ホルモンレセプターリガンド結合ドメイン領域が配列番号：2 に記載のアミノ酸配列の残基 311 から 452 を含むと定義され、前記フラグメントが配列番号：2 に記載のアミノ酸配列の“LBDモチーフ”残基 ASP314、GLN315、LEU318 および LEU319、または等価な残基を保有し、さらに核内ホルモンレセプターリガンド結合ドメインの活性を保有する前記ポリペプチド。

**【請求項 4】**

請求項 1 の(i i i) に記載の機能的等価物であり、配列番号：2 に記載のアミノ酸配列と相同であり、“LBDモチーフ”残基 ASP314、GLN315、LEU318 および LEU319、または等価な残基を保有し、さらに核内ホルモンレセプターリガンド結合ドメインの活性を保有するポリペプチド。

**【請求項 5】**

機能的等価物が、LBDG3 ポリペプチドの核内ホルモンレセプターリガンド結合ドメイン領域と相同である請求項 4 に記載のポリペプチド。

**【請求項 6】**

BLASTバージョン 2.1.3 で NCBI (the National Center for Biotechnology Information ; <http://www.ncbi.nlm.nih.gov/>) によって特定されたデフォルトパラメーター { Blosum62  
マトリックス；ギャップ開放ペナルティー = 11 およびギャップ伸長ペナルティー = 1 }  
を用いて決定したとき、配列番号：2 に記載のアミノ酸配列、または核内ホルモンレセプターリガンド結合ドメイン活性を保有するそのフラグメントと 80 % を超える配列同一性、好ましくは 85 %、90 %、95 %、98 % または 99 % を超える配列同一性を有する、請求項 1 - 5 のいずれか 1 項に記載のフラグメントまたは機能的等価物。

**【請求項 7】**

配列番号：2 のいずれか一つによって与えられるアミノ酸配列を有するポリペプチド、または核内ホルモンレセプターリガンド結合ドメイン活性を保有するそのフラグメントと顕著な構造相同性を示す請求項 1 - 6 のいずれか 1 項に記載の機能的等価物。

**【請求項 8】**

配列番号：2 の配列に由来する 7 つもしくはそれより多い（例えば 8、10、12、14、16、18、20 またはそれより多い）アミノ酸残基からなる、請求項 1 の(i) のポリペプチドと共に抗原決定基を有する請求項 1 - 3 または 6 に記載のフラグメント。

**【請求項 9】**

請求項 1 - 8 のいずれか 1 項に記載のポリペプチドをコードする精製核酸分子。

**【請求項 10】**

配列番号：1 もしくは配列番号：3 に記載の核酸配列を有するか、または重複等価物もしくはそのフラグメントである請求項 9 に記載の精製核酸分子。

**【請求項 11】**

配列番号：1 のヌクレオチド 932 から 1357 を含むか、またはその重複等価物である請求項 50

9 または請求項 1 0 に記載の精製核酸分子のフラグメント。

【請求項 1 2】

高いストリンジエンシー条件下で請求項 9 - 1 1 のいずれか 1 項に記載の核酸分子とハイブリダイズする精製核酸分子。

【請求項 1 3】

請求項 9 - 1 2 のいずれか 1 項に記載の核酸分子を含むベクター。

【請求項 1 4】

請求項 1 3 に記載のベクターで形質転換された宿主細胞。

【請求項 1 5】

請求項 1 - 8 のいずれか 1 項に記載のポリペプチドと特異的に結合し、さらに好ましくは前記ポリペプチドの核内ホルモンレセプターリガンド結合ドメイン活性を阻害するリガンド。 10

【請求項 1 6】

抗体である請求項 1 5 に記載のリガンド。

【請求項 1 7】

請求項 1 - 8 のいずれか 1 項に記載のポリペプチドの発現レベルまたは活性レベルを増加または低下させる化合物。

【請求項 1 8】

請求項 1 - 8 のいずれか 1 項に記載のポリペプチドと結合し、前記ポリペプチドの生物学的作用のいずれをも誘発することがない請求項 1 7 に記載の化合物。 20

【請求項 1 9】

天然または改変された基質、リガンド、酵素、レセプターまたは構造的もしくは機能的摸倣体である請求項 1 7 または 1 8 に記載の化合物。

【請求項 2 0】

疾患の治療または診断で使用することを目的とする請求項 1 - 8 のいずれか 1 項に記載のポリペプチド、請求項 9 - 1 2 のいずれか 1 項に記載の核酸分子、請求項 1 3 に記載のベクター、請求項 1 5 または 1 6 に記載のリガンド、または請求項 1 7 - 1 9 のいずれか 1 項に記載の化合物。 30

【請求項 2 1】

患者の疾患を診断する方法であって、前記患者の組織で請求項 1 - 8 のいずれか 1 項に記載のポリペプチドをコードする天然の遺伝子の発現レベルを評価するか、または請求項 1 - 8 のいずれか 1 項に記載のポリペプチドの活性を評価し、さらに前記発現または活性レベルをコントロールレベルと比較することを含み、前記コントロールレベルと異なるレベルが疾患を示唆する前記疾患の診断方法。 30

【請求項 2 2】

*in vitro*で実施される請求項 2 1 に記載の方法。

【請求項 2 3】

( a ) 請求項 1 5 または請求項 1 6 に記載のリガンドをリガンド - ポリペプチド複合体の形成に適した条件下生物学的サンプルと接触させ；さらに ( b ) 前記複合体を検出する工程を含む、請求項 2 1 または請求項 2 2 に記載の方法。 40

【請求項 2 4】

a ) 患者由来の組織サンプルを、請求項 9 - 1 2 のいずれか 1 項に記載の核酸分子とプローブとの間でハイブリッド複合体を形成し得るストリンジエントな条件下核酸プローブと接触させる工程；

b ) コントロールサンプルと前記プローブを工程 a ) で用いた条件と同じ条件下で接触させる工程；さらに

c ) 前記サンプル中のハイブリッド複合体の存在を検出する工程；  
を含み、コントロールサンプル中のハイブリッド複合体レベルと異なるハイブリッド複合体レベルが患者サンプルで検出されることは疾患を示唆する、請求項 2 1 または請求項 2 2 に記載の方法。 50

**【請求項 25】**

- a ) 患者の組織由来の核酸サンプルを、請求項 9 - 12 のいずれか 1 項に記載の核酸分子とプライマーとの間でハイブリッド複合体を形成し得るストリンジエントな条件下で核酸プライマーと接触させる工程；  
 b ) コントロールサンプルと前記プライマーを工程 a ) で用いた条件と同じ条件下で接触させる工程；  
 c ) サンプル化された前記核酸を増幅する工程；および、  
 d ) 患者サンプルおよびコントロールサンプルの両方から、増幅された核酸レベルを検出する工程；

を含み、コントロールサンプル中の増幅核酸レベルと顕著に異なる増幅核酸レベルが患者サンプルで検出されることは疾患を示唆する、請求項 21 または請求項 22 に記載の方法。  
 10

**【請求項 26】**

以下の工程を含む請求項 21 または請求項 22 に記載の方法：

- a ) 疾患について検査しようとする患者から組織サンプルを得る工程；  
 b ) 請求項 9 - 12 のいずれかの項に記載の核酸分子を前記組織サンプルから単離する工程；および、  
 c ) 核酸分子中で疾患に関連する変異の存在を疾患の徴候として検出することによって患者を疾患について診断する工程。

**【請求項 27】**

核酸分子を増幅して増幅生成物を形成させ、前記増幅生成物中で変異の有無を検出することをさらに含む請求項 26 の方法。

**【請求項 28】**

核酸分子をストリンジエントな条件下で前記核酸分子とハイブリダイズする核酸プローブと接触させてハイブリッド二本鎖分子を形成させ、前記ハイブリッド二本鎖分子は疾患に関連する変異に対応するいずれかの部分でハイブリダイズしない核酸プローブ鎖部分を有し、および、疾患関連変異の有無を表示するものとして前記プローブ鎖のハイブリダイズしない部分の有無を検出することによって患者において変異の有無を検出する請求項 26 または 27 の方法。

**【請求項 29】**

30  
 疾患が以下から選択される請求項 21 - 28 のいずれか 1 項に記載の方法：  
 細胞増殖性疾患（新生物、メラノーマ、肺、結腸直腸、乳房、膵、頭部の腫瘍並びに他の固形腫瘍を含む）、骨髄増殖性疾患（例えば白血病、非ホジキンリンパ腫）、白血球減少症、血小板減少症、血管形成疾患、カポジ肉腫、自己免疫 / 炎症性疾患（アレルギー、炎症性腸疾患、関節炎、乾癬および気道の炎症、喘息および器官の移植拒絶を含む）、心脈管系疾患（高血圧、浮腫、アンギナ、アテローム性硬化症、血栓症、敗血症、ショック、再灌流障害、心不整脈および虚血を含む）、神経学的疾患（中枢神経系疾患、アルツハイマー病、脳損傷、脳卒中、筋萎縮性側索硬化症、不安、抑うつおよび痛みを含む）、発達障害、代謝性障害（真性糖尿病、骨粗しょう症、脂質代謝障害、甲状腺機能亢進、甲状腺機能低下、上皮小体機能亢進、高カルシウム血症、低カルシウム血症、高コレステロール血症、高脂血症および肥満を含む）、腎疾患（糸球体腎炎、腎血管性高血圧を含む）、皮膚疾患（アクネ、湿疹および創傷治癒を含む）、加齢の負の作用、エイズ、感染（ウイルス感染、細菌感染、真菌感染および寄生虫感染を含む）、並びに他の病的状態（特に核内ホルモンレセプターが関与するもの）。

**【請求項 30】**

核内ホルモンレセプタリガンド結合ドメインとしての請求項 1 - 8 のいずれか 1 項に記載のポリペプチドの使用。

**【請求項 31】**

核内ホルモンレセプタリガンド結合ドメイン活性を保有するタンパク質の発現を目的とする請求項 9 - 12 のいずれか 1 項に記載の核酸分子の使用。

10

20

30

40

50

**【請求項 3 2】**

請求項 1 - 8 のいずれか 1 項に記載のポリペプチドを利用して細胞 - 細胞接着を行わせる方法。

**【請求項 3 3】**

請求項 1 - 8 のいずれか 1 項に記載のポリペプチド、請求項 9 - 12 のいずれか 1 項に記載の核酸分子、請求項 13 に記載のベクター、請求項 15 または 16 に記載のリガンド、または請求項 17 - 19 のいずれか 1 項に記載の化合物を含む医薬組成物。

**【請求項 3 4】**

請求項 1 - 8 のいずれか 1 項に記載のポリペプチドまたは請求項 9 - 12 のいずれか 1 項に記載の核酸分子を含むワクチン組成物。

10

**【請求項 3 5】**

以下の疾患の治療用薬物の製造で使用される請求項 1 - 8 のいずれか 1 項に記載のポリペプチド、請求項 9 - 12 のいずれか 1 項に記載の核酸分子、請求項 13 に記載のベクター、請求項 15 または 16 に記載のリガンド、請求項 17 - 19 のいずれか 1 項に記載の化合物、または請求項 3 3 記載の医薬組成物 :

細胞増殖性疾患（新生物、メラノーマ、肺、結腸直腸、乳房、膵、頭部および頸部の腫瘍並びに他の固体腫瘍を含む）、骨髄増殖性疾患（例えば白血病、非ホジキンリンパ腫）、白血球減少症、血小板減少症、血管形成疾患、カポジ肉腫、自己免疫 / 炎症性疾患（アレルギー、炎症性腸疾患、関節炎、乾癬および気道の炎症、喘息および器官の移植拒絶を含む）、心脈管系疾患（高血圧、浮腫、アンギナ、アテローム性硬化症、血栓症、敗血症、ショック、再灌流障害、心不整脈および虚血を含む）、神経学的疾患（中枢神経系疾患、アルツハイマー病、脳損傷、脳卒中、筋萎縮性側索硬化症、不安、抑うつおよび痛みを含む）、発達障害、代謝性障害（真性糖尿病、骨粗しょう症、脂質代謝障害、甲状腺機能亢進、甲状腺機能低下、上皮小体機能亢進、高カルシウム血症、低カルシウム血症、高コレステロール血症、高脂血症および肥満を含む）、腎疾患（糸球体腎炎、腎血管性高血圧を含む）、皮膚疾患（アクネ、湿疹および創傷治癒を含む）、加齢の負の作用、エイズ、感染（ウイルス感染、細菌感染、真菌感染および寄生虫感染を含む）、並びに他の病的状態（特に核内ホルモンレセプターが関与するもの）。

20

**【請求項 3 6】**

請求項 1 - 8 のいずれか 1 項に記載のポリペプチド、請求項 9 - 12 のいずれか 1 項に記載の核酸分子、請求項 13 に記載のベクター、請求項 15 または 16 に記載のリガンド、請求項 17 - 19 のいずれか 1 項に記載の化合物、または請求項 3 3 に記載の医薬組成物を患者に投与することを含む、患者の疾患を治療する方法。

30

**【請求項 3 7】**

天然の遺伝子の発現またはポリペプチドの活性が、健常な対象者における発現または活性レベルと比較したとき罹患対象者で低下する疾患に対して、前記対象者に投与される前記ポリペプチド、核酸分子、ベクター、リガンド、化合物、または組成物がアゴニストである請求項 3 6 に記載の方法。

**【請求項 3 8】**

天然の遺伝子の発現またはポリペプチドの活性が、健常な対象者における発現または活性レベルと比較したとき罹患対象者で上昇する疾患に対して、前記対象者に投与される前記ポリペプチド、核酸分子、ベクター、リガンド、化合物または組成物がアンタゴニストである請求項 3 6 に記載の方法。

40

**【請求項 3 9】**

患者の疾患の治療的処置をモニターする方法であって、請求項 1 - 8 のいずれか 1 項に記載のポリペプチドの発現もしくは活性レベル、請求項 9 - 12 のいずれか 1 項に記載の核酸分子の発現レベルを前記患者由来の組織で一定期間にわたってモニターすることを含み、前記期間にわたってコントロールレベルに向かって前記発現または活性レベルが変化することが前記疾患の軽減の徴候である、前記患者の疾患の治療的処置をモニターする方法。

50

**【請求項 4 0】**

疾患の治療および/または診断において有効な化合物の特定方法であって、前記方法が、請求項1-8のいずれか1項に記載のポリペプチド、請求項9-12のいずれか1項に記載の核酸分子、または請求項13に記載の宿主細胞を、前記ポリペプチドまたは核酸に対する結合親和性を保有すると思われる1つまたは2つ以上の化合物と接触させ、さらに前記核酸分子またはポリペプチドと特異的に結合する化合物を選別することを含む前記有効な化合物の特定方法。

**【請求項 4 1】**

ストリンジエントな条件下で請求項9-12のいずれか1項に記載の核酸分子とハイブリダイズする核酸プローブを含む第一の容器；前記核酸分子の増幅に有用なプライマーを含む第二の容器；および疾患の診断を容易にすることを目的とする前記プローブおよびプライマーの使用のための指示書を含む疾患の診断に有用なキット。

**【請求項 4 2】**

ハイブリダイズしていないRNAを消化するための薬剤を保持する第三の容器をさらに含む請求項41のキット。

**【請求項 4 3】**

核酸分子のアレイを含むキットであって、前記核酸分子の少なくとも一つが請求項9-12のいずれか1項に記載の核酸分子である前記キット。

**【請求項 4 4】**

請求項1-8のいずれか1項に記載のポリペプチドと結合する1つまたは2つ以上の抗体；および前記抗体と前記ポリペプチドとの間の結合反応の検出に有用な試薬を含むキット。

**【請求項 4 5】**

請求項1-8のいずれか1項に記載のポリペプチドをより高レベルまたはより低レベルで発現するように、または前記ポリペプチドを発現しないように形質転換したトランスジェニックまたはノックアウト非ヒト動物。

**【請求項 4 6】**

請求項45に記載の非ヒトトランスジェニック動物を候補化合物と接触させ、さらに前記動物の疾患に対する前記化合物の影響を決定することによって、疾患の治療に有効な化合物をスクリーニングする方法。

**【発明の詳細な説明】****【0001】**

本発明は、本明細書において核内ホルモンレセプターのリガンド結合ドメインとして同定された、CAB55953.1と称される、新規なタンパク質並びに疾患の診断、予防および治療における前記タンパク質およびそのコード遺伝子に由来する核酸配列の使用に関する。

本明細書に引用した全ての刊行物、特許および特許出願は、引用により完全に本明細書に含まれるものとする。

**【0002】****(背景技術)**

薬剤の発見プロセスにおいて、機能ゲノム学(functional genomics)の時代の到来について根幹的な革命が現在進行している。“機能ゲノム学”という用語は、対象のタンパク質配列に機能を帰属させるためにバイオインフォマティクスツールを利用するアプローチに適用される。そのようなツールは、配列データの生成速度が、これらタンパク質配列に機能を割り当てる研究室の能力をはるかに越えるために、ますます必要性を増している。バイオインフォマティクスツールの潜在能力および精度が高まっているために、前記ツールは通常の生化学的特徴付け技術と急速に置き換えられつつある。実際、本発明の同定に用いた高度なバイオインフォマティクスツールは、今や、高い信頼性をもつ結果を産出する能力を有する。

配列データが利用可能になるにつれ、種々の研究機関および企業の組織がそれらを調査し、重要な発見が絶え間なく達成され続けている。しかしながら、研究および薬剤の発見の

ための標的として更に新たな遺伝子およびそれらがコードするポリペプチドを同定し特徴付ける必要性は引き続き存在している。

最近、未知の機能をもつ配列を評価するための注目すべきツールが本発明の出願人によって開発された。このツールは、同時係属国際特許出願第PCT/GB01/01105号の主題であるデータベースシステム（バイオペンジウム（Biopendium）検索データベースと称される）である。このデータベースシステムは、独占的技術を用いて作製され、利用可能な全てのタンパク質または核酸配列の完全な比較から作製された情報を含む集積データリソースから成る。

### 【0003】

別個のデータリソースからこれら配列データを一体化させたその背後の目的は、可能なかぎり多くのデータを、配列それ自体と各配列に関連する情報の両方に關して1つの完全なリソースにまとめることである。各配列と関係を有する全ての利用可能なデータ（入手可能な場合はコードするタンパク質の三次元構造に関するデータを含む）と一緒に統合し、各配列について知られている情報の最大限の利用を可能にし、したがって最も多くの知識に基づく予測がこれら配列の比較から入手することができる。前記データベースで作製され、各配列のエントリーに付隨する注釈は、生物学的に関連がある事柄を前記配列情報に付与する。

このデータリソースは、配列のみからタンパク質機能の正確な予測を可能にした。通常の技術を用いた場合、このような予測は、同じ機能ファミリーに属する他のタンパク質に対して高度な配列同一性（約20% - 30%の同一性）を示すタンパク質についてのみ可能である。既知の機能を有する他の近縁なタンパク質と低度な配列相同性を示すタンパク質に対しては、正確な予測が不可能である。

本件では、その配列が公的に利用可能なデータベースにCAB55953.1として記録されているタンパク質（NCBI GenBankヌクレオチドアクセション番号AL117480およびGenBankタンパク質アクセション番号CAB55953.1）が、核内ホルモンレセプターリガンド結合ドメインファミリーの新規なメンバーとして意味付けられる。

### 【0004】

#### I. 核内ホルモンレセプターリガンド結合ドメインに関する導入部

核内ホルモンレセプター遺伝子スーパーファミリー（表1参照）は、標的遺伝子の転写を調節する構造的に関連するタンパク質をコードする。前記タンパク質には、ステロイドホルモンおよび甲状腺ホルモン、ビタミン並びにリガンドが未だ発見されていない他のタンパク質に対するレセプターが含まれる。核内レセプターは2つの主要なドメイン、DNA結合ドメイン（DBD）およびリガンド結合ドメイン（LBD）から構成されている。DBDは、前記レセプターがモノマー、ホモダイマーまたはヘテロダイマーとして特異的なDNA配列と結合するように指令する。DBDは、核内レセプターでのみ見出される特殊なタイプのジンクフィンガーである。DBDを有する核内レセプターは、PROSITEコンセンサス配列（PS00031）とのマッチングについて検索することによって容易に配列レベルで同定することができる。

リガンド結合ドメイン（LBD）は同系のホルモンと結合しこれに応答する。LBDと結合したリガンドは、すでに結合していた“核内レセプターコリプレッサー”を排斥する構造的変化をひき起こす。続いて前記コリプレッサーによって以前に占有されていた部位は空席になり、“核内レセプターコアクチベーター”で補充される。このリガンドによってひき起こされるコアクチベーターによるコリプレッサーの交換が、リガンド結合が標的遺伝子の転写活性化をもたらすメカニズムである。全てのリガンド結合ドメインはコンセンサス配列、“LBDモチーフ”（表2参照）を含み、前記モチーフは、コリプレッサー結合およびコアクチベーター結合を仲介する。LBDは、今日まで全ての核内ホルモンレセプター標的薬剤のための結合部位であり、したがって新規なリガンド結合ドメインは魅力的な薬剤標的であるので、それらの同定が所望される。リガンド結合ドメインは低い配列同一性（約15%）しか共有しないが、非常に類似する構造を有し、したがって、ゲノムスレッダー（Genome Threader）のような構造に基づく関係付けツール（structure-based relationshi

10

20

30

40

50

p tool)のための理想的なターゲットである。

PROSITEのような基本的な検索ツールを用いて、DBDが存在することによって、さらにそれをもとに推測されたLBDによって、多くのタンパク質配列が核内ホルモンレセプターとしてパブリックドメインすでに注釈付けされている。このために、ゲノムスレッダーによって同定される、核内レセプターとして注釈されてないどの新規LBDも完全にDBDを欠くであろうということが予期される。LBDを有するがDBDを欠くタンパク質の先例は、DAX1によって提供される。したがって、我々は、これらDBDのないヒットを“核内ホルモンレセプター”ではなく“核内ホルモンレセプターリガンド結合ドメイン”を含むものと注釈付けする。

【0005】

10

【表1】

表1：核内ホルモンレセプタースーパーファミリー

ファミリー：ステロイドホルモンレセプター
サブファミリー：グルココルチコイドレセプター
プロゲステロンレセプター
アンドロゲンレセプター
エストロゲンレセプター
ファミリー：甲状腺ホルモンレセプター様因子
サブファミリー：レチノイン酸レセプター (RAR)
レチノイドXレセプター (RXR)
甲状腺ホルモンレセプター
ビタミンDレセプター
NFGI-B
FTZ-F1
ペルオキシソーム増殖因子活性化レセプター (PPAR)
クジソンレセプター
レチノイドオーファンレセプター (ROR)
タイレス (Tailless) / Coup
HNF-4
CF1
Knia
ファミリー：DAX1
サブファミリー：DAX1

20

30

40

【0006】

下表2：“LBDモチーフ”。最上段の数字はモチーフ内の残基の位置を示す。文字は1文字コードによるアミノ酸を示す。1つの縦の欄内の文字は全てモチーフ内のその位置について許容される。例えば、L、I、A、V、M、F、YまたはWは“LBDモチーフ”的最初の位置を占めることができる。位置4と8の間、および位置9と12の間で見出される残基の数には変動が観察されることを述べる。“LBDモチーフ”は、核内ホルモンレセプターリガンド結合ドメインの681の配列を整列化し、残基の保存パターンを同定することによって構築された。

【0007】

【表2】

1	2	3	4	5	6	7	8	9	10	11	12	13
L	任意の2残基	任意の3残基 (又は、2残基、 または4残基)	L	任意の2残基 (又は、1 残基、 または3残 基)	D	Q	任意の2残 基 (又は、1 残基、 または3残 基)	L	L	任意の2残 基 (又は、1 残基、 または3残 基)	A	A
I			I		E	N		I	I		V	V
A			A		R			M	M		F	F
V			V		H			Y	Y		W	W
M			M		K							
F			F		S							
Y			Y		T							
W			W									

10

20

## 【 0 0 0 8 】

I I . 核内ホルモンレセプターおよび疾患

核内ホルモンレセプターは、多様な生理学的機能において役割を果たすことが示されている。それらの多くは疾患プロセスにおいて役割を果たし得る（表3参照）

## 【 0 0 0 9 】

## 表3：核内ホルモンレセプターと疾患

## 【表3】

核内ホルモンレセプター	疾患	
アンドロゲンレセプター	<p>アンドロゲン非感受性症候群 (Lubahn et al. 1989, Proc. Natl. Acad. Sci. USA 86 : 9534–9538)</p> <p>ライフェンスタイン症候群 (Wooster et al. 1992, Nat. Genet. 2 : 132–134)</p> <p>X連鎖劣性脊髄および延髄筋萎縮 (MacLean et al. 1995 Mol. Cell. Endocrinol. 112 : 133–141)</p> <p>男性乳癌 (Wooster et al. 1992, Nat. Genet. 2 : 132–134)</p>	10
グルココルチコイドレセプター	<p>ネルソン症候群 (Karl et al. 1996, J. Clin. Endocrinol. Metab. 81 : 124–129)</p> <p>グルココルチコイド耐性急性T細胞白血病 (Hala et al. 1996, Int. J. Cancer 68 : 663–668)</p>	20
鉱質コルチコイドレセプター	仮性低アルドステロン症 (Chung et al. 1995, J. Clin. Endocrinol. Metab. 80 : 3341–3345)	
エストロゲンレセプター $\alpha$	E R $\alpha$ 発現はヒト乳癌サブセットで上昇する。タモキシフェンの適用は乳癌進行を予防する主要な治療法である。残念ながら、E R $\alpha$ 陽性乳癌の35%はタモキシフェン耐性である (Petrangeli et al. 1994, J. Steroid Biochem. Mol. Biol. 49 : 327–331)	30
ビタミンD3レセプター	ビタミンD3レセプターの変異は、ビタミンD3欠乏に対する表現型 (Rickets) に類似する遺伝性疾患をもたらす (Hughes et al. 1988, Science 242 : 1702–1725)	
レチノイン酸レセプター $\alpha$	急性骨髓性白血病 (Lavau & Dejean 1994, Leukemia 8 : 9–15)	
甲状腺ホルモンレセプター $\beta$	“甲状腺ホルモンに対する普遍的耐性” (GRTH) (Refetoff 1994, Thyroid 4 : 345–349)	
DAX1	X連鎖先天性副腎形成不全 (AHC) および性機能不全 (Ito et al. 1997, Mol. Cell. Biol. 17 : 1476–1483)	40

## 【0010】

したがって、核内ホルモンレセプターのLBDと結合するリガンドによる核内ホルモンレセプターの変化は、疾患の表現型を変化させる手段を提供する。したがって、新規核内ホルモンレセプターのリガンド結合ドメインの同定は、それらタンパク質が上記で同定した疾患や他の症状において役割を果たし得るために強く希求される。したがって新規核内ホルモンレセプターのリガンド結合ドメインの同定は、疾患（特に表3で同定したような疾患

) の治療および診断と密接な関連を有する。

【 0 0 1 1 】

(発明の詳細な説明)

本発明は、CAB55953.1タンパク質が核内ホルモンレセプタリガンド結合ドメインとして機能するという発見を基にしている。

CAB55953.1タンパク質の場合、前記タンパク質配列の残基311-452を含む領域は、ヒトレチノイン酸レセプター (PDBコード1EXA: A) の残基74(Ser255)から残基216(Ala397)と等価なフォールドを採用することが見出された。ヒトレチノイン酸レセプターは、核内ホルモンレセプタリガンド結合ドメインとして機能することが知られている。さらにヒトレチノイン酸レセプターの“LBDモチーフ”残基、ASP258、GLN259、LEU262およびLEU263は、CAB55953.1ではそれぞれASP314、GLN315、LEU318およびLEU319として保存されている。前記の関係は単にヒトレチノイン酸レセプターとの関係ではなく、むしろ全体として核内ホルモンレセプタリガンド結合ドメインファミリー全体との関係である。したがって、ヒトレチノイン酸レセプター (1EXA: A) とCAB55953.1とのゲノムスレッダー (Genome Threader) (商標) アラインメントの参照によって、CAB55953.1のASP314、GLN315、LEU318およびLEU319は“LBDモチーフ”残基を形成すると予測される。

等価なフォールドおよび“LBDモチーフ”残基保存が組み合わされて、CAB55953.1の前記領域の機能的注釈付けを可能とし、したがってこの領域を含むタンパク質が核内ホルモンレセプタリガンド結合ドメイン活性を有するという機能的注釈付けを可能とする。

【 0 0 1 2 】

本発明の第一の特徴の態様ではポリペプチドが提供され、前記ポリペプチドは、

(i) 配列番号：2に記載のアミノ酸配列を含む；

(ii) 核内ホルモンレセプタリガンド結合ドメインの活性を有するか、または(i)のポリペプチドと共に抗原決定基を有する、配列番号：2のフラグメントである；または

(iii) (i)または(ii)の機能的等価物である。

配列番号：2に記載の配列を有するポリペプチドは、以下では“LBDG3ポリペプチド”と称する。

アクセッショ番号CAB55953.1を有する配列は、現在N末端まで延長されている。新しい配列はCAC14946.1と称され、長さは797アミノ酸である。この配列は本明細書中で配列番号：4として示されている。その新しい配列は、もちろん核内ホルモンレセプタリガンド結合ドメイン活性を有するというCAB55953.1ポリペプチド配列の注釈を変更するのではなく、単にそのタンパク質のN末端ドメインを延長するだけである。配列番号：4に記載の配列を有するポリペプチドを、以降は“LBDG3完全長ポリペプチド”と称する。

【 0 0 1 3 】

したがって、本発明の第一の特徴による第二の態様では、ポリペプチドが提供され、前記ポリペプチドは、

(i) 配列番号：4に記載のアミノ酸配列を含む；

(ii) 核内ホルモンレセプタリガンド結合ドメインの活性を有するか、または(i)のポリペプチドと共に抗原決定基を有するそのフラグメントである；または

(iii) (i)または(ii)の機能的等価物である。

好ましくは前記ポリペプチドは、

(i) 配列番号：2または配列番号：4に記載のアミノ酸配列から成る；

(ii) 核内ホルモンレセプタリガンド結合ドメインの活性を有するか、または(i)のポリペプチドと共に抗原決定基を有するそのフラグメントである；または

(iii) (i)または(ii)の機能的等価物である。

本発明の前記特徴にしたがえば、上記(ii)に記載の好ましいポリペプチドフラグメントは、核内ホルモンレセプタリガンド結合ドメインの活性に必要なものと予測される、LBDG3ポリペプチドの領域 (以下では“LBDG3核内ホルモンレセプタリガンド結合ドメイン領域”)を含むか、またはその変種であって“LBDモチーフ”(ASP314、GLN315、LEU31

10

20

30

40

50

8およびLEU319、または等価な残基)を保有する。本明細書で明確にされるように、LBDG3核内ホルモンレセプタリガンド結合ドメイン領域は、LBDG3ポリペプチド配列の残基311と残基452の間に広がっていると考えられる。

本発明の前記特徴はまた、ポリペプチドのフラグメントおよび上記で定義したこれらポリペプチドフラグメントの変種を取り込んだ融合タンパク質を含むが、ただし前記融合タンパク質は核内ホルモンレセプタリガンド結合ドメインとしての活性を保有することを条件とする。

#### 【0014】

第二の特徴では、本発明は、本発明の第一の特徴のポリペプチドをコードする精製核酸分子を提供する。好ましくは前記精製核酸分子は、配列番号：1 (LBDG3ポリペプチドをコードする)または配列番号：3 (LBDG3完全長ポリペプチドをコードする)に記載の核酸配列を有するか、または前記配列の重複等価物またはフラグメントである。好ましい核酸フラグメントは、上記(iii)に記載のポリペプチドフラグメント、好ましくはLBDG3核内ホルモンレセプタリガンド結合ドメイン領域を含むポリペプチドフラグメントをコードするものであるか、上記に定義したこれらフラグメントの変種をコードするものである。

第三の特徴では、本発明は、本発明の第二の特徴の核酸分子と高ストリンジエンシー条件下でハイブリダイズする精製核酸分子を提供する。

第四の特徴では、本発明は、発現ベクターのように本発明の第二または第三の特徴の核酸分子を含むベクターを提供する。

第五の特徴では、本発明は、本発明の第四の特徴のベクターで形質転換された宿主細胞を提供する。

第六の特徴では、本発明は、本発明の第一の特徴のポリペプチドと特異的に結合し、核内ホルモンレセプタリガンド結合ドメインの活性を阻害するリガンドを提供する。

第七の特徴では、本発明は、本発明の第一の特徴のポリペプチドをコードする天然の遺伝子の発現を変化させるか、または本発明の第一の特徴のポリペプチドの活性を調節するために有効な化合物を提供する。

本発明の第七の特徴の化合物は、前記ポリペプチドの遺伝子の発現レベルまたは活性レベルを増加(作動)させるか、または低下(拮抗)させる。重要なことには、それぞれLBDG3ポリペプチドのLBDG3核内ホルモンレセプタリガンド結合ドメイン領域として本明細書で明確にされた領域の機能を同定することによって、核内ホルモンレセプタリガンド結合ドメインが関与する疾患の治療および/または診断に有効な化合物を同定することが可能なスクリーニング方法をデザインすることができる。

#### 【0015】

第八の特徴では、本発明は、診断または治療で使用するための本発明の第一の特徴のポリペプチド、または本発明の第二もしくは第三の特徴の核酸分子、または本発明の第四の特徴のベクター、または本発明の第五の特徴のリガンド、または本発明の第六の特徴の化合物を提供する。

本発明者は、LBDG3のmRNAがヒト脳において有意なレベルで発現していることを発見した。このことは、ヒト疾患の状態とLBDG3のリガンド結合ドメインに対するアゴニストおよびアンタゴニストの開発に潜在的な関連を与えることから注目すべきことであり、以下を含む種々のヒト疾患における治療的処置に対して可能性を提供する：細胞増殖性疾患(新生物、メラノーマ、肺、結腸直腸、乳房、膵、頭部および頸部の腫瘍並びに他の固形腫瘍を含む)、骨髄増殖性疾患(例えは白血病、非ホジキンリンパ腫、白血球減少症、血小板減少症、血管形成疾患、カボジ肉腫)、自己免疫/炎症性疾患(アレルギー、炎症性腸疾患、関節炎、乾癬および気道の炎症、喘息および器官の移植拒絶を含む)、心脈管系疾患(高血圧、浮腫、アンギナ、アテローム性動脈硬化症、血栓症、敗血症、ショック、再灌流障害、心不整脈および虚血を含む)、神経学的疾患(中枢神経系疾患、アルツハイマー病、脳損傷、脳卒中、筋萎縮性側索硬化症、不安、抑うつおよび痛みを含む)、発達障害、代謝性障害(真性糖尿病、骨粗しょう症、脂質代謝異常、甲状腺機能亢進、甲状腺機能

10

20

30

40

50

低下、上皮小体機能亢進、高カルシウム血症、低カルシウム血症、高コレステロール血症、高脂血症および肥満を含む)、腎疾患(糸球体腎炎、腎血管性高血圧を含む)、皮膚疾患(アクネ、湿疹および創傷治癒を含む)、加齢の負の作用、エイズ、感染(ウイルス感染、細菌感染、真菌感染および寄生虫感染を含む)、並びに他の病的状態(特に核内ホルモンレセプターが関与するもの)。

DNA結合ドメインが存在しないリガンド結合ドメインを含むLBDG3のような“非古典的”核内ホルモンレセプターの発見は、脳におけるステロイド(神経ステロイドとして知られる)の広範な効果および一般的にそのような効果は転写活性を必要とする既知の古典的ステロイドホルモン核内レセプターを介さずに媒介されることを一貫して報告している既知文献と一致している。一例を挙げると、神経ステロイドは、特にGABA、NMDAおよびシグマレセプターのようなレセプターの域において神経伝達に影響することが示されている。神経ステロイドは、神経保護的役割を果たすことが示されている。したがってLBDG3に対するアゴニスト(またはアンタゴニスト)の開発を介する治療的処置は、梗塞または出血(脳卒中)、並びに中枢神経系および脊髄の損傷のような脳血管性疾患に続く痴呆、パーキンソン病および神経変性のような神経変性状態の治療において役割を有し得る。さらに神経ステロイドは、認識処理、空間学習および記憶、不安、並びに常習行為パターンを引き起こす欲求のような行為に影響することが示されている。したがってLBDG3に対するアゴニストおよびアンタゴニストの開発は、痴呆、学習困難、不安、常習行為(例えばアルコール症、摂食障害および薬物常習であるが、これだけに限られない)を治療するための治療的処置をもたらし得る。

10

20

30

#### 【0016】

本発明者は、LBDG3が脳において独占的には発現されないことを示した。有意なレベルのmRNAが副腎、卵巣、精巣および胸腺においても見出されている。副腎、卵巣および精巣はステロイドの生合成にとって意味のある部位であり、その活性はLBDG3にとってのステロイド作用における役割(それだけに限られない)と一致している。それらの組織におけるLBDG3の発見は、以下に挙げる疾患(これだけに限られない)の治療のためのアンタゴニストおよびアゴニストの開発を支持する:高血圧、感染症のストレスを含むストレスに対する応答、塩類および水分恒常性の調節、排卵調節を通じての受胎率制御(不妊症および避妊)、着床調節(不妊症および避妊)および精子形成の調節(不妊症および避妊)。さらに、そのような薬物は、良性前立腺肥大、前立腺癌、卵巣癌および精巣癌のようなステロイド感受性腫瘍を治療することにおいても価値があり得る。

胸腺におけるLBDG3の発見は、T細胞の発生における役割と一致している。したがってLBDG3に対して開発したアゴニストおよびアンタゴニストは、以下のような疾患の進行においてT細胞を調節することに役割を果たし得る:自己免疫疾患およびアレルギー(I型真性糖尿病を含む)、慢性関節リウマチ、多発性硬化症、乾癬、糸球体症に起因する腎不全、強皮症、炎症性腸疾患(クローン病および潰瘍性大腸炎)、移植拒絶、喘息、アトピー性皮膚炎、湿疹、エイズ、感染症(ウイルス感染、細菌感染、真菌感染および寄生虫感染を含む)、並びに他の病的状態(特に核内ホルモンレセプターが関与するもの)。

40

#### 【0017】

第九番目の特徴では、本発明は患者で疾患を診断する以下の工程を含む方法を提供する:本発明の第一の特徴のポリペプチドをコードする天然の遺伝子の発現レベル、または本発明の第一の特徴のポリペプチドの活性レベルを前記患者由来の組織で評価し、さらに前記発見または活性レベルをコントロールレベルと比較する工程であって、この場合前記コントロールレベルと異なるレベルは疾患を示唆する。前記の方法は好ましくはin vitroで実施されるであろう。同様な方法は患者での疾患治療のモニタリングに使用することができる。この場合、時間の経過にしたがってポリペプチドまたは核酸分子の発現もしくは活性レベルがコントロールレベルに向かって変化するのは疾患の緩解の指標となる。

本発明の第一の特徴のポリペプチドを検出する好ましい方法は以下の工程を含む:(a)本発明の第六の特徴のリガンド(例えば抗体)を生物学的サンプルとリガンド-ポリペプチド複合体の形成に適した条件下で接触させる工程;および、(b)前記複合体を検出す

50

る工程。

本発明の第九番目の特徴に記載の方法には、例えば短いプローブによる核酸ハイブリダイゼーション法、点変異分析、ポリメラーゼ連鎖反応(PCR)増幅、および異常なタンパク質レベルを検出する抗体を用いる方法といった種々の異なる方法が存在することは、当業者には明らかであろう。同様な方法を短期または長期ベースで用いて、モニターされる疾患の治療を可能にすることができます。本発明はまた前記疾患診断方法で有用なキットも提供する。

#### 【0018】

第十番目の特徴では、本発明は、核内ホルモンレセプタリガンド結合ドメインとしての本発明の第一の特徴のポリペプチドの使用を提供する。本発明はまた、核内ホルモンレセプタリガンド結合ドメインの活性をもつタンパク質の発現のために本発明の第二または第三の特徴に記載の核酸分子の使用を提供する。本発明はまた核内ホルモンレセプタリガンド結合ドメイン活性を作用させる方法を提供し、前記方法は本発明の第一の特徴のポリペプチドを利用する。

10

第十一番目の特徴では本発明は医薬組成物を提供し、前記医薬組成物は、本発明の第一の特徴のポリペプチド、または本発明の第二もしくは第三の特徴の核酸分子、または本発明の第四の特徴のベクター、または本発明の第六の特徴のリガンド、または本発明の第七の特徴の化合物を医薬的に許容できる担体と組合せて含有する。

第十二番目の特徴では、本発明は、疾患の診断または治療を目的とする医薬品の製造で使用するために、本発明の第一の特徴のポリペプチド、または本発明の第二のもしくは第三の特徴の核酸分子、または本発明の第四の特徴のベクター、本発明の第六の特徴のリガンド、または本発明の第七の特徴の化合物を提供する。前記診断または治療される疾患は、例えば、細胞増殖性疾患(新生物、メラノーマ、肺、結腸直腸、乳房、膵、頭部および頸部の腫瘍並びに他の固形腫瘍、骨髄増殖性疾患(例えば白血病、非ホジキンリンパ腫、白血球減少症、血小板減少症、血管形成疾患、カポジ肉腫)、自己免疫/炎症性疾患(アレルギー、炎症性大腸疾患、関節炎、乾癬および気道の炎症、喘息および器官の移植拒絶を含む)、心脈管系疾患(高血圧、浮腫、アンギナ、アテローム性硬化症、血栓症、敗血症、ショック、再灌流障害、心不整脈および虚血を含む)、神経学的疾患(中枢神経系疾患、アルツハイマー病、脳損傷、卒中、筋萎縮性側索硬化症、不安、抑うつおよび痛みを含む)、発達障害、代謝性障害(真性糖尿病、骨粗しょう症、脂質代謝障害、甲状腺機能亢進、甲状腺機能低下、上皮小体機能亢進、高カルシウム血症、低カルシウム血症、高コレステロール血症、高脂血症および肥満を含む)、腎疾患(糸球体腎炎、腎血管性高血圧を含む)、皮膚疾患(アクネ、湿疹および創傷治癒を含む)、加齢の負の作用、エイズ、感染(ウイルス感染、細菌感染、真菌感染および寄生虫感染を含む)、並びに他の病的状態(特に核内ホルモンレセプターが関与するもの)である。

20

#### 【0019】

第十三番目の特徴では、本発明は患者の疾患を治療する方法を提供し、前記方法は、本発明の第一の特徴のポリペプチド、または本発明の第二もしくは第三の特徴の核酸分子、または本発明の第四の特徴のベクター、または本発明の第六の特徴のリガンド、または本発明の第七の特徴の化合物を患者に投与することを含む。

30

本発明の第一の特徴のポリペプチドをコードする天然の遺伝子の発現、または本発明の第一の特徴のポリペプチドの活性が、健常な対象者の発現または活性レベルと比較したとき罹患対象者で低下する疾患の場合、前記患者に投与される前記ポリペプチド、核酸分子、リガンドまたは化合物はアゴニストであろう。逆に、前記天然の遺伝子の発現、または前記ポリペプチドの活性が、健常な対象者の発現または活性レベルと比較したとき罹患対象者で上昇する疾患の場合、前記患者に投与される前記ポリペプチド、核酸分子、リガンドまたは化合物はアンタゴニストであろう。前記アンタゴニストの例にはアンチセンス核酸分子、リボザイムおよびリガンド(例えば抗体)が含まれる。

40

第十四番目の特徴では、本発明は、本発明の第一の特徴のポリペプチドを高レベルで、または低レベルで発現させるために、または全く発現させないために形質転換したトランス

50

ジェニックまたは遺伝子ノックアウト非ヒト動物を提供する。前記トランスジェニック動物は、疾患の研究用モデルとして非常に有用であり、さらに前記疾患の治療または診断に有効な化合物の同定を目的とするスクリーニング方法で用いることができる。

【0020】

本発明を利用するためには用いることができる標準的な技術および方法の要旨は下記で提供される。本発明は、記載した同定の方法論、プロトコル、細胞株、ベクターおよび試薬に限定されないことは理解されよう。本明細書で用いられる専門用語は単に個々の態様を説明するためのものであり、前記用語によって本発明の範囲を限定しようとするものではないこともまた理解されよう。本発明の範囲は添付の請求の範囲の用語によってのみ限定される。

10

本明細書では、ヌクレオチドおよびアミノ酸についての標準的な略語が用いられる。

本発明の実施では別に指示がなければ、分子生物学、微生物学、リコンビナントDNA技術および免疫学の通常の技術が用いられるであろう。前記技術は当業者の技術範囲内である。

前記のような技術は文献で完全に説明されている。特に適切な解説書の例には以下が含まれる：Sambrook Molecular Cloning；A Laboratory Manual, Second Edition (1989)；DNA Cloning, Vol. I and II (D.N. Glover ed. 1985)；Oligonucleotide Synthesis (M.J. Gait ed. 1984)；Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984)；Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984)；Animal Cell Culture (R.I. Freshney ed. 1986)；Immobilized Cells and Enzymes (IRL Press, 1986)；B. Perbal, A Practical Guide to Molecular Cloning (1984)；the Methods in Enzymology series (Academic Press, Inc.)特にVol. 154 & 155；Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory)；Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London)；Scopes, (1987) Protein Purification : Principles and Practice, Second Edition (Springer Verlag, NY)；およびHandbook of Experimental Immunology, Vols. I - IV (D.M. Weir and C.C. Blackwell eds. 1986)。

20

【0021】

本明細書において用いる“ポリペプチド”という用語は、ペプチド結合または改変ペプチド結合によって互いに結合した2つまたは3つ以上のアミノ酸を含む任意のペプチドまたはタンパク質（すなわちペプチドイソスター）が含まれる。この用語は、短鎖（ペプチドおよびオリゴペプチド）および長鎖（タンパク質）の両方を指す。

30

本発明のポリペプチドは成熟タンパク質の形態を有するものでもよく、またプレ-、プロ-またはプレプロ-タンパク質であってプレ-、プロ-またはプレプロ-部分の切断によって活性化され、活性成熟ポリペプチドを生じるタンパク質でもよい。そのようなポリペプチドでは、プレ-、プロ-またはプレプロ-配列はリーダー配列もしくは分泌配列であっても、または成熟ポリペプチド配列の精製のために用いられる配列であってもよい。

本発明の第一の特徴のポリペプチドは融合タンパク質の一部分を形成することができる。例えば、1つまたは2つ以上の付加アミノ酸配列を含むことがしばしば有利である。前記付加アミノ酸配列は、例えばリコンビナント形成時に、分泌もしくはリーダー配列、プロ-配列、精製を促進する配列、またはより高いタンパク質安定性を付与する配列を含んでもよい。あるいは、または前記に加えて、前記成熟ポリペプチドを別の化合物、例えば前記ポリペプチドの半減期を増加させるような化合物（例えばポリエチレングリコール）を融合させることができる。

40

【0022】

ポリペプチドは、天然のプロセス（例えば翻訳後プロセッシング）によって、または当業者に周知の化学的改変技術によって改変された、20の遺伝子コードアミノ酸以外のアミノ酸を含んでいてもよい。本発明のポリペプチドに一般的に存在する公知の改変にはグリコシリ化、脂質付加、硫化、-カルボキシリ化（例えばグルタミン酸残基の）、ヒドロ

50

キシリ化およびADP-リボシリ化がある。他の可能な改変には、アセチル化、アシリ化、アミド化、フラビンの共有結合付加、ヘム部分の共有結合付加、ヌクレオチドまたはヌクレオチド誘導体の共有結合付加、脂質誘導体の共有結合付加、ホスファチジルイノシトールの共有結合付加、架橋、環状化、ジスルフィド結合形成、脱メチル化、共有結合架橋の形成、システインの形成、ピログルタメートの形成、ホルミル化、GPIアンカー形成、ヨード化、メチル化、ミリストイル化、酸化、タンパク分解性プロセッシング、リン酸化、ブレニル化、ラセミ化、セレノイル化、タンパク質へのトランスファーRNA媒介性アミノ酸付加(例えばアルギニル化)およびユビキチン結合が含まれる。

改変は、ペプチド骨格、アミノ酸側鎖およびアミノまたはカルボキシ末端を含むポリペプチド内のいずれの場所に生じてもよい。実際、共有結合改変によるポリペプチドのアミノまたはカルボキシ末端またはその両端の妨害(blockage)は、天然に存在するポリペプチドおよび合成ポリペプチドで一般的であり、そのような改変は本発明のポリペプチドにも存在し得る。

#### 【0023】

ポリペプチド内に生じる改変は多くの場合、ポリペプチドが生成される方法の機能であろう。組換えによって生成されるポリペプチドの場合、大部分の改変の性質および程度は、個々の宿主細胞の改変能力および問題のポリペプチドのアミノ酸配列に存在する改変シグナルによって決定されるであろう。例えば、グリコシリ化パターンは異なる種類の宿主細胞間で変動するであろう。

本発明のポリペプチドは任意の適切な様式で調製することができる。そのようなポリペプチドには、単離された天然に存在するポリペプチド(例えば細胞培養から精製)、組換えにより生成されたポリペプチド(融合タンパク質を含む)、合成により生成されたポリペプチド、または前記方法を併用して生成されたポリペプチドが含まれる。

本発明の第一の特徴の機能的に等価なポリペプチドは、LBDG3ポリペプチドまたはLBDG3完全長ポリペプチドに対して相同なポリペプチドであり得る。2つのポリペプチドは、前記ポリペプチドの一方の配列が他方のポリペプチドの配列に対して充分な同一性または類似性を有する場合、本明細書で用いられる用語のように“相同である”と称される。“同一性”とは、整列化した配列のどの特定の場所においても、アミノ酸残基が前記配列間で同一であることを示す。“類似性”は、整列化した配列のいずれの特定の場所においても、アミノ酸残基が前記配列間で類似の種類であることを示す。同一性および類似性の度合いは容易に計算できる(Computational Molecular Biology, A.M. Lesk ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, D.W. Smith ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, A.M. Griffin and H.G. Griffin eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, G. von Heijne, Academic Press, 1987; およびSequence Analysis Primer, M. Gribskov and J. Devereux eds., M. Stockton Press, New York, 1991)。

#### 【0024】

したがって、相同なポリペプチドには、LBDG3ポリペプチドまたはLBDG3完全長ポリペプチドの天然の生物学的変種(例えば前記ポリペプチドが由來した種における対立形質変種または地理的変種)および変異体(例えばアミノ酸置換、挿入または欠失を含む変異体)が含まれる。前記変異体は、1つまたは2つ以上のアミノ酸残基が保存的または非保存的アミノ酸残基(好ましくは保存的アミノ酸残基)で置換されているポリペプチドを含んでもよく、さらにそのような置換アミノ酸残基は遺伝コードでコードされたものでもそうでなくともよい。典型的な前記の置換は、Ala、Val、LeuおよびIle間で; SerとThr間で; 酸性残基AspとGlu間で; AsnとGln間で、塩基性残基LysとArg間で; または芳香族残基PheとTyr間で生じる。特に好ましいものは、いくつか(すなわち5から10、1から5、1から3、1から2、または単に1つ)のアミノ酸が任意の組合せで置換されたまたは欠失または付加された変種である。特に好ましいものは、タンパク質の特性および活性を変化させないサイレント置換、付加および欠失である。さらにこれに関して特に好ましいものは保存

10

20

30

40

50

的置換である。

前記変異体にはまた、1つまたは2つ以上のアミノ酸残基が置換基を含むポリペプチドが含まれる。

#### 【0025】

典型的には、2つのポリペプチド間で80%を越える同一性（好ましくは特定の領域で）が機能的等価物を示すと考えられる。好ましくは、本発明の第一の特徴の機能的に等価なポリペプチドは、LBDG3ポリペプチドもしくはLBDG3完全長ポリペプチドに関して、またはその活性なフラグメントに関して80%を越える配列同一性を有する。より好ましいポリペプチドは、LBDG3ポリペプチドもしくはLBDG3完全長ポリペプチドまたはその活性なフラグメントに関してそれぞれ85%、90%、95%、98%または99%を越える同一性を有する。

本明細書で言及される同一性のパーセンテージは、BLASTバージョン2.1.3でNCBI(the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>)によって特定されたデフォルトパラメーター { Blosum62マトリックス；ギャップ開放(open)ペナルティー = 11 およびギャップ伸長(extension)ペナルティー = 1 } を用いて決定されるとおりである。

本件では、LBDG3またはLBDG3完全長ポリペプチドの好ましい活性フラグメントは、LBDG3の核内ホルモンレセプタリガンド結合ドメイン領域を含むものであって残基ASP314、GLN315、LEU318およびLEU319、または等価な残基を有する“LBDモチーフ”を保有するものである（LBDG3完全長ポリペプチドでは、当該残基はASP551、GLN552、LEU555およびLEU556である）。“等価な残基”とは、“LBDモチーフ”残基と等価である残基を意味するが、ただし核内ホルモンレセプタリガンド結合ドメイン領域が核内ホルモンレセプタリガンド結合ドメインとしての活性を保持していることを条件とする。例えば、ASP314はGLUによって置換することができる。例えば、GLN315は、ASN、ARG、HIS、LYS、SERまたはTHRと置換することができる。例えば、LEU318は、ILE、ALA、VAL、MET、PHE、TYRまたはTRPと置換することができる。例えば、LEU319は、ILE、ALA、VAL、MET、PHE、TYRまたはTRPと置換することができる。したがって本発明のこの特徴は、LBDG3ポリペプチドの核内ホルモンレセプタリガンド結合ドメイン領域に関して80%を越える同一性、好ましくは85%、90%、95%、98%または99%を越える同一性をそれぞれ有するポリペプチド、および、ASP314、GLN315、LEU318およびLEU319、または等価な残基をもつ“LBDモチーフ”を保有するものを含む。上記で考察したように、LBDG3の核内ホルモンレセプタリガンド結合ドメイン領域は、LBDG3ポリペプチド配列の残基311と残基452の間に広がっていると考えられる。LBDG3完全長ポリペプチドでは、当該境界はアミノ酸残基548および689である。

#### 【0026】

本発明の第一の特徴の機能的に等価なポリペプチドはまた、1つまたは2つ以上の構造的アラインメント技術を用いて同定されたポリペプチドであってもよい。例えば、バイオペンジウム(Biopendium)検索データベースを作製するために用いられた検索ツールの1つの特徴を形成するインファーマチカゲノムスレッダー(商標)(Inpharmatica Genome Reader)技術を用いて（同時係属国際特許出願PCT/GB01/01105を参照されたい）、LBDG3ポリペプチドと比較したとき、低い配列同一性を有するが、LBDG3ポリペプチド配列と顕著な構造的相同性を共有するがゆえに、核内ホルモンレセプタリガンド結合ドメイン活性をもつと予測される、現在は未知の機能を有するポリペプチドを同定することができる。“顕著な構造的相同性”とは、インファーマチカゲノムスレッダー(商標)が、2つのタンパク質またはタンパク質領域が、少なくとも10%、より好ましくは少なくとも20%、30%、40%、50%、60%、70%、80%、90%および前記を越える確実性で構造的相同性を共有することを予測することを意味する。前記インファーマチカゲノムスレッダー(登録商標)の前記確実性の値は以下のように計算される。既知の構造を有する配列をもっぱら使用して、初めに一組の比較をインファーマチカゲノムスレッダー(登録商標)を用いて行った。いくつかの比較は(構造を基準にして)関連することが判明し

10

20

30

40

50

ているタンパク質間で行った。続いてニューラルネットワークを、CATH構造分類 (www.biochem.ucl.ac.uk-bsm/cath) から得られる既知の関係と既知の非関係とを最も良く識別することを必要とすることに基づいて調整 (train) した。これによって 0 と 1 の間のニューラルネットワークスコアが得られた。しかしながら、一方で関連するタンパク質の数および無関係のタンパク質の数は既知であるので、前記ニューラルネットワークの結果を小群に分配し、正確な結果のパーセンテージを経験的に計算することが可能であった。このようにして、バイオペンジウム検索データベースにおける全ての真正の予測はニューラルネットワークスコアが付随しており、信頼百分率は、インファーマチカゲノムスレッダー (登録商標) がいかに良好なトレーニング / テストセットであるかを反映したものである。

10

## 【0027】

LBDG3の構造的相同体は、LBDG3の核内ホルモンレセプタリガンド結合ドメイン領域と構造的相同性を共有し、“LBDモチーフ” 残基 ASP314、GLN315、LEU318およびLEU319 (LBDG3完全長ポリペプチドでは、ASP551、GLN552、LEU555およびLEU556)、または等価な残基を保有するはずである。そのような構造的相同体は、前記ポリペプチド配列と顕著な構造的相同性を共有し、さらに“LBDモチーフ” 残基を保有するがゆえに、核内ホルモンレセプタリガンド結合ドメイン活性を有すると予測される。

本発明の第一の特徴のポリペプチドにはまた、LBDG3ポリペプチドのフラグメント、LBDG3ポリペプチドのフラグメントの機能的等価物、およびLBDG3ポリペプチドの機能的等価物のフラグメントが含まれるが、ただし前記機能的等価物およびフラグメントは核内ホルモンレセプタリガンド結合ドメイン活性を保持するか、またはLBDG3ポリペプチドまたはLBDG3完全長ポリペプチドと共に抗原決定基を有することを条件とする。

20

## 【0028】

本明細書において用いる、“フラグメント” という用語は、LBDG3ポリペプチド、またはその機能的等価物のアミノ酸配列の一部分 (全部ではなく) と同じアミノ酸配列を有するポリペプチドを指す。フラグメントは、前記配列に由来する少なくとも n 個の連続するアミノ酸を含むべきである。さらに、個々の配列に応じて、n は 7 またはそれより大きい (例えば 8、10、12、14、16、18、20 またはそれより大きい)。小さなフラグメントは抗原決定基を構成することができる。

30

本発明のこの特徴の好ましいポリペプチドフラグメントは、LBDG3ポリペプチドのそれぞれLBDG3核内ホルモンレセプタリガンド結合ドメイン領域と本明細書で明確にされる領域を含むフラグメントである。これらの領域は、核内ホルモンレセプタリガンド結合ドメインと注釈付けされた領域である。

LBDG3ポリペプチドの場合、前記領域は残基 311 と残基 452 の間に広がっていると考えられる (LBDG3完全長ポリペプチドでは、残基 548 - 689)。

前記フラグメントの変種は、本発明のこの特徴の態様として含まれるが、ただしこれら変種は核内ホルモンレセプタリガンド結合ドメインとしての活性を保有することを条件とする。

ある観点では、“変種” という用語は前記ポリペプチドフラグメントの伸長型または短縮型を含むことが意図される。

40

## 【0029】

伸長型変種の場合、LBDG3ポリペプチド配列内のこれら境界の C 末端および / または N 末端にさらに付加された残基が前記ポリペプチドフラグメントに含まれるとき、LBDG3ポリペプチドの核内ホルモンレセプタリガンド結合ドメイン領域が正確に折り畳まれ、核内ホルモンレセプタリガンド結合ドメインの活性を示すであろうということは想像に難くない。例えば、LBDG3ポリペプチド配列、または相同な配列に由来する 5、10、20、30、40 または 50 またはそれより多い付加アミノ酸残基は、LBDG3ポリペプチドの核内ホルモンレセプタリガンド結合ドメイン領域の境界の C 末端および / または N 末端のいずれか一方または両方に含まれ、前記ポリペプチドフラグメントは、正確に折り畳まれる能力を損なうことなく核内ホルモンレセプタリガンド結合ドメイン活性を示し得る。

50

LBDG3ポリペプチドの短縮型変種の場合、1つまたは2つ以上のアミノ酸残基をLBDG3ポリペプチドの核内ホルモンレセプターリガンド結合ドメイン領域のC末端またはN末端の一方または両方で欠失させることができるが、ただし“LBDモチーフ”残基（LBDG3の場合、ASP314、GLN315、LEU318およびLEU319；LBDG3完全長ポリペプチドの場合、ASP551、GLN552、LEU555およびLEU556）または等価な残基は無傷のままで維持され、欠失は、前記残基のいずれかが欠失するほど前記ポリペプチド配列内に深く伸長することはない。

### 【0030】

第二の観点では、“変種”という用語は、LBDG3ポリペプチドまたはLBDG3完全長ポリペプチドの核内ホルモンレセプターリガンド結合ドメイン領域と顕著な配列相同性を保有し、かつ“LBDモチーフ”残基（配列番号：2では、ASP314、GLN315、LEU318およびLEU319）または等価な残基を保有する上記で述べたポリペプチドフラグメントの相同体を含むが、ただし前記変種は核内ホルモンレセプターリガンド結合ドメインとしての活性を保持することを条件とする。

相同体には、LBDG3ポリペプチドのうち、それぞれLBDG3核内ホルモンレセプターリガンド結合ドメイン領域と80%を越える同一性を保有するポリペプチド分子が含まれる。同一性パーセンテージは、BLASTバージョン2.1.3でNCBI(the National Center for Biotechnology Information；<http://www.ncbi.nlm.nih.gov/>）によって特定されるデフォルトパラメーター{Blosum62マトリックス；ギャップ開放ペナルティー=11およびギャップ伸長ペナルティー=1}を用いて決定されるとおりである。好ましくは、本発明のこの特徴のポリペプチドフラグメントの変種相同体は、LBDG3ポリペプチドのLBDG3核内ホルモンレセプターリガンド結合ドメイン領域に関してそれぞれ80%を越える同一性を有する。より好ましくは、変種ポリペプチドは、LBDG3ポリペプチドのLBDG3核内ホルモンレセプターリガンド結合ドメイン領域に関してそれぞれ85%、90%、95%、98%または99%を越える同一性をそれぞれ有するが、ただし前記変種は核内ホルモンレセプターリガンド結合ドメインとしての活性を保持することを条件とする。変種ポリペプチドはまた、上記で考察したポリペプチドフラグメントの短縮型相同体も含むが、ただし前記変種は核内ホルモンレセプターリガンド結合ドメインとしての活性を保持することを条件とする。

### 【0031】

本発明の第一の特徴のポリペプチドフラグメントは、例えばインファーマチカゲノムスレッダー（商標）によって同定される、LBDG3ポリペプチド配列のLBDG3核内ホルモンレセプターリガンド結合ドメイン領域によって規定されるポリペプチドフラグメントの構造と顕著な構造的相同性を示すポリペプチドフラグメントであり得る。したがって、LBDG3ポリペプチド配列のLBDG3核内ホルモンレセプターリガンド結合ドメイン領域によって規定されるポリペプチドフラグメントの構造的相同体であるポリペプチドフラグメントは、上記で定義されたフォールドのような、LBDG3ポリペプチドフラグメントによって採用されるフォールドと同じフォールドを採用するはずである。

LBDG3核内ホルモンレセプターリガンド結合ドメイン領域によって規定されるポリペプチドフラグメントの構造的相同体はまた、“LBDモチーフ”残基ASP314、GLN315、LEU318およびLEU319（LBDG3完全長ポリペプチドではASP551、GLN552、LEU555およびLEU556）、または等価な残基も保持するはずである。

### 【0032】

そのようなフラグメントは、“独立的存在(free-standing)”すなわち、他のアミノ酸もしくはポリペプチドの部分でなくともよく、他のアミノ酸もしくはポリペプチドに融合されていなくてもよく、それらフラグメントがその部分または領域を構成するより大きなポリペプチドの内部に含まれていてもよい。より大きなポリペプチドの内部に含まれている場合は、本発明のフラグメントは最も好ましくは連続するただ1つの領域を形成する。例えばある種の好ましい態様は、前記フラグメントのアミノ末端に融合したプレ-および/またはプロ-ポリペプチド領域、および/または前記フラグメントのカルボキシ末端に融合した付加的領域を有するフラグメントに関する。しかしながら、いくつかのフラグメントが単一のより大きなポリペプチドの内部に含まれてもよい。

10

20

30

40

50

本発明のポリペプチドまたはその免疫原性フラグメント（少なくとも1つの抗原決定基を含む）を用いて、例えばポリクローナルまたはモノクローナル抗体といった、前記ポリペプチドに免疫特異的なリガンドを作製することができる。そのような抗体は、本発明のポリペプチドを発現しているクローニングを単離することもしくは同定すること、またはアフィニティークロマトグラフィーによって前記ポリペプチドを精製することに用いることができる。前記抗体はまた、当業者には明らかなように他の利用の中で特に診断的または治療的補助としても用いられる。

#### 【0033】

“免疫特異的”という用語は、前記抗体が、従来技術において他の関連ポリペプチドに対する親和性よりも本発明のポリペプチドに対して実質的に強い親和性を有することを意味する。本明細書で用いる“抗体”という用語は、完全な分子だけでなく問題の抗原決定基と結合することができるそのフラグメント、例えばFab、F(ab')<sub>2</sub>およびFvも意味する。

したがって、そのような抗体は本発明の第一の特徴のポリペプチドと結合する。

ポリクローナル抗体が所望される場合は、選択される哺乳類（例えばマウス、ウサギ、ヤギまたはウマ）は、本発明の第一の特徴のポリペプチドで免疫することができる。動物を免疫するために用いられるポリペプチドは、リコンビナントDNA技術によって誘導するか、または化学的に合成することができる。所望する場合には、前記ポリペプチドは担体タンパク質と結合させることができる。前記ポリペプチドと化学的に結合させることができる一般的に用いられる担体には、ウシ血清アルブミン、サイログロブリンおよびキーホールリンペットヘモシアニンが含まれる。続いて前記担体結合ポリペプチドを用いて動物を免疫することができる。血清は免疫した動物から採取され、既知の方法（例えばイムノアフィニティークロマトグラフィー）にしたがって処理される。

#### 【0034】

本発明の第一の特徴のポリペプチドに対するモノクローナル抗体も、当業者は容易に生成できる。ハイブリドーマ技術を用いてモノクローナル抗体を作製する一般的な方法論は周知である（例えば以下を参照されたい：G. Kohler & C. Milstein, *Nature* 256 : 495 - 497(1975)； Kozbor et al., *Immunology Today* 4 : 72(1983)； Cole et al., 77 - 96 “*Monoclonal Antibodies and Cancertherapy*”, Alan R. Liss, Inc. (1985)）。

本発明の第一の特徴のポリペプチドに対して生成されたモノクローナル抗体のパネル(panel)を種々の特性、すなわちアイソタイプ、エピトープ、親和性などについてスクリーニングすることができる。モノクローナル抗体は、それらを作らせた個々のポリペプチドの精製に特に有用である。あるいは、対象のモノクローナル抗体をコードする遺伝子を、例えば当技術分野で知られるPCR技術によってハイブリドーマから単離し、さらにクローニングし適切なベクターで発現させることができる。

非ヒト可変領域がヒト定常領域と結合または融合されているキメラ抗体（例えば以下を参照されたい：Proc. Natl. Acad. Sci. USA, 84 : 3439(1987)）もまた有用であろう。

#### 【0035】

抗体は、例えばヒト化によって改変して各個体での免疫原性を減少させることができる（例えば以下を参照されたい：Jones et al., *Nature*, 321 : 522(1986)； Verhoeyen et al., *Science*, 239 : 1534(1988)； Kabat et al., *J. Immunol.*, 147 : 1709(1991)； Queen et al., *Proc. Natl. Acad. Sci. USA*, 86 : 10029(1989)； Gorman et al., *Proc. Natl. Acad. Sci. USA*, 88 : 34181(1991)； Hodgson et al., *Bio/Technology* 9 : 421(1991)）。本明細書で用いられる“ヒト化抗体”という用語は、非ヒトドナー抗体の重鎖および/または軽鎖の可変ドメイン中のCDRアミノ酸および選択した他のアミノ酸がヒト抗体の等価なアミノ酸に代えて置換されている抗体分子を指す。したがって、ヒト化抗体はヒトの抗体と密接に類似するがドナー抗体の結合能力を有する。

また別の選択肢では、前記抗体は、2つの異なる抗原結合ドメインを有し、各ドメインは異なるエピトープに向けられている“二重特異性”抗体であってもよい。

ファージディスプレー技術を用いて、本発明のポリペプチドに対する結合活性をもつ抗体をコードする遺伝子を、関連抗体の保有についてスクリーニングされたヒト由来のリンパ

10

20

30

40

50

球のPCR増幅V-遺伝子のレパートリー、または未感作ライブラリーのいずれかから選択することができる (J. McCafferty et al., (1990) *Nature* 348 : 552 - 554; J. Marks et al., (1992) *Biotechnology* 10 : 779 - 783)。前記抗体の親和性は、鎖のシャッフルイングによって改善することもできる (T. Clackson et al., (1991) *Nature* 352 : 624 - 628)。

上記の技術によって作製された抗体は(ポリクローナルであれモノクローナルであれ)、免疫アッセイ、ラジオイムノアッセイ(RIA)または酵素結合免疫吸着アッセイ(ELIZA)で試薬として用いることができるというさらに別の有用性を有する。前記の利用では、これら抗体は分析的に検出可能な試薬(例えば放射性同位元素、蛍光分子または酵素)で標識することができる。

10

### 【0036】

本発明の第二および第三の特徴の好ましい核酸分子は、配列番号：2または配列番号：4に記載のポリペプチド配列および機能的に等価なポリペプチドをコードするものである。前記機能的に等価なポリペプチドには、LBDG3ポリペプチドの活性なフラグメント、例えばLBDG3ポリペプチド配列のLBDG3核内ホルモンレセプタリガンド結合ドメイン領域を含むフラグメントフラグメント、またはその相同体が含まれる。

これら配列の一続きを包含する核酸分子は本発明のこの特徴の好ましい態様を形成する。これらの核酸分子は、本明細書に記載する方法および応用で用いることができる。好ましくは本発明の核酸分子は、本明細書に開示する配列に由来する少なくともn個の連続するヌクレオチドを含む。nは、個々の配列に応じて10またはそれより大きい(例えば12、14、15、18、20、25、30、35、40またはそれより大きい)。

20

本発明の核酸分子は、上記で述べた核酸分子に相補的な配列も含む(例えばアンチセンスまたはプローブとしての目的のために)。

本発明の核酸分子は、RNA(例えばmRNA)、またはDNA(例えばcDNA、合成DNAまたはゲノムDNAを含む)の形態をとることができる。そのような核酸分子は、クローニングによって、化学合成によって、またはそれらを併用して得ることができる。前記核酸分子は、固相ホスホルアミダイト化学合成のような技術を用いるゲノムまたはcDNAライブラリーからの化学合成によって、または生物体から分離することによって調製することができる。RNA分子は一般的にはDNA配列のin vitroまたはin vivo転写によって作製することができる。

30

### 【0037】

核酸分子は二本鎖でも一本鎖でもよい。一本鎖DNAはコード鎖(センス鎖としても知られる)でも、非コード鎖(アンチセンス鎖とも称される)でもよい。

“核酸分子”という用語には、DNAおよびRNAのアナログ(例えば改变骨格を含むもの)、並びにペプチド核酸(PNA)も含まれる。本明細書で用いられる“PNA”という用語はアンチセンス分子または抗遺伝子(anti-gene)作用因子を指し、長さが少なくとも5ヌクレオチドであってアミノ酸残基のペプチド骨格と結合したオリゴヌクレオチドを含む。前記ペプチド骨格は好ましくはリジンで終わり、前記末端リジンは当該組成物に可溶性を付与する。PNAはPEG化(PEGylated)されて細胞内での寿命が延長されてもよい(細胞内では、PNAは優先的に相補性一本鎖DNAおよびRNAと結合して転写物伸長を停止させる(P.E. Nielsen et al. (1993) *Anticancer Drug Des.* 8 : 53 - 63) )。

40

配列番号：2または配列番号：4のポリペプチドまたはその活性フラグメントをコードする核酸分子は、配列番号：1または配列番号：3にそれぞれ示した核酸分子のコード配列と同一でもよい。これらの分子は、遺伝コードの縮退の結果として、配列番号：2または配列番号：4のポリペプチドまたはLBDG3ポリペプチドの活性フラグメント(例えばLBDG3核内ホルモンレセプタリガンド結合ドメイン領域を含むフラグメント)、またはその相同体をコードする多様な配列を有することもできる。LBDG3核内ホルモンレセプタリガンド結合ドメイン領域は、LBDG3ポリペプチド配列の残基311と残基452の間に広がっていると考えられる。LBDG3完全長ポリペプチドでは、当該境界は残基548および689である。したがって配列番号：1では、LBDG3核内ホルモンレセプタリガンド結合ドメイン領域

50

は、ヌクレオチド932から1357を含む核酸分子によってコードされる。配列番号：3では、そのような境界は1642および2067である。前記配列の一続きを包含する核酸分子、および前記配列の相同体は、本発明のこの特徴の好ましい態様を形成する。

【0038】

配列番号：2または配列番号：4のポリペプチドをコードする前記核酸分子には、それ自身で成熟なポリペプチドのコード配列；成熟ポリペプチドおよび付加コード配列（例えばリーダー配列または分泌配列、例えばプロ-、プリ-またはプレプロ-ポリペプチド配列をコードするもの）のためのコード配列；前述の付加的コード配列を伴なう、または伴なわないが、さらに付加的な非コード配列（非コード5'および3'配列を含む）を伴なう成熟ポリペプチドのコード配列が含まれるが、ただしこれらに限定されない。前記の非コード5'および3'配列は、例えば転写される非翻訳配列で、転写（終止シグナルを含む）、リボソーム結合およびmRNA安定性で役割を果たすものである。前記核酸分子は、更なる官能性を提供するアミノ酸のような付加アミノ酸をコードする付加配列を含むこともできる。

本発明の第二および第三の特徴の核酸分子は、本発明の第一の特徴のポリペプチドのフラグメントまたは機能的等価物およびそのフラグメントもコードし得る。

上記で考察したように、LBDG3ポリペプチドの好ましいフラグメントは、LBDG3核内ホルモンレセプタリガンド結合ドメイン領域を含むフラグメント、またはその相同体である。前記核内ホルモンレセプタリガンド結合ドメイン領域は、配列番号：1のヌクレオチド932から1357を含む核酸分子によってコードされる（配列番号：3では、そのような境界は1642および2067である）。

本発明の機能的に等価な核酸分子は、天然に存在する変種（例えば天然に存在する対立形質変種）であっても、または前記分子は天然に存在することが知られていない変種であってもよい。前記のような天然に存在しない核酸分子の変種は、突然変異誘発技術（核酸分子、細胞または生物に対して適用される技術が含まれる）によって達成できる。

このような変種の中では、特にヌクレオチドの置換、欠失または挿入によって前述の核酸分子と異なる変種が挙げられる。置換、欠失または挿入は1つまたは2つ以上のヌクレオチドを含むことができる。変種はコード領域または非コード領域またはその両方が変化してもよい。コード領域における変化は、保存的または非保存的なアミノ酸置換、欠失または挿入をもたらし得る。

【0039】

本発明の核酸分子はまた、多様な理由で、遺伝子生成物（ポリペプチド）のクローニング、プロセッシングおよび/または発現の改変を含む当技術分野で一般的に知られている方法を用いて操作され得る。ランダムフラグメント化によるDNAシャッフリングおよび遺伝子フラグメントおよび合成オリゴヌクレオチドのPCRリアッセンブリーは、ヌクレオチド配列の操作に用いられ得る技術に含まれる。位置特異的突然変異誘発を用いて、新規な制限部位の挿入、グリコシリ化パターンの変更、コドンの優先性の変化、スプライシング変種の生成、変異の導入など、その他を行うことができる。

本発明の第一の特徴のポリペプチドをコードする核酸分子は異種配列に連結され、それによって結合核酸分子が融合タンパク質をコードすることができるようにしてよい。前記のような結合核酸分子は本発明の第二または第三の特徴に包含される。例えば、本発明のポリペプチドの阻害物質のためのペプチドライブライアリーやスクリーニングするために、前記のような結合核酸分子を用いて、市販の抗体によって認識される融合タンパク質を発現させることは有用であろう。融合タンパク質はまた、本発明のポリペプチド配列と異種タンパク質配列との間に位置する切断部位を含むように操作し、それによって前記ポリペプチドを異種タンパク質から切り離して精製することができるようにしてよい。

【0040】

本発明の核酸分子にはまた本発明のポリペプチドをコードする核酸分子と部分的に相補的であり、したがってコード核酸分子とハイブリダイズする（ハイブリダイゼーション）アンチセンス分子が含まれる。そのようなアンチセンス分子（例えばオリゴヌクレオチド）

10

20

30

40

50

は、当業者にはよく知られるように、本発明のポリペプチドをコードする標的核酸を認識し、その標的核酸と特異的に結合してその転写を妨げるようデザインすることができる（例えば以下の文献を参照されたい：J.S. Cohen, Trends in Pharm. Sci., 10: 435(1989); J. Okano, Neurochem. 56: 560(1991); J. O' Connor, Neurochem. 56: 560(1991); Lee et al., Nucleic Acids Res. 6: 3073(1979); Cooney et al., Science 241: 456(1988); Dervan et al., Science 251: 1360(1991)）。

本明細書で用いられる“ハイブリダイゼーション”という用語は、2つの核酸分子が水素結合によって互いに結合することを指す。典型的には、1つの分子が固相支持体に固定され、他方は溶液中で遊離しているであろう。続いて2つの分子を水素結合に適した条件下で互いに接触させる。前記結合に影響する因子には以下が含まれる：溶媒の種類および体積；反応温度；ハイブリダイゼーションの時間；攪拌；液相分子の固相支持体への非特異的結合を妨害する薬剤（デンハルト試薬、またはBLOTT）；分子の濃度；分子の結合速度を増加させる化合物の使用（硫酸デキストランまたはポリエチレングリコール）；およびハイブリダイゼーションに続く洗浄条件のストリンジエンシー（Sambrook et al.（上掲書）を参照されたい）。

#### 【0041】

完全に相補的な分子と標的分子とのハイブリダイゼーションの阻害は、当業者に知られるハイブリダイゼーションアッセイを用いて調べることができる（例えばSambrook et al.（上掲書）を参照されたい）。したがって、実質的に相同な分子は、文献（G.M. Wahl and S.L. Berger, 1987, Methods Enzymol. 152: 399 - 407; A.R. Kimmel, 1987, Methods Enzymol. 152: 507 - 511）に開示されたように完全に相同な分子と標的分子との結合を種々のストリンジエンシー条件下で競合させ阻害するであろう。

“ストリンジエンシー”とは、異なる分子の結合よりも非常に類似した分子の結合に適したハイブリダイゼーション反応の条件を指す。高ストリンジエンシーのハイブリダイゼーション条件は、以下を含む溶液（50%のホルムアミド、5倍のSSC（150mM NaCl、15mMクエン酸三ナトリウム）、50mMリン酸ナトリウム（pH 7.6）、5倍のデンハルト溶液、10%の硫酸デキストラン、および20 μg / mLの変性せん断サケ精子DNA）中で42度一晩インキュベーションし、続いてフィルターを0.1倍のSSCで約65度洗浄すると定義される。低ストリンジエンシー条件は、ハイブリダイゼーション反応が35度で実施されることを含む（Sambrook et al.（上掲書）を参照されたい）。好ましくは、ハイブリダイゼーションに用いられる条件は高ストリンジエンシーを構成するものである。

#### 【0042】

本発明のこの特徴の好ましい態様は、LBDG3ポリペプチド（配列番号：2）またはLBDG3完全長ポリペプチド（配列番号：4）をコードする核酸分子の全長にわたって少なくとも80%同一である核酸分子、および前記のような核酸分子に対して実質的に相補的な核酸分子である。好ましい活性フラグメントは、LBDG3ポリペプチド配列のそれぞれLBDG3核内ホルモンレセプタリガンド結合ドメイン領域を含むフラグメントである。したがって、好ましい核酸分子は、LBDG3ポリペプチド配列の核内ホルモンレセプタリガンド結合ドメイン領域をコードする核酸分子の全長にわたって少なくとも80%同一であるものを含む。

本明細書で言及される同一性パーセンテージは、BLASTバージョン2.1.3でNCBI（the National Center for Biotechnology Information；<http://www.ncbi.nlm.nih.gov/>）によって特定されたデフォルトパラメーターを用いて決定されるとおりである。

好ましくは、本発明のこの特徴の核酸分子は、配列番号：1に示された配列を有する核酸分子、この配列のヌクレオチド932 - 1357を含む領域の全長にわたって少なくとも80%同一の領域を含む。本発明のこの特徴の他の好ましい核酸分子は、配列番号：1に示された配列を有する核酸分子、この配列のヌクレオチド2913 - 3515を含む領域の全長にわたって少なくとも80%同一の領域を含むか、またはこれら核酸領域のいずれかと相補的である核酸分子を含む。この場合、前記の全長にわたって少なくとも90%、好ましくは少なくとも95%、より好ましくは少なくとも98%または99%同一の核酸分子が特に好ま

10

20

30

40

50

しい。これに関して、好ましい態様は、LBDG3ポリペプチドと実質的に同じ生物学的機能または活性を保持するポリペプチドをコードする核酸分子である。

#### 【0043】

本発明はまた、以下の工程を含む、本発明の核酸分子を検出する方法を提供する：(a)二重鎖を形成するハイブリダイゼーション条件下で本発明の核酸プローブを生物学的サンプルと接触させる工程；および(b)形成された前記の全ての二重鎖を検出する工程。本発明にしたがって利用することができるアッセイに関する下記でさらに考察するよう 10 に、上記で述べた核酸分子をRNA、cDNAまたはゲノムDNAのためのハイブリダイゼーションプローブとして用い、LBDG3ポリペプチドまたはLBDG3完全長ポリペプチドをコードする完全長のcDNAおよびゲノムクローンを単離し、さらに前記ポリペプチドをコードする遺伝子と高い配列類似性を有する相同遺伝子またはオーソログ遺伝子のcDNAまたはゲノムクローンを単離することができる。

これに関しては、当技術分野で公知の他の技術の中で特に以下の技術を利用することができる。これらの技術は例示として下記で考察される。DNAのシークエンシングおよび解析のための方法は周知で、当技術分野では一般的に利用可能であり、本明細書で考察される本発明の態様の多くを実施するために実際に用いることができる。そのような方法では、DNAポリメラーゼIのKlenowフラグメント、シークエンサー(US Biochemical Corp., Cleveland, OH)、Taqポリメラーゼ(Perkin Elmer)、耐熱性T7ポリメラーゼ(Amersham, Chicago, IL)、またはポリメラーゼと校正エキソヌクレアーゼの組合せ(例えば市販(Gibco/BRL, Gaithersburg, MD)のELONGASE增幅キットで見出されるようなもの)のような酵素を利用することができる。好ましくは、シークエンシングプロセスは、例えばハミルトンマイクロラブ(Hamilton Micro Lab)2200(Hamilton, Reno, NV)、ペルティエサーマルサイクラー(Peltierthermal Cycler)PTC200(MJ Research, Watertown, MA)、A B Iカタリスト並びに373および377DNAシークエンサー(Perkin Elmer)のような機器を用いて自動化することができる。

#### 【0044】

LBDG3ポリペプチドの機能と等価な機能を有するポリペプチド、特にLBDG3ポリペプチドのLBDG3核内ホルモンレセプターリガンド結合ドメイン領域と同等の機能をもつポリペプチドをコードする核酸分子を単離する方法の1つは、当技術分野で知られている標準的な方法を用いる、天然のプローブまたは人工的にデザインしたプローブによるゲノムライブラリーまたはcDNAライブラリーの探索である(例えば以下の文献を参照されたい：“Current Protocols in Molecular Biology”, Ausubel et al.(eds). Greene Publishing Association and John Wiley Interscience, New York, 1989, 1992)。特に有用なプローブは、適切なコード遺伝子(配列番号：1)、特に配列番号：1のヌクレオチド932-1357の領域に由来する核酸配列に対応する、または前記配列と相補的である、少なくとも15、好ましくは少なくとも30、さらに好ましくは少なくとも50の連続塩基を含むプローブである。

前記のようなプローブは分析的な検出が可能な試薬で標識して前記プローブの識別を容易にすることができる。有用な試薬には、放射性同位元素、蛍光色素、および検出可能な生成物の形成を触媒し得る酵素が含まれるが、ただしこれらに限定されない。これらのプローブを用いて、当業者は、ヒト、哺乳類または他の動物供給源から対象のタンパク質をコードするゲノムDNA、cDNAまたはRNAポリヌクレオチドの相補的なコピーを単離し、近縁配列、例えば前記のファミリー、タイプおよび/またはサブタイプに属するまた別のメンバーについて、前記の供給源をスクリーニングすることができるであろう。

#### 【0045】

多くの場合、単離されるcDNA配列は不完全で、ポリペプチドをコードする領域は短く(通常は5'末端で)切断されているであろう。完全長cDNAを得るために、または短いcDNAを伸長させるために、いくつかの方法が利用可能である。そのような配列は、部分的なヌクレオチド配列を用い、上流の配列(例えばプロモーターおよび調節エレメント)を検出するための当技術分野で公知の種々の方法を用いて伸長させることができる。例えば、使用

10

20

30

40

50

され得るある方法は、cDNA末端迅速増幅法（RACE；例えば以下を参照されたい：Frohman et al., Proc. Natl. Acad. Sci. USA(1988) 85: 8998 - 9002）に基づく。前記技術の最近の改変（例えばマラソン（Marathon）（商標）技術（Clontech Laboratories Inc.）により例示される）は、より長いcDNAの検索を顕著に単純化した。わずかに異なる技術（“制限部位”PCRと称される）では、普遍的プライマーを用いて既知の遺伝子座に近接する未知の核酸配列が検索される（G. Sarkar (1993) PCR Methods Applic. 2: 318 - 322）。逆PCRもまた、既知の領域に基づく多様なプライマーを用いた配列の増幅または伸長に用いることができる（T. Triglia et al. (1988) Nucleic Acids Res. 16: 8186）。使用できる別の方法は捕捉PCRで、前記は、ヒトおよび酵母の人工染色体DNAで既知配列に近接するDNAフラグメントのPCR増幅を含む（M. Lagerstrom et al. (1991) PCR Methods Applic. 1: 111 - 119）。未知の配列を検索するために利用可能な別の方法はパークーの方法である（J.D. Parker et al. (1991) Nucleic Acids Res. 19: 3055 - 3060）。さらに、ゲノムDNAを歩行させるためにPCR、入れ子（nested）プライマーおよびプロモーターファインダー（PromoterFinder）（商標）ライブラリー（Clontech, Palo Alto, CA）を用いてもよい。この方法ではライブラリーのスクリーニングが不要で、イントロン／エクソン結合部を見出すことに有用である。

#### 【0046】

完全長cDNAをスクリーニングする場合、より大きなcDNAを包含するためにサイズ選択を実施したライブラリーを用いることが好ましい。さらにまた、遺伝子の5'領域を含む配列をより多く含むという点でランダムプライミングした（random-primed）ライブラリーが好ましい。ランダムプライミングしたライブラリーの使用は、オリゴd（T）ライブラリーが完全長cDNAを生成できない状況で特に好ましいであろう。ゲノムライブラリーは、5'非転写調節領域に配列を伸長させるために有用であろう。

本発明のある態様では、染色体局在化のために本発明の核酸分子を用いることができる。この技術では、核酸分子は個々のヒト染色体上の特定の位置に対して特異的に標的化され、個々のヒト染色体上の特定の位置にハイブリダイズさせることができる。本発明の関連配列の染色体上へのマッピングは、遺伝子関連疾患に関する配列の相関性確認において重要な工程である。いったん染色体の正確な位置に配列がマッピングされたら、前記配列の染色体上の物理的な位置を遺伝子地図データと相關させることができる。そのようなデータは、例えば以下で見出すことができる：V. McKusick, Mendelian Inheritance in Man（ジョーンズホプキンス大学、ウェルチ医学図書館を通じてオンラインで入手可能である）。同じ染色体領域にマッピングされた遺伝子と疾患との関係を、次に連鎖解析（物理的に近接する遺伝子の同時遺伝（coinheritance））によって同定する。これにより、ポジショナルクローニングまたは他の遺伝子発見技術を用いて疾患遺伝子を検索する研究者に貴重な情報が提供される。いったん疾患または症候群の位置が遺伝連鎖によって特定のゲノム領域で大まかに限局されたら、前記領域にマッピングされるいすれの配列も、更なる解析のための関連または調節遺伝子となることができる。前記核酸分子はまた、正常な個体、キャリア個体または罹患個体間で転座、逆位などによる染色体位置上の相違を検出するために用いることができる。

#### 【0047】

本発明の核酸分子はまた組織分布同定（tissue localisation）のために貴重である。そのような技術は、ポリペプチドをコードするmRNAの検出によって組織中の前記ポリペプチドの発現パターンの決定を可能にする。これらの技術にはin situハイブリダイゼーション技術およびヌクレオチド増幅技術（例えばPCR）が含まれる。これらの研究から得られる結果は、生物内での前記ポリペプチドの正常な機能を示唆する。さらに、変異遺伝子によってコードされるmRNAの発現パターンと正常mRNA発現パターンとの比較研究によって、変異ポリペプチドの疾患における役割に対する貴重な洞察が提供される。そのような不適切な発現は時間的、位置的または量的性質を有する場合もある。

本発明のベクターは本発明の核酸分子を含み、クローニングベクターでも発現ベクターでもよい。本発明のベクターで形質転換、トランスフェクトまたは形質導入され得る本発明

10

20

30

40

50

の宿主細胞は、原核細胞でも真核細胞でもよい。

本発明のポリペプチドは、宿主細胞内に含まれるベクター中の前記ポリペプチドをコードする核酸分子の発現によってリコンビナント形態で調製することができる。前記のような発現方法は当業者によく知られており、以下の文献により詳細に記述され得る：Sambrook et al. (上掲書) およびFeRNAAndez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression", Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto)。

#### 【0048】

一般的には、要求される宿主でポリペプチドを生成させるために核酸分子の維持、増殖または発現に適したいずれの系またはベクターも用いることができる。周知であり日常的である種々の技術のいずれによっても（例えば前掲書（Sambrook et al.）に記載されたようなもの）、適切なヌクレオチド配列を発現系に挿入することができる。一般的には、コード遺伝子は制御エレメント（例えばプロモーター、リボソーム結合部位（細菌での発現の場合）、および場合によってオペレーター）の制御下に置かれ、それによって所望のポリペプチドをコードするDNA配列を形質転換宿主細胞でRNAに転写させることができる。

適切な発現系の例には、例えば染色体系、エピソーム系およびウイルス由来系、例えば以下に由来するベクターが含まれる：細菌プラスミド、バクテリオファージ、トランスポゾン、酵母エピソーム、挿入エレメント、酵母染色体エレメント、ウイルス、例えばバキュロウイルス、パポーバウイルス（例えばSV40）、ワクシニアウイルス、アデノウイルス、鶏痘ウイルス、仮性狂犬病ウイルスおよびレトロウイルス、または上記の組合せ、例えばプラスミドとバクテリオファージの遺伝子エレメントに由来するもの（例えばコスミドおよびファージミドを含む）。ヒト人工染色体（HAC）もまた、プラスミドに包含させ発現させるには大きいDNAフラグメントを搬送するために用いることができる。

#### 【0049】

特に適切な発現系には、リコンビナントバクテリオファージ、プラスミドまたはコスミドDNA発現ベクターで形質転換された微生物（例えば細菌）；酵母発現ベクターで形質転換された酵母；ウイルス発現ベクター（例えばバキュロウイルス）を感染させた昆虫細胞系；ウイルス発現ベクター（例えばカリフラワーモザイクウイルス、CaMV；タバコモザイクウイルス、TMV）または細菌発現ベクター（例えばTiまたはpBR322プラスミド）で形質転換した植物細胞系；または動物細胞系が含まれる。無細胞翻訳系もまた本発明のポリペプチドの生成に用いることができる。

本発明のポリペプチドをコードする核酸分子の宿主細胞への導入は、多くの標準的な実験室マニュアル（例えば、Davis et al., Basic Methods in Molecular Biology (1986) および上掲書（Sambrook et al.））に記載された方法によって達成できる。特に適切な方法には、リン酸カルシウムトランスフェクション、DEAEデキストラン仲介トランスフェクション、トランスペクション(transvection)、マイクロインジェクション、陽イオン脂質仲介トランスフェクション、エレクトロポレーション、トランスタクション、擦過ローディング（scrape loading）、弾道導入または感染が含まれる（以下を参照されたい：Sambrook et al. (1989) 上掲書；Ausbel et al. (1991) 上掲書；Spector, Goldman &leinward, (1998)）。

#### 【0050】

コード核酸分子は、所望であれば、例えばシグナルペプチドまたはリーダー配列のような制御配列をコードする配列を（例えば翻訳ポリペプチドの小胞体内、細胞周辺腔または細胞外環境への分泌のために）含んでいても含んでいなくてもよい。このシグナルは前記ポリペプチドにとって内因性であっても異種シグナルであってもよい。リーダー配列は、翻訳後プロセッシングで細菌宿主によって取り除くことができる。

コントロール配列の他に、宿主細胞の増殖に関連して前記ポリペプチドの発現の調節を可能にする調節配列を付加することが望ましい場合がある。調節配列の例は、化学的または物理的刺激（調節化合物の存在を含む）または多様な温度もしくは代謝条件に応答して遺伝子の発現を増加させたり低下させたりする配列である。調節配列は、ベクターの非翻訳

10

20

30

40

50

領域、例えばエンハンサー、プロモーター並びに5'および3'非翻訳領域である。これらは、宿主細胞タンパク質と相互作用して、転写および翻訳を実行する。そのような調節配列は、その強さおよび特異性を変化させることができる。用いられるベクター系および宿主に依存して、多くの適切な転写および翻訳エレメント（構成性および誘発性プロモーターを含む）を用いることができる。例えば、細菌系でクローニングするときは、誘発性プロモーター、例えばBluescriptファージミド（Stratagene, La Jolla, CA）またはpSprotl（商標）プラスミド（Gibco BRL）などのハイブリッドlacZプロモーターを用いることができる。バキュロウイルスピリヘドリン(polyhedrin)プロモーターは昆虫細胞で用いることができる。植物細胞ゲノムに由来するプロモーターまたはエンハンサー（例えば熱ショック、RUBISCOおよび貯蔵タンパク質遺伝子）または植物ウイルスに由来するプロモーターまたはエンハンサー（例えばウイルスプロモーターまたはリーダー配列）は、ベクターへクローニングすることができる。哺乳類細胞系では、哺乳類遺伝子由来または哺乳類ウイルス由来のプロモーターが好ましい。配列の多数コピーを含む細胞株の作製が必要な場合、SV40またはEBVをベースにしたベクターを適切な選択マーカーとともに用いることができる。10

#### 【0051】

発現ベクターは、特定の核酸コード配列を適切な調節配列とともにベクター内に配置させることができるように構築される。前記コード配列の調節配列に関する位置および向きは、前記コード配列が調節配列の“制御下”で転写されるような位置および向きである（すなわちコントロール配列にてDNA分子と結合するRNAポリメラーゼは前記コード配列を転写する）。いくつかの事例では、前記配列を適切な向きで制御配列に付属させることができるように（すなわちリーディングフレームを維持するために）、前記配列を改変する必要があるであろう。20

コントロール配列および他の調節配列は、ベクターへの挿入の前に核酸コード配列に連結させることができる。あるいは、コード配列は、コントロール配列および適切な制限部位を既に含む発現ベクターへ直接クローニングすることができる。

長期的なリコンビナントポリペプチドの高収量の生成のためには、安定な発現が好ましい。例えば、対象のポリペプチドを安定に発現する細胞株は、ウイルスの複製起点および/または内因性発現エレメント並びに選択マーカー（同じベクターまたは別個のベクターに存在する）を含む発現ベクターを用いて形質転換させることができる。ベクターの導入に続き、選択培地に切り替える前に細胞を栄養(enriched)培地で1-2日間増殖させることができる。選択マーカーの目的は、選択に対する耐性を付与することで、選択マーカーの存在によって、導入された配列をうまく発現する細胞の増殖および回収が可能になる。安定に形質転換された細胞の耐性クローンは、細胞の種類に適した組織培養技術を用いて増殖させることができる。30

#### 【0052】

発現のための宿主として利用可能な哺乳類細胞株は当技術分野で公知であり、米国菌培養収集所（American Type Culture Collection, ATCC）から入手可能な多くの不朽化細胞株が含まれる。前記には、例えばチャイニーズハムスター卵巣細胞(CHO)、HeLa、ベイビーハムスター腎(BHK)、サル腎(COS)、C127、3T3、BHK、HEK293、ボウズ(Bowes)メラノーマおよびヒト肝細胞癌（例えばHepG2）細胞および多数の細胞株が含まれるが、これだけに限られない。40

バキュロウイルス系では、バキュロウイルス/昆虫細胞発現系のための材料は、特にインビトロジエン（Invitrogen, San Diego, CA）からキットの形態で（“MaxBac”キット）商業的に入手可能である。前記の技術は一般的に当業者に知られており、文献には完全に記載されている（Summers & Smith, Texas Agricultural Experiment Station Bulletin No.1555(1987)）。この系での使用に特に適切な宿主細胞には昆虫細胞、例えばドロソフィラ(Drosophila) S2およびスピドプテラ(Spodoptera) Sf9細胞が含まれる。

当技術分野で公知である多くの植物細胞培養および植物体(whole plant)遺伝子発現系が存在する。適切な植物細胞遺伝子発現系の例には米国特許第5,693,506号、5,659,122号お50

より5,608,143号に記載されたものが含まれる。植物細胞培養における遺伝子発現の別の例は、文献に記載されている (Zenk (1991) *Phytochemistry* 30 : 3861 - 3863)。

特に、プロトプラストを単離し、これを培養して完全な再生植物を形成することが可能な植物は全て利用することができ、それによって移入遺伝子を含む完全な植物が回収される。特に、サトウキビ、サトウダイコン、綿花、果実および他の樹木、マメ類および野菜の主要な種の全てを含む(ただしこれらに限定されない)全ての植物は、培養細胞または組織から再生させることができる。

#### 【0053】

特に好ましい細菌宿主細胞の例には連鎖球菌、ブドウ球菌、大腸菌 (*E. coli*)、ストレプトマイセスおよびバチルス・ズブチリス (*Bacillus subtilis*) 細胞が含まれる。

10

真菌での発現に特に適切な宿主細胞の例には酵母細胞(例えば *S. cerevisiae*)およびアスペルギルス細胞が含まれる。

形質転換細胞株の回収に用いることができる多くの選択系は、当技術分野で公知である。例えば、単純ヘルペスウイルスチミジンキナーゼ (M. Wigler et al. (1977) *Cell* 11 : 22 3 - 32) およびアデニンホスホリボシルトランスフェラーゼ (I. Lowy et al. (1980) *Cell* 22 : 817 - 23) の遺伝子が含まれ、前記はそれぞれ tk- または *aprt* ± 細胞で用いることができる。

さらにまた、抗代謝物質耐性、抗生物質耐性または除草剤耐性を選択基準として用いてよい。例えばジヒドロ葉酸レダクターゼ (DHFR) はメトトレキセートに対する耐性を付与し (M. Wigler et al. (1980) *Proc. Natl. Acad. Sci.* 77 : 3567 - 70)、*npt* はアミノグリコシド系ネオマイシンおよび G-418 に耐性を付与し (F. Colbere-Garapin et al. (1981) *J. Mol. Biol.* 150 : 1 - 14)、さらに *als* または *pat* はそれぞれクロロスルフロン (chlorsulfuron) およびホスフィノトリシン (phosphinotricin) アセチルトランスフェラーゼに対する耐性を付与する。さらに別の選択可能な遺伝子が報告されており、当業者にはそれらの例は明白であろう。

20

#### 【0054】

マーカー遺伝子の発現の有無は対象の遺伝子も存在することを示唆するが、対象の遺伝子の存在および発現を確認する必要があり得る。例えば、関連する配列がマーカー遺伝子配列内に挿入されている場合、適切な配列を含む形質転換細胞はマーカー遺伝子機能が存在しないことによって識別することができる。あるいは、マーカー遺伝子は、ただ 1 つのプロモーターの制御下で本発明のポリペプチドをコードする配列とともに直列に配置することができる。通常、誘発または選択に応答するマーカー遺伝子の発現は、直列遺伝子の発現も示している。

30

あるいは、本発明のポリペプチドをコードする核酸配列を含み、前記ポリペプチドを発現する宿主細胞は、当業者に知られている多様な方法で同定することができる。前記方法には、DNA - DNA または DNA - RNA ハイブリダイゼーションおよびタンパク質バイオアッセイ、例えば蛍光活性化細胞ソーティング (FACS) またはイムノアッセイ技術(例えば酵素結合免疫吸着アッセイ (ELISA) および放射性イムノアッセイ (RIA)) が含まれ(ただしこれらに限定されない)、核酸またはタンパク質の検出および/または定量のためのメンブレン、溶液またはチップをベースとする技術が含まれる(例えば以下を参照されたい: R. H ampton et al. (1990) *Serological Methods, A Laboratory Manual*, APS Press, St Paul, MN; および D.E. Maddox et al. (1983) *J. Exp. Med.* 158 : 1211 - 1216)。

40

#### 【0055】

多様な標識および結合技術が当業者に知られており、種々の核酸およびアミノ酸アッセイで用いることができる。本発明のポリペプチドをコードする核酸分子に関連する配列を検出するための標識ハイブリダイゼーションプローブまたは PCR プローブの作製手段には、標識したポリヌクレオチドを用いるオリゴ標識、ニックトランスレーション、末端標識または PCR 増幅が含まれる。あるいは、本発明のポリペプチドをコードする配列をベクターにクローニングして mRNA プローブを作製することができる。そのようなベクターは当技術分野で公知で、商業的に入手可能であり、適切な RNA ポリメラーゼ(例えば T7、T3 または S

50

P6) および標識ヌクレオチドを添加することにより *in vitro* で RNA プローブを合成することに用いることができる。これらの方法は、種々の市販キット (Pharmacia & Upjohn (Kalamazoo, MI); Promega (Madison, WI); U.S. Biochemical Corp., (Cleaveland, OH)) を用いて実施することができる。

適切なレポーター分子または標識 (前記は検出を容易にするために用いることができる) には、放射性核種、酵素および蛍光、化学発光または色素生産性物質の他に基質、コファクター、阻害剤、磁性粒子などが含まれる。

#### 【0056】

本発明の核酸分子はまたトランスジェニック動物 (特にげっ歯類動物) の作製に用いることができる。そのようなトランスジェニック動物は本発明の別の特徴を構成する。これは、体細胞の改変によって局部的に、または遺伝性改変を導入する生殖細胞系療法によって実施することができる。前記のようなトランスジェニック動物は、本発明のポリペプチドのモジュレーターとして有効な薬剤分子のための動物モデルを作製するために特に有用であろう。

ポリペプチドは、周知の方法によってリコンビナント細胞培養物から回収し精製することができる。前記周知の方法には、硫安またはエタノール沈澱、酸性抽出、陰イオンまたは陽イオン交換クロマトグラフィー、リン酸化セルロースクロマトグラフィー、疎水性相互作用クロマトグラフィー、アフィニティークロマトグラフィー、ヒドロキシルアパタイトクロマトグラフィーおよびレクチンクロマトグラフィーが含まれる。高性能液体クロマトグラフィーは、精製に特に有用である。単離および精製の間にポリペプチドが変性した場合には、タンパク質のリフォールディングのためによく知られている技術を用いて活性な構造を再生することができる。

#### 【0057】

所望の場合には、可溶性タンパク質の精製を容易にするポリペプチドドメインをコードするヌクレオチド配列に本発明のポリペプチドをコードする配列を連結させることにより特殊化したベクター構築物も、タンパク質の精製を容易にするために用いることができる。そのような精製促進ドメインの例には、金属キレートペプチド (例えば固定化金属上での精製を可能にするヒスチジン - トリプトファンモジュール、固定化免疫グロブリン上での精製を可能にするプロテイン A ドメイン、および FLAGS 伸長 / アフィニティー精製システム (Immunex Corp., Seattle, WA) で用いられるドメイン) が含まれる。切断可能なリンカーパ配列 (例えば X A 因子またはエンテロキナーゼ (Invitrogen, San Diego, CA) に特異的なもの) を精製ドメインと本発明のポリペプチドとの間に包含させて、精製を容易にすることに用いてもよい。そのような発現ベクターの1つは、チオレドキシンまたはエンテロキナーゼ切断部位に先行するいくつかのヒスチジン残基と融合させた本発明のポリペプチドを含む融合タンパク質の発現を提供する。ヒスチジン残基は、IMAC (固定金属イオンアフィニティークロマトグラフィー; J. Porath et al. (1992) *Prot. Exp. Purif.* 3: 263 - 281) により精製を容易にし、一方、チオレドキシンまたはエンテロキナーゼ切断部位は融合タンパク質からポリペプチドを精製するための手段を提供する。融合タンパク質を含むベクターについての考察は以下で提供される: D.J. Kroll et al. (1993) *DNA Cell Biol.* 12: 441 - 453)。

#### 【0058】

スクリーニングアッセイで使用するためにポリペプチドを発現させる場合は、前記ポリペプチドを発現する宿主細胞の表面で前記ポリペプチドを生成させることができが一般には好ましい。この場合、宿主細胞はスクリーニングアッセイで使用する前に、例えば蛍光活性化細胞ソーティング (FACS) またはイムノアフィニティー技術のような技術を用いて収穫することができる。ポリペプチドが培養液中に分泌される場合は、前記培養液を回収して発現されたポリペプチドを回収および精製することができる。ポリペプチドが細胞内で生成される場合、ポリペプチドを回収する前に、先ず初めに細胞を溶解させねばならない。

本発明のポリペプチドを用いて、種々の薬剤スクリーニング技術のいずれかで化合物ライブラリーをスクリーニングすることができる。そのような化合物は、本発明のポリペプチド

10

20

30

40

50

ドの遺伝子発現レベルまたは活性レベルを活性化(作動)させ、または阻害(拮抗)することができる。それらは本発明のさらなる特徴を形成する。好ましい化合物は、本発明の第一の特徴のポリペプチドをコードする天然の遺伝子の発現を変化させることに有効であるか、または本発明の第一の特徴のポリペプチドの活性を調節することに有効である。

アゴニスト化合物またはアンタゴニスト化合物は、例えば細胞、無細胞調製物、化学物質ライブラリーまたは天然生成物の混合物から単離することができる。これらのアゴニストまたはアンタゴニストは天然または改変された基質、リガンド、酵素、レセプターまたは構造的もしくは機能的模倣物質であってよい。前記のスクリーニング技術の適切な概論のためには以下を参照されたい: Coligan et al. (1991) *Current Protocols in Immunology* 1(2) : Chapter 5.

10

#### 【0059】

おそらく良好なアンタゴニストと考えられる化合物は、本発明のポリペプチドと結合し、結合したときに前記ポリペプチドの生物学的作用を誘発しない分子である。潜在的なアンタゴニストには、本発明のポリペプチドと結合し、それによって本発明のポリペプチドの活性を阻害または消滅させる小型有機分子、ペプチド、ポリペプチドおよび抗体が含まれる。そのようなやり方で、前記ポリペプチドと正常な細胞の結合分子との結合が阻害され、その結果前記ポリペプチドの正常な生物学的活性が阻害され得る。

このようなスクリーニング技術で用いられる本発明のポリペプチドは溶液中で遊離しても、固相支持体に固定されても、細胞表面に保持されても、または細胞内に位置してもよい。一般に、このようなスクリーニングの方法は、前記のポリペプチドを発現している適切な細胞または細胞膜を用いることを含み、前記細胞または細胞膜をテスト化合物と接触させて、結合または機能的な応答の刺激もしくは阻害を観察する。続いて前記テスト化合物と接触させた細胞の機能的応答を、前記テスト化合物と接触させなかつたコントロール細胞と比較する。前記のようなアッセイによって、テスト化合物が前記ポリペプチドの活性化によって発生するシグナルをもたらすか否かを、適切な検出系を用いて評価する。活性化の阻害剤は、一般的には既知のアゴニストの存在下でアッセイされ、テスト化合物の存在下でアゴニストによる活性化の影響が観察される。

20

#### 【0060】

或いは、単純な結合アッセイを用いてもよい。この場合、テスト化合物のポリペプチド保持表面への付着が直接または間接的にテスト化合物と結合させた標識手段によって検出されるか、または標識競合物質を用いた競合を含むアッセイで検出される。別の態様では、競合薬剤のスクリーニングアッセイを用いることができる。この場合、ポリペプチドと特異的に結合することができる中和抗体がテスト化合物と結合について競合する。このようにして、前記抗体を用いて前記ポリペプチドに対して特異的な結合親和性を保有する一切のテスト化合物の存在を検出することができる。

30

前記ポリペプチドをコードするmRNAの細胞内の生成に対する添加テスト化合物の影響を検出するアッセイをデザインすることもできる。例えば、当技術分野で公知の標準的な方法によりモノクローナルまたはポリクローナル抗体を用いてポリペプチドの分泌レベルまたは細胞結合レベルを測定するELISAを構築することができ、それを用いて適切に操作した細胞または組織からのポリペプチド生成を阻害または増強し得る化合物について検索することができる。続いて、前記ポリペプチドとテストされる化合物との結合複合体の生成を測定することができる。

40

#### 【0061】

使用され得る薬剤スクリーニングのための別の技術は、対象の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高速大量処理スクリーニングを提供する(国際特許出願WO 84/03564を参照されたい)。前記方法では、多数の異なる小型のテスト化合物が固相基質上で合成され、次に本発明のポリペプチドと反応させられ洗浄され得る。ポリペプチドを固定する方法の1つは非中和抗体を使用することである。続いて、当技術分野で周知の方法を用いて結合ポリペプチドを検出することができる。精製ポリペプチドはまた、前述の薬剤スクリーニング技術で使用するためにプレートに直接被覆することができる。

50

本発明のポリペプチドを用いて、膜結合レセプターまたは可溶性レセプターを当技術分野で公知の標準的なレセプター結合技術により同定することができる。前記標準的な技術は、例えばリガンド結合アッセイおよび架橋アッセイであり、そのようなアッセイでは、ポリペプチドは放射性同位体で標識されているか、化学的に改変されているか、またはその検出もしくは精製を容易にするペプチド配列と融合されており、推定上のレセプター供給源（例えば細胞の組成物、細胞膜、細胞上清、組織抽出物または体液）とインキュベートされる。結合有効性は、生物物理的技術、例えば表面プラズモン共鳴および分光法を用いて測定することができる。結合アッセイは、レセプターの精製およびクローニングのために用いることができるが、ポリペプチドとそのレセプターとの結合に競合する前記ポリペプチドのアゴニストおよびアンタゴニストを同定するためにも用いることができる。スクリーニングアッセイを実施する標準的方法は当技術分野ではよく理解されている。 10

#### 【 0 0 6 2 】

本発明はまた、上記で述べたアゴニスト、アンタゴニスト、リガンド、レセプター、基質、酵素を同定する方法で有用なスクリーニングキットを含む。

本発明は、上記アゴニスト、アンタゴニスト、リガンド、レセプター、基質および酵素、並びに上記で述べた方法によって発見される、本発明のポリペプチドの活性または抗原性を調節する他の化合物を含む。

本発明はまた、本発明のポリペプチド、核酸、リガンドまたは化合物とともに適切な医薬担体を含む医薬組成物を提供する。これらの組成物は、下記で詳細に説明するように、治療用もしくは診断用試薬として、ワクチンとして、または他の免疫原性組成物として適切であり得る。 20

本明細書で用いられる専門用語にしたがえば、ポリペプチド、核酸、リガンドまたは化合物 { X } を含む組成物は、組成物中の  $X + Y$  の合計の少なくとも 85 質量 % が X である場合は不純物（本明細書中では Y ）を“実質的に含まない”。好ましくは X は、組成物中の  $X + Y$  の合計の少なくとも約 90 質量 % を、より好ましくは少なくとも約 95 質量 %、98 質量 % または 99 質量 % を構成する。

#### 【 0 0 6 3 】

医薬組成物は、好ましくは治療的に有効な量の本発明のポリペプチド、核酸分子、リガンドまたは化合物を含むべきである。本明細書で用いられる“治療的に有効な量”という用語は、標的疾患または症状を治療、緩和もしくは予防するために、または検出可能な治療効果もしくは予防効果を示すために必要な治療薬剤の量を指す。いずれの化合物についても、治療的に有効な投与量は、最初に細胞培養アッセイ（例えば新生物細胞培養アッセイ）または動物モデル（通常はマウス、ウサギ、イヌまたはブタ）のいずれかで見積もることができる。動物モデルは、適切な濃度範囲および投与経路の決定にも用いることができる。次にそのような情報を用いて、ヒトで有用な投与量および投与経路を決定することができる。 30

ヒト対象者に対する正確な有効量は、疾患状態の重篤度、対象者の全身の健康状態、対象者の年齢、体重および性別、食事、投与時間および投与回数、併用薬剤、反応感受性および治療に対する許容性 / 応答性に依存するであろう。前記の量は日常的検査により決定でき、それは臨床医の判断の範囲内であろう。一般には有効用量は 0.01mg/kg から 50mg/kg、好ましくは 0.05mg/kg から 10mg/kg であろう。本組成物は個別にまたは他の薬剤、医薬品またはホルモンと一緒に患者に投与することができる。 40

#### 【 0 0 6 4 】

医薬組成物はまた、治療薬の投与のために医薬的に許容できる担体を含むことができる。そのような担体には、抗体および他のポリペプチド、遺伝子並びに他の治療薬剤（例えばリポソーム）が含まれるが、ただし担体はそれ自体で前記組成物を投与される個体に有害な抗体の産生を誘発せず、さらに不都合な毒性をもたらすことなく投与されることを条件とする。適切な担体は大型でゆっくりと代謝される巨大分子、例えばタンパク質、多糖類、ポリ乳酸、ポリグリコール酸、重合アミノ酸、アミノ酸コポリマーおよび不活性ウイルス粒子であり得る。 50

それらの中には医薬的に許容できる塩、例えば鉛酸塩（塩酸塩、臭化水素酸塩、リン酸塩、硫酸塩などのような）；および有機酸の塩（酢酸塩、プロピオン酸塩、マロン酸塩、安息香酸塩などのような）を用いることができる。医薬的に許容できる担体についての綿密な考察は以下のテキストで入手可能である：Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991)。

治療用組成物中の医薬的に許容できる担体は、さらに液体、例えば水、生理食塩水、グリセロールおよびエタノールを含むことができる。さらに、助剤（例えば湿润剤または乳化剤、pH緩衝物質など）が前記組成物中に存在していてもよい。そのような担体は、患者が摂取できるように前記医薬組成物を錠剤、ピル、糖衣錠、カプセル、液剤、ゲル、シロップ、スラリー、懸濁剤などとして製剤化することを可能にする。

10

#### 【0065】

いったん製剤化されたら、本発明の組成物は直接対象者に投与することができる。治療される対象者は動物で、特にヒト対象者が治療され得る。

本発明で用いられる医薬組成物は、多数の経路（経口、静脈内、筋肉内、動脈内、骨髄内、硬膜下腔内、心室内、経皮適用（例えばWO98/20734を参照）、皮下、腹腔内、鼻内、腸内、局所、舌下、腔内または直腸的手段が含まれるが、ただしこれらに限定されない）によって投与できる。遺伝子錠またはハイポスプレーもまた、本発明の医薬組成物の投与に用いることができる。典型的には、本治療用組成物は、注射用物質（液体溶液または懸濁剤のいずれか）として調製できる。注射前に液体賦形剤中で溶液または懸濁剤とするために適した固体を調製することもできる。

20

本組成物の直接的デリバリーは一般に、皮下、腹腔内、静脈内または筋肉内に注射によって達成されるか、または組織の間隙腔に搬送されるであろう。前記組成物はまた、病巣に投与してもよい。投薬治療は単回投与スケジュールでも複数回投与スケジュールでもよい。

本発明のポリペプチドの活性が特定の疾患状態で過剰である場合には、いくつかのアプローチが利用可能である。あるアプローチは、医薬的に許容できる担体とともに上記のような阻害化合物（アンタゴニスト）を、前記ポリペプチドの機能阻害に有効な量で対象者に投与することを含む（前記化合物による前記機能阻害は、例えばリガンド、基質、酵素、レセプターの結合を遮断するか、または第二のシグナルを阻害し、それによって異常な症状を緩和することによって達成される）。好ましくは、前記アンタゴニストは抗体である。最も好ましくは、そのような抗体は、先に記載したようなその免疫原性を最少にするキメラ抗体および／またはヒト化抗体である。

30

#### 【0066】

別のアプローチでは、問題のリガンド、基質、酵素、レセプターに対する結合親和性を保持するポリペプチドの可溶形を投与することができる。典型的には、前記ポリペプチドは、関連部分を保持するフラグメントの形態で投与することができる。

また別のアプローチでは、前記ポリペプチドをコードする遺伝子の発現は、発現遮断技術を用いて、例えば内部で生成されるかまたは別々に投与されるアンチセンス核酸分子を使用して（上記で述べたように）阻害することができる。遺伝子発現の改変は、ポリペプチドをコードする遺伝子の制御領域、5'領域または調節領域（シグナル配列、プロモーター、エンハンサーおよびイントロン）に対して相補的な配列またはアンチセンス分子（DNA、RNAまたはPNA）をデザインすることによって達成できる。同様に、阻害は“三重らせん”塩基対方法論を用いて達成することができる。三重らせん対形成は、ポリメラーゼ、転写因子または調節分子の結合のために二重らせんが充分に開く能力を阻害することから有用である。三重DNAを用いる最近の治療の進歩は文献に記載されている（J.E. Gee et al. (1994) In: B.E. Huber & B.I. Carr, Molecular and Immunologic Approaches, Future Publishing Co., Mt. Kisco, NY）。相補的配列またはアンチセンス分子をデザインし、リボソームに対する結合を妨げて転写を妨害することによってmRNAの翻訳を遮断することもできる。そのようなオリゴヌクレオチドは投与されてもよいし、またin vivoでの発現によりin situで生成させてもよい。

40

50

## 【0067】

さらに、本発明のポリペプチドの発現は、そのコードmRNA配列に特異的なリボザイムを用いることによって妨げることができる。リボザイムは、天然または合成であり得る触媒的活性型のRNAである（例えば以下を参照されたい：N. Usman et al., *Curr. Opin. Struct. Biol.* (1996) 6(4) : 527 - 533）。合成リボザイムをデザインして、選択した位置でmRNAを特異的に切断し、それによってmRNAの機能的ポリペプチドへの翻訳を妨げることができる。リボザイムは、通常RNA分子で見出されるような天然のリン酸リボース骨格および天然の塩基から合成することができる。或いは、リボザイムは、非天然の骨格（例えば2'-0-メチルRNA）を用いて合成して、リボヌクレアーゼ分解から保護することができる。これらは改変塩基を含んでいてもよい。

10

RNA分子は細胞内安定性および半減期を増加させるために改変することができる。可能な改変には、RNA分子の5'および/または3'末端へのフランкиング配列の付加、または分子の骨格内でホスホジエステル結合に代わるホスホロチオエートまたは2'-0-メチルの使用が含まれるが、ただしこれらに限定されない。この概念はPNAの生成にも受け継がれ、さらに非慣用塩基（例えばイノシン、ケオシン(gueosine)およびブトシン(butosine)（アセチル-、メチル-、チオ-も同様に）、および同様に改変された形態のアデニン、シチジン、グアニン、チミンおよびウリジンで、これらは内因性エンドヌクレアーゼによって容易に認識されないようなものである）の包含によってPNA分子の全てに前記概念を広げることができる。

20

## 【0068】

本発明のポリペプチドおよびその活性の過小発現に関連する異常な状態を治療するためには、いくつかのアプローチも利用可能である。あるアプローチは、前記ポリペプチドを活性化する化合物（すなわち上記で述べたアゴニスト）の治療的に有効な量を対象者に投与し、異常な状態を緩和することを含む。あるいは、本ポリペプチドの治療量を適切な医薬担体と組合せて投与し、ポリペプチドの相対的な生理学的バランスを回復させてもよい。遺伝子治療を用い、対象者の関連する細胞による本ポリペプチドの内因性生成を行わせることができる。遺伝子治療は、不完全な遺伝子を修正した治療用遺伝子と置き換えることによって、前記ポリペプチドの不適切な生成を永久的に治療するために用いられる。

本発明の遺伝子治療は *in vivo* または *ex vivo* で実施できる。*Ex vivo* 遺伝子治療は、患者の細胞の単離および精製、治療用遺伝子の導入、および遺伝的に改変した細胞を患者に戻して導入することを必要とする。対照的に、*in vivo* 遺伝子治療は、患者の細胞の単離および精製を必要としない。

30

治療用遺伝子は患者に投与するために典型的には“封入されて”いる。遺伝子デリバリー賦形剤はリボソームのような非ウイルス、または、例えばK.L. Berkner (1992) *Currtop. Microbiol. Immunol.*, 158: 39 - 66に記載されたアデノウイルスのような複製欠損ウイルスまたはN. Muzyczka (1992) *Currtop. Microbiol. Immunol.*, 158: 97 - 129および米国特許第5,252,479号に記載されたアデノ付随ウイルス(AAV)ベクターであり得る。

例えば、本発明のポリペプチドをコードする核酸分子は、複製欠損レトロウイルスベクターで発現させるために操作を施すことができる。次にこの発現構築物を単離し、前記ポリペプチドをコードするRNAを含有するレトロウイルスプラスミドベクターで形質導入したパッケージ細胞に導入することができる。その結果、前記パッケージ細胞は対象の遺伝子を含有する感染性ウイルス粒子を産生することができるようになる。これらのプロデューサー細胞は *in vivo* で細胞を操作するために、さらに *in vivo* でポリペプチドを発現させるために対象者に投与することができる（以下を参照されたい：Gene Therapy and Other Molecular Genetic-based Therapeutic Approaches, Chapter 20（およびその中に引用された文献）, “Human Molecular Genetics” (1996) T. Strachan & A.P. Read, BIOS Scientific Publishers Ltd.）。

40

## 【0069】

別のアプローチは“裸のDNA”的投与で、この場合、治療用遺伝子は血流または筋肉組織に直接注射される。

50

本発明のポリペプチドまたは核酸分子が疾患をひき起こす原因物質である場合には、本発明は、前記疾患をひき起こす原因物質に対する抗体を生成するワクチンとして用いることができる。

本発明のワクチンは予防的（すなわち感染を防ぐ）または治療的（すなわち感染後の疾患を治療する）である。そのようなワクチンは、免疫性を付与する抗原、免疫原、ポリペプチド、タンパク質または核酸を、通常は上記で述べた医薬的に許容できる担体と組合せて含む。前記担体には、それ自体で組成物を投与される個体に対して有害な抗体の產生を誘発しない担体のいずれもが含まれる。さらに、これらの担体は免疫刺激剤（“アジュvant”）として機能してもよい。さらにまた、前記抗原または免疫原は、細菌の類毒素（例えばジフテリア、破傷風、コレラ、H. ピロリ菌（*pyroli*））および他の病原体と結合させることができる。10

ポリペプチドは胃で分解されるので、ポリペプチドを含むワクチンは好ましくは非経口的に（例えば皮下、筋肉内、静脈内または皮内注射）投与される。非経口投与に適した製剤には、水性および非水性滅菌注射溶液（前記は抗酸化剤、緩衝剤、抗菌剤および製剤を受容者の血液に対し等張にする溶質を含むことができる）、並びに水性および非水性滅菌懸濁剤（前記は分散剤または粘稠剤を含むことができる）が含まれる。

#### 【0070】

本発明のワクチン製剤は単位用量または複数単位用量容器で提供されてもよい。例えば封入したアンプルおよびバイアルは、使用直前に滅菌された液状担体を添加することだけを必要とする凍結乾燥状態で保存することができる。投与量はワクチンの比活性に依存し、日常的な検査で容易に決定することができる。20

本発明はまた、診断薬としての本発明の核酸分子の使用に関する。本発明の核酸分子により特徴付けられ、機能不全に付随する遺伝子の変異型の検出は、前記遺伝子の過小発現、過剰発現または位置的もしくは時間的発現の変化から生じる疾患の診断、または疾患に対する感受性の診断に付け加えることができるか、またはそれらを明確にすることができる診断ツールを提供する。前記遺伝子に変異を保有する個体は、種々の技術によってDNAレベルで検出することができる。

診断のための核酸分子は対象者の細胞、例えば血液、尿、唾液、組織生検または剖検材料から入手できる。ゲノムDNAを直接検出に用いてもよいし、PCR、リガーゼ連鎖反応（LCR）、鎖置換増幅（SDR）または他の増幅技術を分析前に用いることによって、ゲノムDNAを酵素的に増幅してもよい（以下の文献を参照されたい：Saiki et al., *Nature* 324 : 163 - 166(1986)；Bej et al., *Crit. Rev. Biochem. Molec. Biol.*, 26 : 301 - 334(1991)；Birkenmeyer et al., *J. Virol. Meth.*, 35 : 117 - 126(1991)；Van Brunt, J., *Bio/Technology*, 8 : 291 - 294(1990)）。30

#### 【0071】

ある態様では、本発明のこの特徴は、本発明のポリペプチドをコードする天然の遺伝子の発現レベルを評価し、さらに前記発現レベルをコントロールのレベルと比較することを含む、患者における疾患を診断する方法を提供する（ここで前記コントロールレベルと異なるレベルは疾患を示唆する）。前記方法は以下の工程を含む：

a ) 本発明の核酸分子と核酸プローブとの間でハイブリッド複合体の形成を可能にするストリンジエント条件下で、患者由来の組織サンプルを前記核酸プローブと接触させる工程；40

b ) コントロールサンプルを工程a ) で用いた条件と同じ条件下で前記プローブと接触させる工程；および、

c ) 前記サンプル中のハイブリッド複合体の存在を検出する工程；

この場合、コントロールサンプル中のハイブリッド複合体レベルと異なるハイブリッド複合体レベルが患者サンプルで検出されることは、疾患を示唆する。

本発明のさらなる特徴は、以下の工程を含む診断方法を含む：

a ) 疾患について検査される患者から組織サンプルを入手する工程；

b ) 前記組織サンプルから本発明の核酸分子を単離する工程；および、50

c) 前記核酸分子内の疾患に付随する変異の存在を検出することによって患者を疾患について診断する工程。

【0072】

上記に記載した方法における核酸分子の検出を補助するために、増幅工程、例えばPCRの使用を含むことができる。

正常な遺伝子型と比較すると、増幅生成物におけるサイズの変化によって、欠失および挿入が検出される。点変異は、増幅DNAを本発明の標識RNAとハイブリダイズさせるか、あるいは本発明の標識アンチセンスDNA配列とハイブリダイズさせることによって同定することができる。完全にマッチした配列は、RNase消化によって、または溶融温度における差異を評価することによって、ミスマッチを有する二重鎖と区別することができる。DNAをストリングエントな条件下で前記DNAとハイブリダイズする核酸プローブと接触させてハイブリッド二重鎖分子を形成させ(前記ハイブリッド二重鎖は、疾患に付随する変異に対応するいずれかの部分で前記核酸プローブ鎖のハイブリダイズしていない部分を有する)、DNA鎖の対応する部分における疾患付随変異の有無を示すものとして、前記プローブ鎖のハイブリダイズしていない部分の有無を検出することによって、患者における変異の有無を検出することができる。

前記のような診断は特に出生前検査で有用であり、新生児検査においてすら有用である。

【0073】

点変異、および参照遺伝子と“変異”遺伝子間で異なる他の配列は、他の周知の技術、例えば直接DNAシークエンシングまたは一本鎖構造多型性(Orita et al., *Genomics*, 5: 874 - 879(1989))によって同定できる。例えば、シークエンシングプライマーは、二本鎖PCR生成物または改変PCRによって作製された一本鎖テンプレート分子とともに用いることができる。配列決定は、放射能標識ヌクレオチドを用いる通常の方法によって、または蛍光タグを用いる自動シークエンシング法によって実施される。クローン化DNAセグメントを、特異的DNAセグメントを検出するためのプローブとして用いることもできる。この方法の感受性は、PCRと併用したとき極めて増強される。さらに、点変異および他の配列の変動(例えば多型性)は、例えば対立遺伝子特異的オリゴヌクレオチドをただ1つのヌクレオチドが異なる配列のPCR増幅に用いることによって、上記のように検出することができる。

DNA配列の相違はまた、変性剤の存在下または非存在下でのゲル内のDNAフラグメントの電気泳動移動度における変化によって、または直接DNAシークエンシング(例えば、Myers et al., *Science* (1985) 230: 1242)によって検出することができる。特定の位置における配列の変化はまた、ヌクレアーゼ保護アッセイ(例えばRNaseおよびS1保護)または化学切断法(Cotton et al., *Proc. Natl. Acad. Sci. USA* (1985) 85: 4397 - 4401を参照)によって明らかにできる。

【0074】

通常のゲル電気泳動およびDNAシークエンシングの他に、ミクロ欠損、異数性、転座、逆位のような変異は、*in situ*分析によっても検出できる(例えば以下を参照されたい: Keller et al., *DNA Probes*, 2nd Ed., Stockton Press, New York, N.Y., USA(1993))。すなわち、細胞内のDNAまたはRNA配列は、それらを単離および/またはメンブレン上に固定する必要なしに変異について分析することができる。蛍光*in situ*ハイブリダイゼーション(FISH)は、現在のところ最も一般的に用いられている方法で、FISHに関する多数の概論が存在する(例えば以下を参照されたい: Trachuck et al., *Science*, 250: 559 - 562(1990); およびTrask et al., *Trends Genet.* 7: 149 - 154(1991))。

本発明の別の態様では、本発明の核酸分子を含むオリゴヌクレオチドプローブのアレイを構築し、遺伝的変種、変異および多型性の効率的スクリーニングを実施することができる。アレイ技術方法はよく知られており汎用的に応用することができ、遺伝子発現、遺伝連鎖および遺伝的多様性を含む分子遺伝学における種々の疑問に取り組むために用いることができる(例えば以下を参照されたい: M. Chee et al., *Science* (1996) 274: 610 - 613)。

## 【0075】

ある態様では、前記アレイは以下の文献に記載された方法にしたがって調製され使用される（PCT出願W095/11995（Chee et al.）；D.J. Lockhart et al.（1996）Nat. Biotech. 14 : 1675 - 1680；M. Schena et al.（1996）Proc. Natl. Acad. Sci. 93 : 10614 - 10619）。オリゴヌクレオチド対は2つから100万個を越える範囲で変動し得る。前記オリゴマーは、光誘導化学方法を用いて基板上の指定領域で合成される。基板は紙、ナイロンまたは他のタイプのメンブレン、フィルター、チップ、ガラススライドもしくは他の適切な任意の固相支持体でよい。別の特徴では、オリゴヌクレオチドは、PCT特許出願（W095/251116, Baldeschweiler et al.）に記載されたように化学的結合方法およびインクジェット適用装置を用いて基板表面で合成することができる。別の特徴では、ドット（またはスロット）プロットに類似する“格子化（gridded）”アレイを用い、真空系、熱結合方法、UV結合方法、機械的または化学的結合方法を用いて基質表面にcDNAフラグメントまたはオリゴヌクレオチドを配置し、連結させることができる。アレイ（例えば上記で述べたようなもの）は手動で、または利用可能な装置（スロットプロットまたはドットプロット装置）、材料（適切な固相支持体すべて）および機械（ロボット機器を含む））を用いて作製することができ、8、24、96、384、1536または6144個のオリゴヌクレオチド、または2つから100万個を越える範囲の他のいずれの数をも含むことができる（このことはアレイ自体を商業的に入手可能な計測器の有効利用に向くものとしている）。 10

## 【0076】

上記で考察した方法の他に、対象者に由来するサンプルから、ポリペプチドまたはmRNAの異常な増加または低下レベルを決定することを含む方法によって、疾患を診断することができる。発現低下または発現増加は、ポリヌクレオチドの定量のために当技術分野で周知の方法のいずれか、例えば核酸増幅（例を挙げるとPCR、RT-PCR、RNase保護、ノザンプロット法）および他のハイブリダイゼーション方法を用いてRNAレベルで測定することができる。 20

宿主に由来するサンプルで本発明のポリペプチドレベルを決定することに用いることができるアッセイ技術は当業者によく知られており、さらに上記でいくらか詳細に考察されている（ラジオイムノアッセイ、競合結合アッセイ、ウェスタンプロット分析およびELISAアッセイを含む）。本発明のこの特徴では以下の工程を含む診断方法が提供される：（a）上記に記載したリガンドを、リガンド-ポリペプチド複合体の形成に適した条件下で生物学的サンプルと接触させる工程；および（b）前記複合体を検出する工程。 30

例えばELISA、RIAおよびFACSのようなポリペプチドレベルを測定するためのプロトコルは、さらにポリペプチド発現の変化レベルまたは異常レベルを診断するための基礎を提供することができる。ポリペプチド発現の正常値または標準値は、正常な哺乳類対象体（好ましくはヒト）から得られた体液または細胞抽出物を、複合体形成に適した条件下で前記ポリペプチドに対する抗体と混合することによって確立される。標準的な複合体形成量は種々の方法、例えば分光測定方法によって定量することができる。 40

## 【0077】

本発明のポリペプチドと特異的に結合する抗体は、前記ポリペプチドの発現によって特徴付けられる症状または疾患の診断のために、または本発明のポリペプチド、核酸分子、リガンドおよび他の化合物を用いて治療されている患者をモニターするアッセイにおいて、用いることができる。診断目的に有用な抗体は、治療薬として上記で述べたのと同じ様式で調製することができる。前記ポリペプチドについての診断アッセイは、前記抗体および標識を用いてヒトの体液または細胞もしくは組織の抽出物中のポリペプチドを検出する方法を含む。前記抗体は変更して、または変更せずに用いることができ、さらにそれらをレポーター分子と共有結合または非共有結合によって結合させることによって標識することができる。当技術分野で公知の多様なレポーター分子を用いることができる（それらのいくつかは上記に記載されている）。 50

生検組織由来の対象者、コントロールおよび疾患サンプルで発現されたポリペプチドの量は、標準値と比較される。標準値と対象者の値との間の偏差は疾患診断のためのパラメー

タを確立する。診断アッセイを用いて、ポリペプチド発現の有無および過剰を識別し、治療的処置の間のポリペプチドレベルの調節をモニターすることができる。そのようなアッセイはまた、動物実験、臨床試験または個々の患者の治療モニタリングにおける個々の治療的処置方法の有効性を評価することに用いることができる。

【0078】

本発明の診断キットは以下を含むことができる：

- (a) 本発明の核酸分子；
- (b) 本発明のポリペプチド；または
- (c) 本発明のリガンド。

本発明のある特徴では、診断キットは、ストリンジエントな条件下で本発明の核酸分子とハイブリダイズする核酸プローブを含む第一の容器；核酸分子を増幅するために有用なプライマーを含む第二の容器；および疾患の診断を容易にするために前記プローブおよびプライマーの使用についての指示書を含むことができる。前記キットはさらに、ハイブリダイズしていないRNAを消化するための試薬を保持する第三の容器を含むことができる。本発明の別の特徴では、診断キットは核酸分子のアレイを含み、前記核酸分子の1つが本発明の核酸分子であってもよい。

本発明のポリペプチドを検出するために、診断キットは以下を含むことができる：本発明のポリペプチドと結合する1つまたは2つ以上の抗体；および前記抗体とポリペプチドとの間の結合反応の検出に有用な試薬。

【0079】

そのようなキットは、以下のような疾患または疾患に対する感受性を診断することに有用であろう：特に細胞増殖疾患（新生物、メラノーマ、肺、結腸直腸、乳房、膵、頭部および頸部の腫瘍並びに他の固形腫瘍を含む）、骨髄増殖性疾患（例えば白血病、非ホジキンリンパ腫、白血球減少症、血小板減少症、血管形成疾患、カポジ肉腫）、自己免疫／炎症性疾患（アレルギー、炎症性腸疾患、関節炎、乾癬および気道の炎症、喘息および器官の移植拒絶を含む）、心脈管系疾患（高血圧、浮腫、アンギナ、アテローム性動脈硬化症、血栓症、敗血症、ショック、再灌流障害、心不整脈および虚血を含む）、神経学的疾患（中枢神経系疾患、アルツハイマー病、脳損傷、脳卒中、筋萎縮性側索硬化症、不安、抑うつおよび痛みを含む）、発達障害、代謝性障害（真性糖尿病、骨粗しょう症、脂質代謝障害、甲状腺機能亢進、甲状腺機能低下、上皮小体機能亢進、高カルシウム血症、低カルシウム血症、高コレステロール血症、高脂血症および肥満を含む）、腎疾患（糸球体腎炎、腎血管性高血圧を含む）、皮膚疾患（アクネ、湿疹および創傷治癒を含む）、加齢の負の作用、エイズ、感染（ウイルス感染、細菌感染、真菌感染および寄生虫感染を含む）、並びに他の病的状態（特に核内ホルモンレセプターが関与するもの）。

本発明の種々の特徴および態様は、特にLBDG3ポリペプチドおよびLBDG3完全長ポリペプチドへの言及を有する実施例を介してこれからより詳細に説明されるであろう。

細部の改変は本発明の範囲を逸脱することなく実施できることは理解されよう。

【0080】

（実施例）

実施例：1 CAB55953.1(LBDG3)

遠縁の新規な核内ホルモンレセプターリガンド結合ドメインの検索を開始するために、原型となるファミリーメンバー、ホモサピエンスレチノイン酸レセプター-2のリガンド結合ドメインを選択する。より具体的には、前記検索は、タンパク質データバンク（PDB）（Research Collaboratory for Structural Bioinformaticsによって運用されている）から得られた構造を用いて開始する。

選択した構造は、ホモサピエンスレチノイン酸レセプター-2のリガンド結合ドメインである（PDBコード1EXA：A；図1参照）。

1EXA：Aの類縁体についてバイオペンジウム（Biopendium）の検索を実施し（ターゲットマイニングインターフェースを用いる）、4462のゲノムスレッダーの結果が返される。この4462のゲノムスレッダーの結果には、典型的な核内ホルモンレセプターリガンド結合ド

10

20

30

40

50

メイン（例えばホモサピエンスレチノイン酸レセプター -1の残基182-471の間で見出されるもの）（図2 Aの矢印参照）の例が含まれる。

核内ホルモンレセプタリガンド結合ドメインを含むことが判明しているタンパク質の中で、核内ホルモンレセプタリガンド結合ドメインを含むと注釈付けされていないタンパク質、CAB55953.1 (LBDG3; 図2 Bの矢印参照) が現れる。インファーマチカゲノムスレッダーは、配列CAB55953.1 (LBDG3) の領域（残基311-452の間）をホモサピエンスレチノイン酸レセプター -2のリガンド結合ドメインと類似の構造を有すると同定した。ホモサピエンスレチノイン酸レセプター -2のリガンド結合ドメインと類似する構造を保有することは、CAB55953.1 (LBDG3) の残基311-452は核内ホルモンレセプタリガンド結合ドメインとして機能することを示唆する。ゲノムスレッダーはこれを94%の信頼度で同定する。

1EXA: Aの類縁体についてのバイオペンジウム (Biopendium) の検索（ターゲットマイニングインターフェースを用いる）によって、850のフォワード(Forward)PSI-BLASTの結果も返される。フォワードPSI-BLAST（図2 Cを参照されたい）は前記関係を同定できず、インファーマチカゲノムスレッダーだけがCAB55953.1 (LBDG3) は核内ホルモンレセプタリガンド結合ドメインを含むと同定できる。

#### 【0081】

CAB55953.1 (LBDG3) についてパブリックドメインデータベースで何が判明しているかを評価するために、重複配列ディスプレーページ (Redundant Sequence Display Page) (図3) を調査する。CAB55953.1 (LBDG3) が核内ホルモンレセプタリガンド結合ドメインを含むと同定するPROSITEヒットもPRINTSヒットも得られない。PROSITEおよびPRINTSは、類似ファミリーのタンパク質を記述するのに役立つデータベースである。両データベースから核内ホルモンレセプタリガンド結合ドメインのヒットが返ってこないことは、PROSITEまたはPRINTSを用いた場合、CAB55953.1 (LBDG3) が核内ホルモンレセプタリガンド結合ドメインを含むものとして同定できないことを意味する。

他のいずれかのパブリックドメインの注釈付け手段がCAB55953.1 (LBDG3) は核内ホルモンレセプタリガンド結合ドメインを含むと注釈付けすることができるか否かを明らかにするために、CAB55953.1 (LBDG3) タンパク質配列をインタープロ (InterPro) ウェブサイトでPFAMデータベース（アラインメントおよび隠れマルコフモデルのタンパク質ファミリーデータベース（Protein Family Database of Alignment and hidden Markov models））に対して検索する（図4参照）。核内ホルモンレセプタリガンド結合ドメインを含むとCAB55953.1 (LBDG3) を注釈付けているPFAM-Aマッチングは存在しない。したがってPFAMでは、CAB55953.1 (LBDG3) が核内ホルモンレセプタリガンド結合ドメインを含むことが同定されない。

次に、CAB55953.1 (LBDG3) についてパブリックドメインにおいて知られる更なる情報が存在するかを検討するために、全米バイオテクノロジー情報センター (NCBI) Genebankタンパク質データベースを調査した。Genebankは、タンパク質配列および遺伝子配列の寄託のためのUSパブリックドメインデータベースである（図5）。CAB55953.1 (LBDG3) はホモサピエンス配列であり、そのGenebankタンパク質IDはCAB55953.1であり、長さは560アミノ酸である。CAB55953.1 (LBDG3) はMIPS (Am Klopferspitz 18a, D-82152, Martinsried, Germany) の科学者グループによってクローン化された。この遺伝子についてのパブリックドメイン情報は、核内ホルモンレセプタリガンド結合ドメインを含むと前記遺伝子を注釈付けしていない。

#### 【0082】

したがって、全てのパブリックドメインの注釈付けツールを用いて、CAB55953.1 (LBDG3) は核内ホルモンレセプタリガンド結合ドメインを含むと注釈付けされないのであろうと結論することができる。インファーマチカゲノムスレッダーのみが、前記タンパク質は核内ホルモンレセプタリガンド結合ドメインを含むと注釈付けすることができる。ここで逆検索を実施する。CAB55953.1 (LBDG3) を今度はクエリー配列としてバイオペンジウムで用いる（図6 A参照）。インファーマチカゲノムスレッダーは、CAB55953.1 (LB

10

20

30

40

50

DG3) の残基311-452をホモサピエンスレチノイン酸レセプター -2のリガンド結合ドメインと同じ構造を有すると信頼度94%で同定する(図6Bの矢印参照)。ホモサピエンスレチノイン酸レセプター -2のリガンド結合ドメイン(1EXA:A)は、元々のクエリー配列であった。PSI-BLASTの正の繰返しはこの結果を返さない(図6C)。この関係を明らかにすることは、インファーマチカゲノムスレッダーだけである。

ホモサピエンスレチノイン酸レセプター -2リガンド結合ドメイン、1EXA:Aの配列は、CAB55953.1(LBDG3)の配列アラインメントを調査するために前記に対して選択される。類似構造を有すると同定されたタンパク質に対するクエリータンパク質のAIEyeアラインメント(図7)を観察することは、相同領域の可視化に役立つ。

ホモサピエンスレチノイン酸レセプター -2リガンド結合ドメインは“LBDモチーフ”を含み、前記モチーフは、今日までに注釈付けされた全ての核内ホルモンレセプターリガンド結合ドメインで見出されている。“LBDモチーフ”は核内ホルモンレセプターのコアクチベーターおよびコリプレッサーを補充することに関与する。6つの残基；PHE251、LEU254、ASP258、GLN259、LEU262およびLEU263は、ホモサピエンスレチノイン酸レセプター -2リガンド結合ドメインにおいてこのモチーフを構成する(図7の四角枠を参照)。しかしながら、それら6残基のうちの4つ(ASP258、GLN259、LEU262およびLEU263)のみがゲノムスレッダーアラインメントの範囲内に存在する。したがってそれら4残基のみが、核内ホルモンレセプターリガンド結合ドメインを含むというCAB55953.1(LBDG3)のゲノムスレッダーの注釈付けを強固にすることに用いることができる。CAB55953.1(LBDG3)の4残基(ASP314、GLN315、LEU318およびLEU319)は、ホモサピエンスレチノイン酸レセプター -2リガンド結合ドメインにおける“LBDモチーフ”残基の4つ(ASP258、GLN259、LEU262およびLEU263)全てと正確にマッチする(ゲノムスレッダーアラインメント領域内において)。このことは、CAB55953.1(LBDG3)はホモサピエンスレチノイン酸レセプター -2リガンド結合ドメインと類似する核内ホルモンレセプターリガンド結合ドメインを含むことを示している。

### 【0083】

同定されたタンパク質が実際にクエリー配列と関連することを保証するために、可視化プログラム“LigEye”(図8A)および“iRasmol”(図8B)を用いる。これらの可視化ツールは、既知の小型阻害分子が既知タンパク質構造の活性部位で相互作用するアミノ酸を表示することによって前記既知タンパク質構造の活性部位を同定する。これらの相互作用は、直接的水素結合を介するか、または疎水性相互作用を介する。このようにして、活性部位のフォールド/構造が、同定した相同体と選択した既知構造のタンパク質との間で保存されているか否かを理解することができる。ホモサピエンスレチノイン酸レセプター -2リガンド結合ドメインのLigEyeによる調査は、R-3-フルオロ-4-[2-ヒドロキシ-2-(5,5,8,8-テトラメチル-5,6,7,8-テトラヒドロ-ナフタレン(naphthalen)-2-イル)-アセチルアミノ]-安息香酸{略してBms270394と表記する}と結合する14残基(図7の丸付き文字)を明らかにする。しかしながら、これら14残基のうちの9残基(LEU271、MET272、ILE275、ARG278、PHE288、SER289、GLY303、PHE304およびLEU307)のみがゲノムスレッダーアラインメントの範囲内にある。したがって、これら9残基のみが、核内ホルモンレセプターリガンド結合ドメインを含むというCAB55953.1(LBDG3)のゲノムスレッダーの注釈付けを強固にするために用いることができる。

これら9残基の中の6つ(LEU271、MET272、ILE275、PHE288、PHE304およびLEU307)は疎水性であり、これらはホモサピエンスレチノイン酸レセプター -2リガンド結合ドメインの疎水性ポケットを形成している(Bms270394は、この疎水性の空洞に結合する)。これらの疎水性残基のうちの2つ(ILE275およびPHE304)は、CAB55953.1(LBDG3)において完全に保存されている(ILE331およびPHE363；図7を参照)。さらに、ホモサピエンスレチノイン酸レセプター -2リガンド結合ドメインのLEU271、MET272、PHE288およびLEU307は、それぞれCAB55953.1(LBDG3)では疎水性残基ILE327、LEU328、LEU349およびTYR366によって置換されている；(図7の穴あき丸付き文字)。結合ポケットの輪郭を形成する6つの疎水性残基のうちの6つが疎水性を保存することは(ゲノムスレッダーアラインメント

10

20

30

40

50

領域の範囲内で)、CAB55953.1(LBDG3)が疎水性ステロイド様リガンドと結合するであろうことを示している。このことは、実際インファーマチカゲノムスレッダーによって予測されたように、CAB55953.1(LBDG3)はホモサピエンスレチノイン酸レセプター-2リガンド結合ドメインと類似の様式で折り畳まれることを示し、したがって核内ホルモンレセプターリガンド結合ドメインを含むと同定される。

#### 【0084】

CAB55953.1(LBDG3)の逆最大化(reverse-maximaised)PSI-BLASTによって、マス・マスキュラス(Mus musculus)(AAF36513.1、図60矢印1を参照)およびダニオ・レリオ(Danio rerio)(AAF36514.1、図60矢印2、およびAAF36515.1、図60矢印3)における相同体を同定する。

CAB55953.1(LBDG3)、AAF36513.1(マス・マスキュラス相同体)、AAF36514.1(ダニオ・レリオ相同体)およびAAF36515.1(第二のダニオ・レリオ相同体)をAIEyeで整列化して調査した(図9)。

AIEyeは、リガンド結合ポケットの輪郭を形成することが予測されるCAB55953.1(LBDG3)の6つの疎水性残基(ILE327、LEU328、ILE331、LEU349、PHE363およびTYR366)がAAF36513.1(マス・マスキュラス相同体)、AAF36514.1(ダニオ・レリオ相同体)およびAAF36515.1(第二のダニオ・レリオ相同体)において全て正確に保存されていることを明らかにしている。さらにその上、CAB55953.1(LBDG3)における“LBDモチーフ”とマッチしている(ゲノムスレッダーアラインメントの範囲内で)4残基(ASP314、GLN315、LEU318およびLEU319)は、AAF36513.1(マス・マスキュラス相同体)、AAF36514.1(ダニオ・レリオ相同体)およびAAF36515.1(第二のダニオ・レリオ相同体)において全て正確に保存されている。タンパク質の機能に必須の残基は、前記タンパク質の相同体において保存されるであろう。したがって、予測されるCAB55953.1(LBDG3)核内ホルモンレセプターリガンド結合ドメインの機能に必須であろうと思われる“LBDモチーフ”および疎水性残基のマス・マスキュラス相同体および2つのダニオ・レリオ相同体における正確な保存は、CAB55953.1(LBDG3)が核内ホルモンレセプターリガンド結合ドメインを含むとの注釈付けを強く支持する。

#### 【0085】

図10は、NCBIユニジーン(UniGene)データベースから作製されたレポートである。このデータベースは、種々のヒト組織由来の発現配列タグ(EST)の収集であり、タンパク質に対し、その配列がデータベース中に存在することを条件として、全身組織分布を与えることに用いられ得る。CAB55953.1(LBDG3)はデータベース中に存在しており、次に記述するような組織において発現されていることが示されている:大動脈、血液、脳、乳房、CAN、結腸、眼、生殖細胞、心臓、腎臓、肺、筋肉、鼻、卵巣、臍臓、上皮小体、末梢神経系、胎盤、前立腺、胃、精巣、子宮、全胚、正常羊膜、膀胱腫瘍、脳、乳房、正常乳房、頸部、結腸、結腸-ins、正常結腸、デニス-ドラッシュ(denis-drash)、epid-腫瘍、眼、尿生殖路、頭部-頸部、正常頭部、腎臓、腎臓腫瘍、胚、胚腫瘍、リンパ、筋肉、神経腫瘍、卵巣、臍臓、胎盤、ピネット(pnet)、正常前立腺、皮膚、胸腺、プール(pooled)、甲状腺、子宮、子宮腫瘍、全血。

パブリックドメイン情報を用いて、CAB55953.1(LBDG3)を計算機により(in silico)細胞遺伝学的遺伝子座にマッピングできる。CAB55953.1(LBDG3)は、NCBIサイトにおいてePCRによりゲノムフラグメントAL13285.35の中に見出され得る(図12矢印)AL117480によってコードされている(図11)。AL13285.35はマッピングされており、20q11.21-12の間に位置している(図13矢印)。したがって、CAB55953.1(LBDG3)をコードする遺伝子も20q11.21-12の間に位置しているはずである。

自己免疫性甲状腺疾患(AITD)グレーブス病の感受性遺伝子は、20q11.2にマッピングされている(図14)。このことは、甲状腺ホルモン(および関連分子)が核内ホルモンレセプターリガンド結合ドメインに結合することが知られていることから、特に興味深い。ユニジーンのレポートは、CAB55953.1(LBDG3)が甲状腺で発現していることも示している。20q11.2-q12間のゲノム領域の増幅は、例えば胃癌といった種々の癌と関連がある(

10

20

30

40

50

図15）。したがってCAB55953.1 (LBDG3) 遺伝子は前記疾患状態に関連し得ることが推測され、それによりCAB55953.1 (LBDG3) 遺伝子またはそのコードタンパク質を標的として前記疾患状態を診断および/または治療する薬物の開発への道を開き得るであろう。特に、核内ホルモンレセプタリガンド結合ドメインを含むという前記遺伝子の同定は、例えば核内ホルモンレセプタリガンド結合ドメインアゴニストまたはアンタゴニストの使用を通じて、CAB55953.1 (LBDG3) タンパク質に対して特異的な薬物の開発を容易にする。

### 【0086】

#### 実施例2

提唱したLBDの組織発現を定量することを目的として、タックマン(Taqman)RT-PCR定量を用いた。タックマン3' - 5' エキソヌクレアーゼアッセイは、2つのプライマー認識配列間の部位で標的テンプレートとハイブリダイズする二重標識蛍光発生プローブの核酸分解を含む過程によって、PCR単位複製配列の形成にシグナルを送る(米国特許第5,876,930号を参照)。ABIプリズム(Prism)7000は、各増幅サイクルの間に、生成した単位複製配列の量に関する化学量論的な前記シグナルの検出および定量的測定を自動化して行う。PCR分析のために必要な時間および労働量の実質的な節減を提供することに加えて、この技術は反応における標的配列の単純化および非常に高精度な定量化を可能にする。

非罹病器官から調製したヒトRNAは、Ambion Europe (Huntingdon, UK) またはClontechから購入した。オリゴヌクレオチドプライマーおよびプローブは、Primer Expressソフトウェア (Applied Biosystems, Foster City CA) を用いて、40~60%のGC含有量を有し、プローブの5'末端にG-ヌクレオチドが存在せず、4つより多くの連続するGが存在しないように設計した。次にBLAST(登録商標)(Basic Local Alignment Search Tool, Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. : J Mol Biol 1990 Oct 5 ; 215(3) : 403 - 10)を用いて各プライマーおよびプローブを解析した。各オリゴヌクレオチドは、公的に利用可能なデータベースであるユニジーン(Unigene)およびゴールデンパス(GoldenPath)において提示された他の既知cDNA配列または既知ゲノム配列と比較した場合に3 bpより高い特異性で標的配列を認識したという結果を確認した。そのプライマーおよびプローブの配列は、次のようである：

CAC14946.1 (LBDG3) 順方向プライマー : AAC ATC CCA GAG GTG GAA GCT

CAC14946.1 (LBDG3) 逆方向プライマー : CAG ACG AGT TAA TAA GCC CCT CTT

CAC14946.1 (LBDG3) プローブ : CAT CAC ACA CCA AAC TCC CGG TGT T

### 【0087】

18sの前最適化プライマーおよびプローブは、Applied Biosystems, Foster City, CAから購入した。プローブは、最も5'の塩基および最も3'の塩基において、それぞれ蛍光レポーター色素(例えば、6カルボキシ-フルオレセイン{FAM};  $\lambda_{em}$  = 518 nm)および蛍光クエンチャーカラム(carboxytetram) - エチル - ローダミン{TAMRA};  $\lambda_{em}$  = 582 nm)と共有結合させた。全てのプライマーおよびプローブは、Perkin Elmer, Germanyから入手した。プライマー/プローブ濃度を50 nMから900 nMの範囲で滴定して、効率的なPCR反応のための最適濃度を決定した。プライマーおよびプローブの最適濃度は、増幅された標的遺伝子に依存して100 nMから900 nMの間を変動する。cDNAはPE Applied Biosystems, Foster City CAが提供している構成成分を用いて調製した。RNaseフリーの0.5 mlチューブに、50  $\mu$  lの反応物を調製する。その反応物は、総RNA 500 ng; 1×逆転写酵素緩衝液; 5.5 mM MgCl<sub>2</sub>; 1 mM dNTP; ランダムヘキサマー 2.5  $\mu$  l; RNase阻害剤20U; および逆転写酵素 62.5Uを含む。

0.5 mlの薄壁光学グレードのPCR96ウェルプレート(PE Applied Biosystems, Foster City CA)に25  $\mu$  lの反応物を調製した。その反応物は、1×最終濃度のタックマン・ユニバーサルマスターミックス(Universal Master Mix)(PE Applied Biosystems, Foster City CAが独自に開発したAmpliTaq Gold DNAポリメラーゼ、AmpEraseX UNG、UTPを有するdNTP、受動的参照色素および最適化した緩衝液構成成分の混合物); 100 nM タックマンプローブ; 300 nM 順方向プライマー; 900 nM 逆方向プライマーおよびcDNAテンプレート 15

10

20

30

40

50

ngを含んでいた。

#### 【 0 0 8 8 】

ABIプリズム7000の操作の標準的手順または同様の検出系を用いた。これには、反応体積を50から25  $\mu$ lに変えたことを除き、例えばABIプリズム7000での全デフォルトプログラム設定の使用が含まれていた。熱サイクリング条件は、50 で2分間、95 で10分間に続いて、95 で15秒間および60 で1分間の40サイクルで構成した。サイクル閾値(Ct)決定(すなわち、PCR産物の合成により生じるレポーター色素蛍光がバックグラウンド蛍光レベルより著しく高くなるのに必要なサイクル数の非整数計算)は、デフォルトパラメータを用い、各反応に対して機器により自動的に行つた。同じcDNAサンプル中の標的配列およびリボソーム18s(参照)配列に対するアッセイは、別々の反応チューブで行った。各実験では、典型的な組織サンプルの50 ngから0.78 ngのcDNAテンプレートで標準曲線化を行つた。その標準曲線から、各テストサンプルにおける実際の標的cDNAまたは18s cDNAの出発量を決定する。

各サンプル中の標的cDNAのレベルは、胃における標的の発現レベルに正規化した。各サンプル中の18s cDNAのレベルは、胃における18sの発現レベルに正規化した。次にそのデータを、任意で1に設定した胃のcDNAにおける発現レベルに比例する正規化標的配列の倍率発現として表した。各実験に対して、転写物を2-4回定量化した。前記データは、実験のセットについての平均値  $\pm$  SEMを示している。

#### 【 0 0 8 9 】

詳細な説明に記述するように、CAC14946.1(LBDG3)に特異的なプライマー/プローブを用いて、タックマンRT-PCRを結腸cDNAの2倍希釈物に対して行つた。図16は投入量cDNAの対数に対してプロットしたCt値を示しており、前記範囲の投入量cDNA濃度にわたって線形関係が見られることを図示している。テストCt値からcDNAの出発量を算出するために、標準曲線の直線回帰分析を用いた。

詳細な説明に記述するように、18s rRNAに特異的なプライマー/プローブを用いて、タックマンRT-PCRを結腸cDNAの2倍希釈物に対して行つた。図17は投入量cDNAのlog値に対してプロットしたCt値を示しており、前記範囲の投入量cDNA濃度にわたって線形関係が見られることを図示している。テストCt値からcDNAの出発量を算出するために、標準曲線の直線回帰分析を用いた。

詳細な説明に記述するように、CAC14946.1(LBDG3)および18s rRNAに対して特異的なプライマー/プローブを用い、指定したcDNA15 ngを用いてタックマンRT-PCRを行つた。典型的な組織サンプルの50 ngから0.78 ngのcDNAテンプレートを用いて、標的および内部標準についての標準曲線化も行つた。標準曲線の直線回帰分析を用いて、各テストサンプル中の標的cDNAまたは18s cDNAの実際の出発量を算出するためにそのCt値を使用した。

各サンプル中の標的cDNAレベルは、比較サンプル(この場合は胃である)中の標的の発現レベルに正規化した。各サンプル中の18s cDNAレベルも、胃の18s発現レベルに正規化した。次に、CAC14946.1(LBDG3)の発現レベルを18sの発現レベルに正規化した。図13は、任意で1に設定した胃のcDNAにおける発現レベルに関して正規化した標的配列の倍率発現を表している。各サンプルは、3つの個別の実験で定量化した。図18は、複数の実験についての平均値  $\pm$  SEMを示す。

#### 【 0 0 9 0 】

LBDG3のmRNAがヒト脳において有意なレベルで発現されているという発見は注目すべきことであり、そのことがヒト疾患の状態に潜在的な関連を与えることから、LBDG3のリガンド結合ドメインに対するアゴニストおよびアンタゴニストの開発は以下に挙げる種々のヒト疾患における治療的処置の可能性を提供する：細胞増殖性疾患(新生物、メラノーマ、肺、結腸直腸、乳房、臍、頭部および頸部の腫瘍、および他の固形腫瘍を含む)、骨髄増殖性疾患(例えば白血病、非ホジキンリンパ腫、白血球減少症、血小板減少症、血管形成疾患、カポジ肉腫)、自己免疫/炎症性疾患(アレルギー、炎症性腸疾患、関節炎、乾癬および気道の炎症、喘息および器官の移植拒絶を含む)、心脈管系疾患(高血圧、浮腫、アンギナ、アテローム性動脈硬化症、血栓症、敗血症、ショック、再灌流障害、心不整脈

10

20

30

40

50

および虚血を含む)、神経学的疾患(中枢神経系疾患、アルツハイマー病、脳損傷、脳卒中、筋萎縮性側索硬化症、不安、抑うつおよび痛みを含む)、発達障害、代謝性障害(真性糖尿病、骨粗しょう症、脂質代謝障害、甲状腺機能亢進、甲状腺機能低下、上皮小体機能亢進、高カルシウム血症、低カルシウム血症、高コレステロール血症、高脂血症および肥満を含む)、腎疾患(糸球体腎炎、腎血管性高血圧を含む)、皮膚疾患(アクネ、湿疹および創傷治癒を含む)、加齢の負の作用、エイズ、感染(ウイルス感染、細菌感染、真菌感染および寄生虫感染を含む)、並びに他の病的状態(特に核内ホルモンレセプターが関与するもの)。

DNA結合ドメインが存在しないリガンド結合ドメインを含むLBDG3のような“非古典的”核内ホルモンレセプターの発見は脳におけるステロイド(神経ステロイドとして知られる)の広範囲な効果を一貫して報告している既知文献と一致しており、それらの効果は、一般的には転写活性を必要とする既知の古典的ステロイドホルモン核内レセプターを介さずに媒介される。一例を挙げると、神経ステロイドは、特にGABA、NMDAおよびシグマレセプターのようなレセプターの域において神経伝達に影響することが示されている。神経ステロイドは、神経保護的役割を果たすことが示されている。したがってLBDG3に対するアゴニスト(またはアンタゴニスト)の開発を介する治療的処置は、梗塞または出血(脳卒中)、並びに中枢神経系および脊髄の損傷のような脳血管性疾患に続く痴呆、パーキンソン病および神経変性のような神経変性状態の治療において役割を有し得る。さらに神経ステロイドは、認識処理、空間学習および記憶、不安、並びに常習行為パターンを引き起こす欲求のような行為に影響することが示されている。したがってLBDG3に対するアゴニストおよびアンタゴニストの開発は、痴呆、学習困難、不安、常習行為(例えばアルコール症、摂食障害および薬物常習であるが、これだけに限られない)を治療するための治療的処置をもたらし得る。

#### 【0091】

LBDG3は脳において独占的には発現されない。有意なレベルのmRNAが副腎、卵巣、精巣および胸腺においても見出されている。副腎、卵巣および精巣はステロイドの生合成にとって意味のある部位であり、その活性はLBDG3にとってのステロイド作用における役割(それだけに限られない)と一致している。それらの組織におけるLBDG3の発見は、以下に挙げる疾患(これだけに限られない)の治療のためのアンタゴニストおよびアゴニストの開発を支持する:高血圧、感染症のストレスを含むストレスに対する応答、塩類および水分恒常性の調節、排卵調節を通じての受胎率制御(不妊症および避妊)、着床調節(不妊症および避妊)および精子形成の調節(不妊症および避妊)。さらに、そのような薬物は、良性前立腺肥大、前立腺癌、卵巣癌および精巣癌のようなステロイド感受性腫瘍を治療することにおいて価値があり得る。

胸腺におけるLBDG3の発見は、T細胞の発生における役割と一致している。したがってLBDG3に対して開発したアゴニストおよびアンタゴニストは、以下のような疾患の進行においてT細胞を調節することに役割を果たし得る:自己免疫疾患およびアレルギー(I型真性糖尿病を含む)、慢性関節リウマチ、多発性硬化症、感染、糸球体症に起因する腎不全、強皮症、炎症性腸疾患(クローン病および潰瘍性大腸炎)、移植拒絶、喘息、アトピー性皮膚炎、湿疹、エイズ、感染症(ウイルス感染症、細菌感染症、真菌感染症および寄生虫感染症を含む)、並びに他の病的状態(特に核内ホルモンレセプターが関与するもの)。

#### 【0092】

##### 実施例3

本明細書中でCAB55953.1(LBDG3)の3つの相同体を同定し、CAB55953.1(LBDG3)と整列化して、相同体の全体にわたって主要な機能性残基の保存を例証した(例えば予測される“LBDモチーフ”的残基;図6Cを参照)。それら相同体は、マス・マスキュラスAAH03931.1、ダニオ・レリオAAF36515.1およびダニオ・レリオAAF36514.1である。相同体の全体にわたる前記残基の保存は、進化的圧力がそれら特定の残基に特異的に適用されることを示しており、“LBDモチーフ”残基としての注釈付けを支持している。このことは、CAB55953.1(LBDG3)が核内ホルモンレセプターリガンド結合ドメインを含むという注釈付けに

10

20

30

40

50

対して支持を与える。

マス・マスキュラス相同体 (AAH03931.1) は、アミノ酸レベルでCAB55953.1 (LBDG3) に対し98%の配列同一性を有する。ダニオ・レリオ相同体 (AAF36514.1) は、アミノ酸レベルでCAB55953.1 (LBDG3) に対し74%の配列同一性を有する。ダニオ・レリオ相同体 (AAF36515.1) は、アミノ酸レベルでCAB55953.1 (LBDG3) に対し69%の配列同一性を有する。より不一致な他のCAB55953.1 (LBDG3) の相同体は、CAB55953.1 (LBDG3) が核内ホルモンレセプターリガンド結合ドメインを含むという注釈付けのさらなるテストに役立つであろう：全体配列同一性が30%に向かって低下する場合であっても、機能的に重要である残基（例えば“LBDモチーフ”の残基）はなお保存されるであろうことが予測される。

より不一致な2つの新しいCAB55953.1 (LBDG3) の相同体もNCBI月間(month)データベースにおいて同定された。CAB55953.1 (LBDG3) の相同体は、海生脊索動物オイコプルーラ (Oikopleura) のゲノムBAC配列AF374376.1から同定された。このオイコプルーラ相同体はCAB55953.1 (LBDG3) と44%の配列同一性を共有しており、これをLBDG3-Oikoと称する。

CAB55953.1 (LBDG3) の別の相同体は、脊索動物チオナ (Ciona) に由来する転写物の5' (AL666329.1) および3' (AL664923.1) EST配列から同定された。そのEST配列はオーバーラップしておらず、そのギャップは配列中で“X”によって示されている。このチオナ相同体はCAB55953.1 (LBDG3) と37%の配列同一性を共有しており、これをLBDG3-Cionaと称する。図19は、ホモサピエンスレチノイン酸レセプター リガンド結合ドメイン (1EXA: A) を有するCAB55953.1 (LBDG3) とマス・マスキュラス (AAH03931.1) 、ダニオ・レリオ (AAF36515.1およびAAF36514.1) 、チオナ (LBDG3-Ciona) およびオイコプルーラ (LBDG3-Oiko) に由来するCAB55953.1 (LBDG3) 相同体とのアライメントを示している。

#### 【0093】

ホモサピエンスレチノイン酸レセプター リガンド結合ドメインは、今日までに注釈付けられた全ての核内ホルモンレセプターリガンド結合ドメインにおいて見出されている“LBDモチーフ”を含む。“LBDモチーフ”は核内ホルモンレセプターのコアクチベーターおよびコリプレッサーを補充することに関与する。6つの残基PHE251、LEU254、ASP258、LEU262およびLEU263は、ホモサピエンスレチノイン酸レセプター リガンド結合ドメインにおいて前記モチーフを構成する（図19の黒い囲みで印をしたc、d、e、f、gおよびhを参照）。上で考察したように、その“LBDモチーフ”残基のうちの4つは、CAB55953.1 (LBDG3) で保存されている。1EXA: AのASP258はCAB55953.1 (LBDG3) ではASP314として保存されている。1EXA: AのGLN259はCAB55953.1 (LBDG3) ではGLN315として保存されている。1EXA: AのLEU262はCAB55953.1 (LBDG3) ではLEU318として保存されている。1EXA: AのLEU263はCAB55953.1 (LBDG3) ではLEU319として保存されている。顕著なことに、それら4残基は5つのCAB55953.1 (LBDG3) 相同体において全て保存されている (LBDG3-OikoにおけるLEU318に対するTYRの保存的置換を除いて) (図19の黒い囲みで印をしたc、d、e、f、gおよびhを参照)。特に全体配列同一性がLBDG3-Cionaに対する37%と同じように低い場合に、それら4残基の特異的な保存は、それら4残基がCAB55953.1 (LBDG3) および5つのCAB55953.1 (LBDG3) 相同体において重要な役割を果たすことに対する強力な証拠を提供する。このことは、ASP314、GLN315、LEU318およびLEU319が“LBDモチーフ”残基として機能していることと一致し、CAB55953.1 (LBDG3) が核内ホルモンレセプターリガンド結合ドメインを含むという注釈付けを支持する。

“LBDモチーフ”残基に加えて、既知の核内ホルモンレセプターリガンド結合ドメインにおいてよく保存されている他の残基が存在する。CAB55953.1 (LBDG3) が核内ホルモンレセプターリガンド結合ドメインを含むという注釈付けのさらなるテストは、“LBDモチーフ”の外側の重要な残基がCAB55953.1 (LBDG3) および5つのCAB55953.1 (LBDG3) 相同体で保存されているかどうかを解析することである。

#### 【0094】

重要な保存残基に関するデータの一つの便利な供給源は、Homstradデータベースである (Mizuguchi K, Deane CM, Blundell TL, Overington JP. (1998) HOMSTRAD: a database of protein structure alignments for homologous families. Protein Science 7: 2469-50

2471)。Homstrad ( HOMologous STRucture Alignment Database ) は、残基が特定のタンパク質ファミリーの全般にわたって保存されていることを実証し、その構造上の役割を詳述する。図20は、既知の核内ホルモンレセプターリガンド結合ドメイン構造の選択に関するHomstradレポートを示している。

Homstradレポートは、ホモサピエンスレチノイン酸レセプター 構造2LBDへのエントリーを含む。2LBDおよび1EXA : Aは、ホモサピエンスレチノイン酸レセプター リガンド結合ドメインの2つの等価構造物であり、その結合リガンドにおいてのみ異なっている ( 2LBD はオールトランス型レチノイン酸を含み、1EXA : Aはbms270394を含む )。図20の矢印aを参照すると、ホモサピエンスレチノイン酸レセプター のPHE244が、他の核内ホルモンレセプターリガンド結合ドメインにおいては溶媒非接近性の芳香族残基として保存されている溶媒非接近性残基であることが示され得る ( 1LBD ホモサピエンスレチノイドXレセプター 、1PRG : A ホモサピエンスペルオキシソーム増殖因子活性化レセプター 、3ERT : A ホモサピエンスエストロゲンレセプター 、および1A28 : A ホモサピエンスプロゲステロンレセプター )。既知の核内ホルモンレセプターリガンド結合ドメインファミリーの全般にわたる前記溶媒非接近性芳香族残基の保存は、その残基がリガンド結合ドメインフォールドにおいて重要な構造上の役割を果たすことを示している。ゲノムスレッダーアラインメントでは、ホモサピエンスレチノイン酸レセプター のPHE244は、CAB55953.1 ( LBDG3 ) のPHE305として保存されている ( 図19囲みa )。この主要な溶媒非接近性芳香族残基の保存は、CAB55953.1 ( LBDG3 ) が核内ホルモンレセプターリガンド結合ドメインを含むという注釈付けを支持する。その保存の意義は、PHE305が全5つのCAB55953.1 ( LBDG3 ) 相同体においてPHEとして保存されているという知見によって高められる ( 図19囲みa )。このことは、CAB55953.1 ( LBDG3 ) が核内ホルモンレセプターリガンド結合ドメインを含むという注釈付けをさらに支持する。

#### 【 0 0 9 5 】

図20の矢印bを参照すると、ホモサピエンスレチノイン酸レセプター のLYS246は、他の核内ホルモンレセプターリガンド結合ドメインでは溶媒接近性の正荷電残基として保存されている溶媒接近性残基であることが示され得る ( 1LBD ホモサピエンスレチノイドXレセプター 、1PRG : A ホモサピエンスペルオキシソーム増殖因子活性化レセプター 、3ERT : A ホモサピエンスエストロゲンレセプター 、および1A28 : A ホモサピエンスプロゲステロンレセプター )。既知の核内ホルモンレセプターリガンド結合ドメインファミリーの全般にわたる前記溶媒接近性の正荷電残基の保存は、その残基がリガンド結合ドメイン機能において重要な役割を果たすことを示している。実際、実験データは、前記のような溶媒露出する正荷電残基が核内ホルモンレセプターコアクチベーターの補充において“電荷クランプ”として機能することを示している ( A. K. Shiao, D. Barstad, P. M. Loria, Lin Cheng, P. J. Kushner, D. A. Agard, and G. L. Greene, The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen Cell 1998 95 : 927 )。ゲノムスレッダーアラインメントでは、ホモサピエンスレチノイン酸レセプター のLYS246は、CAB55953.1 ( LBDG3 ) の正荷電ARG307によって保存的に置換されている ( 図19囲みb )。この主要な正荷電残基の保存は、CAB55953.1 ( LBDG3 ) が核内ホルモンレセプターリガンド結合ドメインを含むという注釈付けを支持する。その保存の意義は、ARG307が全5つのCAB55953.1 ( LBDG3 ) 相同体でARGとして保存されているという知見によって高められる ( 図19囲みb )。このことは、CAB55953.1 ( LBDG3 ) が核内ホルモンレセプター結合ドメインを含むという注釈付けをさらに支持する。

#### 【 0 0 9 6 】

図20の矢印iを参照すると、ホモサピエンスレチノイン酸レセプター のASP269は、他の4つの核内ホルモンレセプターリガンド結合ドメインのうちの3つにおいて酸性残基として保存されていることが示され得る ( 1LBD ホモサピエンスレチノイドXレセプター 、1PRG : A ホモサピエンスペルオキシソーム増殖因子活性化レセプター および3ERT : A ホモサピエンスエストロゲンレセプター )。既知の核内ホルモンレセプターリガンド結合ド

10

20

30

40

50

メインの多数における前記酸性残基の保存は、その残基がリガンド結合ドメインフォールドにおいて重要な役割を果たすことを示している。ゲノムスレッダーアラインメントでは、ホモサピエンスレチノイン酸レセプターのASP269は、CAB55953.1(LBDG3)のGLU325として保存的に置換されている(図19囲みi)。この酸性残基の保存は、CAB55953.1(LBDG3)が核内ホルモンレセプター結合ドメインを含むという注釈付けを支持する。その保存の意義は、GLU325が全5つのCAB55953.1(LBDG3)相同体で酸性の(GLUまたはASP)として保存されているという知見によって高められる(図19囲みi)。このことは、CAB55953.1(LBDG3)が核内ホルモンレセプター結合ドメインを含むという注釈付けをさらに支持する。

## 【0097】

10

ゲノムスレッダーは、CAB55953.1(LBDG3)が核内ホルモンレセプター結合ドメインを含むと注釈付けしている。複合二次構造予測プログラムPSI-PRED(Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195-202)は、その注釈付けをテストする独立な方法を提供する。図21は、ホモサピエンスレチノイン酸レセプター(1EXA:A)の結晶構造に由来する実際の既知二次構造と比較したホモサピエンスレチノイン酸レセプター配列(“LBDモチーフ”領域において)についてのPSI-PRED出力を示している。PSI-PREDは、ホモサピエンスレチノイン酸レセプターの配列からヘリックスの3および5を正しく予測していることが理解できる。PSI-PREDは70%の精度を有することが実証されている。CAB55953.1(LBDG3)が実際にリガンド結合ドメインフォールドを採用する場合は、ホモサピエンスレチノイン酸レセプターのリガンド結合ドメインフォールドに対して同一の二次構造を有するべきであろう。図21は、CAB55953.1(LBDG3)(“LBDモチーフ”領域において)についてのPSI-PRED出力を示している。それをホモサピエンスレチノイン酸レセプター(1EXA:A)の既知二次構造と比較した場合、CAB55953.1(LBDG3)は1EXA:Aの3および5に対応する2つのヘリックスを採用すると予測されることが明らかに理解され得る。このことは、CAB55953.1(LBDG3)が核内ホルモンレセプター結合ドメインを含むというゲノムスレッダーの注釈付けの独立な確証である。ホモサピエンスレチノイン酸レセプター、1EXA:Aの構造における5に特有な特徴は、5がヘリックスに沿った中途で“ねじれ”を有することである。その5ヘリックスのねじれは、他の既知のリガンド結合ドメインにおいても観察される。5ヘリックスのねじれは、リガンド結合ポケットを形成することにおいて重要である。5ヘリックスのねじれは図21で矢印として記されており、ホモサピエンスレチノイン酸レセプターのPSI-PRED出力ではヘリックス予測の確率における小さくぼみとして反映されていることが理解できる。興味深いことに、CAB55953.1(LBDG3)も正確にその位置でヘリックス予測の確率におけるくぼみを示している。

20

## 【0098】

30

ホモサピエンスレチノイン酸レセプター、1EXA:Aの既知構造は、3および5を連結する3残基長の $3_{10}$ ヘリックス(4)を含む(ホモサピエンスレチノイン酸レセプターについてのPSI-PRED出力から、PSI-PREDがそのプログラムを用いて期待されるであろう前記 $3_{10}$ ヘリックスの予測を行わないことが理解され得る)。CAB55953.1(LBDG3)と1EXA:Aのゲノムスレッダーアラインメントは、1EXA:Aの $3_{10}$ ヘリックスと整列化した領域においてCAB55953.1(LBDG3)中にギャップを挿入する。このことは、3および5に対応するヘリックスは $3_{10}$ ヘリックスによって連結されないが、代わりに配列(GLY308-THR309-THR310)を有する短いループ領域によって連結されることを示している。実際、前記 $3_{10}$ ヘリックスの予測された不在は、CAB55953.1(LBDG3)中の“LBDモチーフ”に由来する先頭の2残基(1EXA:AのPHE251およびLEU254)の不在を説明している。

40

CAB55953.1(LBDG3)およびホモサピエンスレチノイン酸レセプターについての出力と整列化された5つのCAB55953.1(LBDG3)相同体についてのPSI-PRED出力が図22で示されている。顕著なことに、全5つの相同体もホモサピエンスレチノイン酸レセプターの3および5に対応するヘリックスを採用することが予測される(LBDG3-Cionaは“5”

50

ヘリックスのC末端について非常に弱い予測を示しているが、このことはPSI-PRED計算の間に多数のPSI-BLASTヒットとして取り上げることができないLBDG3-Ciona配列のより短い長さによって説明され得る）。全5つの相同体がCAB55953.1（LBDG3）のPSI-PRED予測と等価なPSI-PRED予測を有するという知見は、CAB55953.1（LBDG3）がホモサピエンスレチノイン酸レセプター リガンド結合ドメインの 3および 5に対応する2つの ヘリックスを採用するであろうという予測に意義を加える。このことは、CAB55953.1（LBDG3）のPSI-PRED出力をCAB55953.1（LBDG3）が核内ホルモンレセプターリガンド結合ドメインを含むというゲノムスレッダーの注釈付けの独立な確証としてさらに強める。

## 【0099】

CAB55953.1（LBDG3）が核内ホルモンレセプターリガンド結合ドメインを含むというゲノムスレッダーの注釈付けの更なるテストは、LBDフォールドを採用しているCAB55953.1（LBDG3）の相同モデルを構築する試みである。CAB55953.1（LBDG3）の相同モデルは、ホモサピエンスレイノインXレセプター 構造1LBDを有するCAB55953.1（LBDG3）のゲノムスレッダーアラインメントをモデラー（Modeller）バージョン5（Sali, A and Blundell T. L. Comparative protein modelling by satisfaction of spatial restraint. *J. Mol. Biol.* 234, 779-815, 1993）にかけることによって構築した。図23に相同モデルの全体図を提示する。CAB55953.1（LBDG3）は、1LBD構造上へモデルを作ることに成功しており、LBDフォールドを作り上げる交差格子(cross-latticed) ヘリックスの典型的な構成を示していることが理解され得る。図24は2つの主要な ヘリックス“ 3” および“ 5”（既知の核内ホルモンレセプターリガンド結合ドメイン構造における 3および 5に対応する）に焦点をあてている。（前記相同モデルにおける“ 3” および“ 5” ヘリックスは、図21において、PSI-PREDによりCAB55953.1（LBDG3）に対して予測された2つのヘリックスである。）図25は予測されるコアクチベーター結合部位についてズームインしている。既知のLBD構造では、コアクチベーター結合部位は、リガンド結合ドメイン表面上の 3および 5の間に形成された溝である。図26は同じ図であるが、前記の溝に結合しているコアクチベーターの予測される様式を図示するために加えられた相互作用コアクチベーターへリックスのアニメーションを有する図を示している。CAB55953.1（LBDG3）で予測される“ LBDモチーフ ”の4つの保存残基、ASP314、GLN315、LEU318およびLEU319は濃灰色で記され、それらは全てどのような衝突または立体障害もなしにその役割を果たすために適切な方向で存在する。

## 【0100】

例えばGLN315はコアクチベーターへリックスのC末端と相互作用するために適切な位置に存在し、このことはエストロゲンレセプター がSRC-1のコアクチベーターへリックスに結合する場合にエストロゲンレセプター の等価な残基GLN375について同じように観察される（A.K. Shiao, D. Barstad, P. M. Loria, Lin Cheng, P. J. Kushner, D. A. Agard, and G. L. Greene, The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen Cell 1998 95: 927）。同様に、LEU319は溝に突出してLXXLLコアクチベーターへリックスに存在する数個のロイシンと疎水性接触面を作っており、このことはエストロゲンレセプター がSRC-1のコアクチベーターへリックスに結合する場合にエストロゲンレセプター の等価な残基LEU379について同じように観察される（A.K. Shiao, D. Barstad, P. M. Loria, Lin Cheng, P. J. Kushner, D. A. Agard, and G. L. Greene, The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen Cell 1998 95: 927）。“ LBDモチーフ ”の外側の残基も、その役割を果たすのに適切な方向で配置されている。例えばARG307は溝に突出してコアクチベーターへリックスのC末端上で“ 電荷クランプ ”として作動し、このことはエストロゲンレセプター がSRC-1のコアクチベーターへリックスと結合する場合にエストロゲンレセプター の等価な残基LYS362について同じように観察される（A.K. Shiao, D. Barstad, P. M. Loria, Lin Cheng, P. J. Kushner, D. A. Agard, and G. L. Greene, The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen Cell 1998 95: 927）。

JP 2005-500010 A 2005.1.6

xifex Cell 1998 95: 927)。リガンド結合ドメインとしてのCAB55953.1(LBDG3)の相同モデリングは、CAB55953.1(LBDG3)がリガンド結合ドメインを含むというゲノムスレッダーの注釈付けを支持する。

【図面の簡単な説明】

【0 1 0 1】

【図1】バイオペンジウムのターゲットマイニングインターフェイスのフロントエンドである。データベースの検索は、PDBコード“1EXA:A”を用いて開始する。

【図2 A】1EXA:Aを用いた検索に対するインファーマチカゲノムスレッダーの結果の抜粋が示されている。矢印は、典型的な核内レセプターリガンド結合ドメインを有するホモサピエンスレチノイン酸レセプター-1を示している。

【図2 B】1EXA:Aを用いた検索に対するインファーマチカゲノムスレッダーの結果の抜粋が示されている。矢印はCAB55953.1(LBDG3)を示している。

【図2 C】1EXA:Aを用いた検索に対するフォワードPSI-BLASTの結果の完全なリストが示されている。CAB55953.1(LBDG3)は同定されない。

【図3】CAB55953.1(LBDG3)に対する重複配列表示の結果のページである。

【図4】CAB55953.1(LBDG3)に対するインタープロPFAM検索結果を示す。

【図5】CAB55953.1(LBDG3)に対するNCBIタンパク質レポートである。

【図6 A】バイオペンジウムデータベースのフロントエンドである。データベースの検索はクエリー配列としてCAB55953.1(LBDG3)を用いて開始する。

【図6 B】クエリー配列としてCAB55953.1(LBDG3)を用いた検索のインファーマチカゲノムスレッダーの結果の抜粋である。矢印は1EXA:Aを指している。

【図6 C】クエリー配列としてCAB55953.1(LBDG3)を用いて得られた逆最大化PSI-BLASTの結果の抜粋である。

【図7】CAB55953.1(LBDG3)および1EXA:AのAIEye配列アラインメントを示す。

【図8 A】R-3-フルオロ-4-[2-ヒドロキシ-2-(5,5,8,8-テトラメチル-5,6,7,8-テトラヒドロ-ナフタレン-2-イル)-アセチルアミノ]-安息香酸{略してBms270394と表記する}(394450)のホモサピエンスレチノイン酸レセプター-2のリガンド結合ドメイン、1EXA:Aとの相互作用部位を図示している、1EXA:Aに関するLigEyeである。

【図8 B】1EXA:A、ホモサピエンスレチノイン酸レセプター-2のリガンド結合ドメインのiRasMol図である。

【図9】CAB55953.1(LBDG3)、AAF36513.1、AAF36514.1およびAAF36515.1のAIEye配列アラインメントである。

【図10】DKFZP727M231についてのNCBIタンパク質レポートである。

【図11】DKFZP727M231についてのNCBI配列リストである。

【図12】AL117480をゲノムフラグメントAL13285.35で見出すことができたことを示している電子PCR結果である。

【図13】AL13285.35を20q11.21-12で見出すことができたことを示しているNCBIヌクレオチドレポートである。

【図14】20q11.2に自己免疫性甲状腺疾患(AITD)グレーブス病の感受性遺伝子をマッピングすることを示しているPubMed検索結果である。

【図15】20q11.2-q12の間のゲノム領域の増幅が種々の癌に対して関連していることを示しているPubMed検索結果である。

【図16】結腸cDNAにおける標的CAC14946.1(LBDG3)反応についての線形ダイナミックレンジである。

【図17】結腸cDNAにおける内部標準18s rRNA反応についての線形ダイナミックレンジである。

【図18】18の正常ヒト組織におけるCAC14946.1(LBDG3)の正規化した発現である。

【図19】CAB55953.1(LBDG3)を有するホモサピエンスレチノイン酸レセプター構造1EXA:Aのゲノムスレッダーアラインメントである。アラインメントに含まれるものは、CAB55953.1(LBDG3)の5つの相同体；マス・マスキュラスAAH03931.1、ダニオ・レリオAA

10

20

30

40

50

F36515.1およびAAF36514.1、チオナ LBDG3-Cionaおよびオイコブルーラ LBDG3-0ikoである。特徴的なリガンド結合ドメイン残基は、a-iで標識した黒い囲みによって記されている。

【図20】既知の核内ホルモンレセプターリガンド結合ドメイン構造の選択についてのHomstradレポートである。1LBD=ホモサピエンスレチノイドXレセプター、2LBD=ホモサピエンスレチノイン酸レセプター、1PRG:A=ホモサピエンスペルオキシソーム増殖因子活性化レセプター、3ERT:A=ホモサピエンスエストロゲンレセプター、および1A28:A=ホモサピエンスプロゲステロンレセプター。各残基の構造上の役割は、the Key to Homstradに記述した小文字、大文字、太字、下線又はイタリックによって記述される。

【図21】ホモサピエンスレチノイン酸レセプター 配列 (H.s. RAR 既知2<sup>0</sup>) およびCA B55953.1 (LBDG3) 配列 (LBDG3予測2<sup>0</sup>) についての、“LBDモチーフ”の領域におけるPSI-PRED二次構造予測プログラムの結果である。(その2つのPSI-PRED出力は、ホモサピエンスレチノイン酸レセプターを有するCAB55953.1 (LBDG3) のゲノムスレッダーラインメントを参照することによって、PHE305以降で整列化されている。)灰色の陰影はヘリックスの予測を示しており(カラムが高くなるにつれて、ねじれ予測の確信が高まる)、黒色の陰影は二次構造予測が行われていないことを示している。図の上段 (H.s. RAR 既知2<sup>0</sup>) は、ホモサピエンスレチノイン酸レセプター (1EXA:A) の結晶構造に由来する実際の既知二次構造を描いている。ヘリックス5において実験的に観察されたねじれは、矢印によって記されている。

【図22】ホモサピエンスレチノイン酸レセプター 配列 (H.サピエンス RAR)、CAB55953.1 (LBDG3) 配列 (H.サピエンス LBDG3) および5つのCAB55953.1 (LBDG3) 相同体(マス・マスキュラス AAH03931.1、ダニオ・レリオ AAF36515.1およびAAF36514.1、チオナ LBDG3-Cionaおよびオイコブルーラ LBDG3-0iko)についてのPSI-PRED二次構造予測プログラムの結果である。(図21におけるように、PSI-PRED出力は、ホモサピエンスレチノイン酸レセプターを有するCAB55953.1 (LBDG3) のゲノムスレッダーラインメントを参照することによって、PHE305以降で整列化されている。さらにRAR 残基GLY250-LEU254はスペースの制限のためにRAR 出力から切り落とされており、それはPR0249およびSER255の間の白線として記されている。)

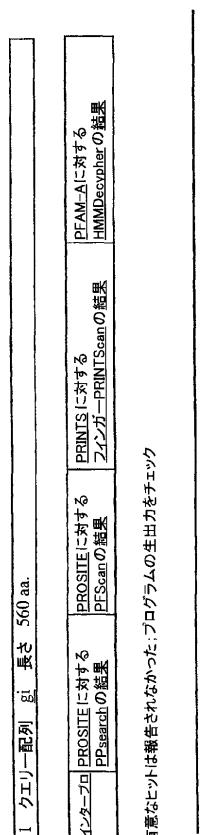
【図23】リガンド結合ドメインフォールドを採用しているCAB55953.1 (LBDG3) の相同モデルの全体図である。特に注目される残基は黒色で示されている。

【図24】リガンド結合ドメインフォールドを採用しているCAB55953.1 (LBDG3) の相同モデルの図であり、予測されるヘリックス“3”および“5”を囲む領域のみを示している。灰色の矢印はN-末端からC-末端に連続するポリペプチド鎖の方向を記す。

【図25】リガンド結合ドメインフォールドを採用しているCAB55953.1 (LBDG3) の相同体の予測されるコアクチベーター結合部位の拡大図である。特に注目される残基は黒色で示されている。

【図26】リガンド結合ドメインフォールドを採用しているCAB55953.1 (LBDG3) の相同体の予測されるコアクチベーター結合部位の拡大図である。特に注目される残基は黒色で示されている。コアクチベーターヘリックスのアニメーションは、相同モデルに対する予測結合様式を図示するために加えられている。

【 図 4 】



## インターネット検索結果

【 図 9 - 1 】

AI Eye 出力 (2月26日, 2001 5:04 PM)

16

CAB5951: MAAPAAQAG PYWGRRATAAAYVGRPBNILLQGRGIVRAVQFETTELTKDQSKYVPOV  
 ....MAATLQOOGVGGKTRNITRISQAGRFSOTQPELLEDRKAWHGIVPOV  
 ....MATRIGF SQTQTRKPCKNFMHKRGQVGEITRSPVQVALLERDKACOVGIVPOV  
 60 90 100 110 120 130 140 150  
 CAB5951: 100 110 120 130 140 150  
 CAB5951: 160 170 180 190 200 210 220 230  
 CAB5951: 240 250 260 270 280 290 300 310  
 CAB5951: 320 330 340 350 360 370 380 390  
 CAB5951: 400 410 420 430 440 450 460 470  
 CAB5951: 480 490 500 510 520 530 540 550  
 CAB5951: 560 570 580 590 600 610 620 630  
 CAB5951: 640 650 660 670 680 690 700 710  
 CAB5951: 720 730 740 750 760 770 780 790  
 CAB5951: 800 810 820 830 840 850 860 870  
 CAB5951: 880 890 900 910 920 930 940 950  
 CAB5951: 960 970 980 990 1000 1010 1020 1030  
 CAB5951: 1040 1050 1060 1070 1080 1090 1100 1110  
 CAB5951: 1120 1130 1140 1150 1160 1170 1180 1190  
 CAB5951: 1200 1210 1220 1230 1240 1250 1260 1270  
 CAB5951: 1280 1290 1300 1310 1320 1330 1340 1350  
 CAB5951: 1360 1370 1380 1390 1400 1410 1420 1430  
 CAB5951: 1440 1450 1460 1470 1480 1490 1500 1510  
 CAB5951: 1520 1530 1540 1550 1560 1570 1580 1590  
 CAB5951: 1600 1610 1620 1630 1640 1650 1660 1670  
 CAB5951: 1680 1690 1700 1710 1720 1730 1740 1750  
 CAB5951: 1760 1770 1780 1790 1800 1810 1820 1830  
 CAB5951: 1840 1850 1860 1870 1880 1890 1900 1910  
 CAB5951: 1920 1930 1940 1950 1960 1970 1980 1990  
 CAB5951: 2000 2010 2020 2030 2040 2050 2060 2070  
 CAB5951: 2080 2090 2100 2110 2120 2130 2140 2150  
 CAB5951: 2160 2170 2180 2190 2200 2210 2220 2230  
 CAB5951: 2240 2250 2260 2270 2280 2290 2300 2310  
 CAB5951: 2320 2330 2340 2350 2360 2370 2380 2390  
 CAB5951: 2400 2410 2420 2430 2440 2450 2460 2470  
 CAB5951: 2480 2490 2500 2510 2520 2530 2540 2550  
 CAB5951: 2560 2570 2580 2590 2600 2610 2620 2630  
 CAB5951: 2640 2650 2660 2670 2680 2690 2700 2710  
 CAB5951: 2720 2730 2740 2750 2760 2770 2780 2790  
 CAB5951: 2800 2810 2820 2830 2840 2850 2860 2870  
 CAB5951: 2880 2890 2900 2910 2920 2930 2940 2950  
 CAB5951: 2960 2970 2980 2990 3000 3010 3020 3030  
 CAB5951: 3040 3050 3060 3070 3080 3090 3100 3110  
 CAB5951: 3120 3130 3140 3150 3160 3170 3180 3190  
 CAB5951: 3200 3210 3220 3230 3240 3250 3260 3270  
 CAB5951: 3280 3290 3300 3310 3320 3330 3340 3350  
 CAB5951: 3360 3370 3380 3390 3400 3410 3420 3430  
 CAB5951: 3440 3450 3460 3470 3480 3490 3500 3510  
 CAB5951: 3520 3530 3540 3550 3560 3570 3580 3590  
 CAB5951: 3600 3610 3620 3630 3640 3650 3660 3670  
 CAB5951: 3680 3690 3700 3710 3720 3730 3740 3750  
 CAB5951: 3760 3770 3780 3790 3800 3810 3820 3830  
 CAB5951: 3840 3850 3860 3870 3880 3890 3900 3910  
 CAB5951: 3920 3930 3940 3950 3960 3970 3980 3990  
 CAB5951: 4000 4010 4020 4030 4040 4050 4060 4070  
 CAB5951: 4080 4090 4100 4110 4120 4130 4140 4150  
 CAB5951: 4160 4170 4180 4190 4200 4210 4220 4230  
 CAB5951: 4240 4250 4260 4270 4280 4290 4300 4310  
 CAB5951: 4320 4330 4340 4350 4360 4370 4380 4390  
 CAB5951: 4400 4410 4420 4430 4440 4450 4460 4470  
 CAB5951: 4480 4490 4500 4510 4520 4530 4540 4550  
 CAB5951: 4560 4570 4580 4590 4600 4610 4620 4630  
 CAB5951: 4640 4650 4660 4670 4680 4690 4700 4710  
 CAB5951: 4720 4730 4740 4750 4760 4770 4780 4790  
 CAB5951: 4800 4810 4820 4830 4840 4850 4860 4870  
 CAB5951: 4880 4890 4900 4910 4920 4930 4940 4950  
 CAB5951: 4960 4970 4980 4990 5000 5010 5020 5030  
 CAB5951: 5040 5050 5060 5070 5080 5090 5100 5110  
 CAB5951: 5120 5130 5140 5150 5160 5170 5180 5190  
 CAB5951: 5200 5210 5220 5230 5240 5250 5260 5270  
 CAB5951: 5280 5290 5300 5310 5320 5330 5340 5350  
 CAB5951: 5360 5370 5380 5390 5400 5410 5420 5430  
 CAB5951: 5440 5450 5460 5470 5480 5490 5500 5510  
 CAB5951: 5520 5530 5540 5550 5560 5570 5580 5590  
 CAB5951: 5600 5610 5620 5630 5640 5650 5660 5670  
 CAB5951: 5680 5690 5700 5710 5720 5730 5740 5750  
 CAB5951: 5760 5770 5780 5790 5800 5810 5820 5830  
 CAB5951: 5840 5850 5860 5870 5880 5890 5900 5910  
 CAB5951: 5920 5930 5940 5950 5960 5970 5980 5990  
 CAB5951: 6000 6010 6020 6030 6040 6050 6060 6070  
 CAB5951: 6080 6090 6100 6110 6120 6130 6140 6150  
 CAB5951: 6160 6170 6180 6190 6200 6210 6220 6230  
 CAB5951: 6240 6250 6260 6270 6280 6290 6300 6310  
 CAB5951: 6320 6330 6340 6350 6360 6370 6380 6390  
 CAB5951: 6400 6410 6420 6430 6440 6450 6460 6470  
 CAB5951: 6480 6490 6500 6510 6520 6530 6540 6550  
 CAB5951: 6560 6570 6580 6590 6600 6610 6620 6630  
 CAB5951: 6640 6650 6660 6670 6680 6690 6700 6710  
 CAB5951: 6720 6730 6740 6750 6760 6770 6780 6790  
 CAB5951: 6800 6810 6820 6830 6840 6850 6860 6870  
 CAB5951: 6880 6890 6900 6910 6920 6930 6940 6950  
 CAB5951: 6960 6970 6980 6990 7000 7010 7020 7030  
 CAB5951: 7040 7050 7060 7070 7080 7090 7100 7110  
 CAB5951: 7120 7130 7140 7150 7160 7170 7180 7190  
 CAB5951: 7200 7210 7220 7230 7240 7250 7260 7270  
 CAB5951: 7280 7290 7300 7310 7320 7330 7340 7350  
 CAB5951: 7360 7370 7380 7390 7400 7410 7420 7430  
 CAB5951: 7440 7450 7460 7470 7480 7490 7500 7510  
 CAB5951: 7520 7530 7540 7550 7560 7570 7580 7590  
 CAB5951: 7600 7610 7620 7630 7640 7650 7660 7670  
 CAB5951: 7680 7690 7700 7710 7720 7730 7740 7750  
 CAB5951: 7760 7770 7780 7790 7790 7800 7810 7820  
 CAB5951: 7830 7840 7850 7860 7870 7880 7890 7900  
 CAB5951: 7910 7920 7930 7940 7950 7960 7970 7980  
 CAB5951: 7990 8000 8010 8020 8030 8040 8050 8060  
 CAB5951: 8070 8080 8090 8090 8100 8110 8120 8130  
 CAB5951: 8140 8150 8160 8170 8180 8190 8190 8200  
 CAB5951: 8210 8220 8230 8240 8250 8260 8270 8280  
 CAB5951: 8290 8300 8310 8320 8330 8340 8350 8360  
 CAB5951: 8370 8380 8390 8390 8400 8410 8420 8430  
 CAB5951: 8440 8450 8460 8470 8480 8490 8490 8500  
 CAB5951: 8510 8520 8530 8540 8550 8560 8570 8580  
 CAB5951: 8590 8600 8610 8620 8630 8640 8650 8660  
 CAB5951: 8670 8680 8690 8690 8700 8710 8720 8730  
 CAB5951: 8740 8750 8760 8770 8780 8790 8790 8800  
 CAB5951: 8810 8820 8830 8840 8850 8860 8870 8880  
 CAB5951: 8890 8900 8910 8920 8930 8940 8950 8960  
 CAB5951: 8970 8980 8990 8990 9000 9010 9020 9030  
 CAB5951: 9040 9050 9060 9070 9080 9090 9090 9100  
 CAB5951: 9110 9120 9130 9140 9150 9160 9170 9180  
 CAB5951: 9190 9200 9210 9220 9230 9240 9250 9260  
 CAB5951: 9270 9280 9290 9290 9300 9310 9320 9330  
 CAB5951: 9340 9350 9360 9370 9380 9390 9390 9400  
 CAB5951: 9410 9420 9430 9440 9450 9460 9470 9480  
 CAB5951: 9490 9500 9510 9520 9530 9540 9550 9560  
 CAB5951: 9570 9580 9590 9590 9600 9610 9620 9630  
 CAB5951: 9640 9650 9660 9670 9680 9690 9690 9700  
 CAB5951: 9710 9720 9730 9740 9750 9760 9770 9780  
 CAB5951: 9790 9800 9810 9820 9830 9840 9850 9860  
 CAB5951: 9870 9880 9890 9890 9900 9910 9920 9930  
 CAB5951: 9940 9950 9960 9970 9980 9990 9990 10000

G. 9-2

Alleye 出力 (2月26日, 2001 3:04 PM)

【図 9 - 2】

AIEye 中国 (1月15日, 2001 2:18 PM)

〔 図 7 〕

EIG 7

1 クエリ一覧列	gi	長さ	560 aa.
イクターフロ	PROSITEに対する PSSearchの結果	PROSITEに対する PSScanの結果	PRINT フィン

XML/TXTフォーマット清  
アリケーションは正常に終了し

EIG 7

【図10】

FIG. 10

NCBI

UniGene

Homo sapiens  
DKFZP727M231 protein

PubMed Entrez BLAST OMIM Taxonomy Structure

Search Human for display as html Go

switch to text mode

ユニジーンクラスター Hs.168073 DKFZP727M231

ユニジーン DKFZP727M231 タンパク質

SEE ALSO

遺伝子リンク: 26133  
相同遺伝子: Hs.168073

選択したモデル生物タンパク質の類似性

生物、タンパク質および同一性バーセンテージ、並びに整列化領域の長さ

H.サビエンス: PIR:T17263-T17263 仮想タンパク質 100 % / 559 aa  
DKFZP727M231.1 ヒト

M.マスコット: PID:g7106862-AFI30458\_1 Trp4-閲覧 98 % / 559 aa  
タンパク質 TAPI

マッピング情報

染色体: 20  
UniSTS 全体: 1923  
UniSTS 全体: sISG4612 ゲノム コンテキスト: マップビュー  
UniSTS 全体: sISG3028

発現情報

cDNA 供給源: 大動脈、血液、脳、乳房、CNS、結腸、眼、生殖細胞、心臓、腎臓、肺、筋肉、鼻、卵巣、膀胱、上皮小体、末梢神経系、胎盤、前立腺、胃、精巢、子宮、全胚、正常羊膜、脛、乳房、正常乳房、頸部、結腸、正常結腸、デニスドラッサン(denis-drash)、眼、原生胎盤、頭部-頸部、腎臓、肺、肺腫瘍、リンパ、筋肉、神經腫瘍、卵巣、膀胱、胎盤、ビネット(pnet)、正常前立腺、皮膚、胸腺、ブルー(pooled)、甲状腺、子宮、子宮腫瘍、全血

ユニジーン オーガニズム

SAGE: タグマッピングに対する遺伝子

mRNA/ 遺伝子配列 (3)

【図14】

FIG. 14

The University of Chicago press

染色体20q11.2に対する新しいグレーブス病感受性遺伝子座マップ。自己免疫性甲状腺疾患の遺伝学のためのインターナショナル・コンソーシアム

Tomer Y, Barbesino G, Greenberg DA, Concepcion E, Davies TF

USAニューヨークのマウント・シナイ(Mount Sinai)医科大学、医学部、内分泌学および代謝部局  
yotomer@smplink.mssm.edu

自己免疫性甲状腺疾患(AITD)はグレーブス病(GD)および橋本甲状腺炎とい2つの関連疾患を含み、免疫調節の混乱により甲状腺にに対して免疫攻撃を生じる。AITDは多因子であり、遺伝学的に感受性である個人において発生する。しかししながら、その感受性についての遺伝的原因はまだ判明していない。近年我々は、AITDの感受性遺伝子を同定することを目的として、AITD患者の全ゲノム連鎖研究を開始した。我々は、高度に多型でありかつ高密度な間隔のマイクロサテライトマークーを用いて(マークー間の距離<10 cM)、50の多型(multiplex)データセット、複数世代のAITDアフィア(323の個人)について研究した。二点および多点バラメトリック法の使用(古典的LODスコア解析)によって、連鎖解析を行った。染色体20を研究する間に、我々はGDI(強度関連付けられた遺伝子座を染色体20q11.2に見出した。劣性の遺伝様式およびLODの浸透率が想定されるマークーD20S195では、3.2の最大二点LODスコアが得られた。最大ノンバラメトリックLODスコアは2.4(P=0.00043)であった。このスコアもマークーD20S195で得られた。多点連鎖解析により、マークーD20S195とD20S107との間の6 cMの間隔に対して、3.5の最大LODスコアが得られた。我々のサンプルでは、不均一性(heterogeneity)の証拠は存在しなかった。我々の見解では、これらの結果は連鎖についての強力な証拠を示しており、染色体20q11.2に重要なGD-感受性遺伝子の存在を示唆している。

PMID: 9837828, UI: 99057513

□12: *Cancer Res* 1998 Oct 1;58(19):4260-3 Related Articles, Books, LinkOut

Chromosomal amplification is associated with cisplatin resistance of human male germ cell tumors.

Rao PH, Houldsworth J, Palanisamy N, Murty VV, Reuter VE, Motzer RJ, Bost GJ, Chaganti RS

Laboratory of Cancer Genetics, Sloan-Kettering Institute for Cancer Research, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

Chemotherapy resistance of tumors is an important biological and clinical problem. Studies from many tumor types have indicated that resistance may be based on multiple genetic pathways. Human male germa cell

【図15】

FIG. 15

NCBI PubMed

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM

Search PubMed for 20q11+ Clear

Display Abstract Save Text Order Add to Clipboard

Entrez PubMed Show 20 Items 1-20 of 26 Page 1 of 2 Select page: 1 2

PubMed Services □1: *Cancer Genet Cytogenet* 2000 Nov;123(1):27-34 Related Articles, Books, LinkOut

ELSEVIER SCIENCE

GATEWAY TO FULLTEXT ARTICLES

比較ゲノムハイブリダイゼーションにより検出された、62の原発性腎癌における反復性染色体変化

Guan X, Fu S, Xie J, Fang Y, Sham JS, Du B, Zhou H, Lu S, Wang B, Liu Y, Liang Q, Li X, Du B, Ning X, Du J, Li P, Trent JM

香港大学、クイーン・メアリー病院、臨床腫瘍学部門  
Room 129, Professorial Block, Pokfulam Road, Hong Kong, China.

[Medline record in process]

62の原発性腎癌における反復性染色体変化を検出するために、比較ゲノムハイブリダイゼーション(CGH)を用いた。8q(31のケース、50%)、20q11.2-q12.1:最小増加(gain)領域を有する20q(29のケース、47%)、13q22に最小増加領域を有する13q(21のケース、34%)および3q(19のケース、31%)の増加を含む種々のランダムでない染色体変化が共通して観察された。最も高頻度で減少した領域には、18p(20のケース、37%)、17p(21のケース、33%)および1p(14のケース、23%)が含まれた。6のケースで高コピー数の増加(DNA配列増幅)が検出された。3つのケースで8q23-q24.2および20q11.2-q12の増幅が観察された。20qの増加および19pの減少は、それらの領域由来の対応する細胞質人工染色体(BAC)クローニングを行い、蛍光インサインハイブリダイゼーションによって確認した。本研究で同定された染色体領域の増加および減少は、腎の腫瘍形成に関連する候補領域を提供する。

PMID: 11120330, UI: 20570089

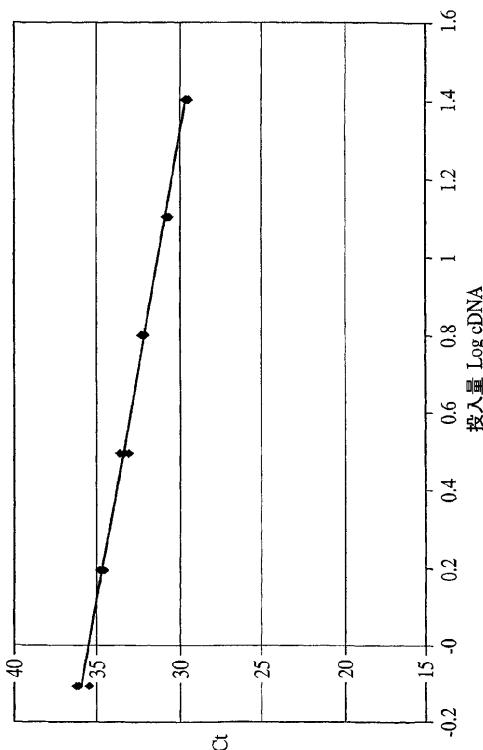
□2: *Mol Genet Metab* 2000 Sep-Oct;71(1-2):66-9 Related Articles, Books, LinkOut

IDEAL

Genetic determinants of graves disease.

【図16】

FIG. 16



【図17】

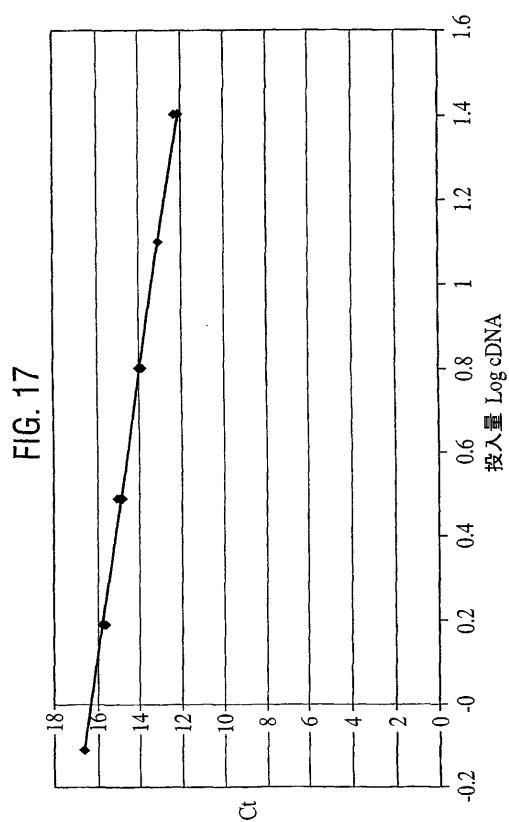
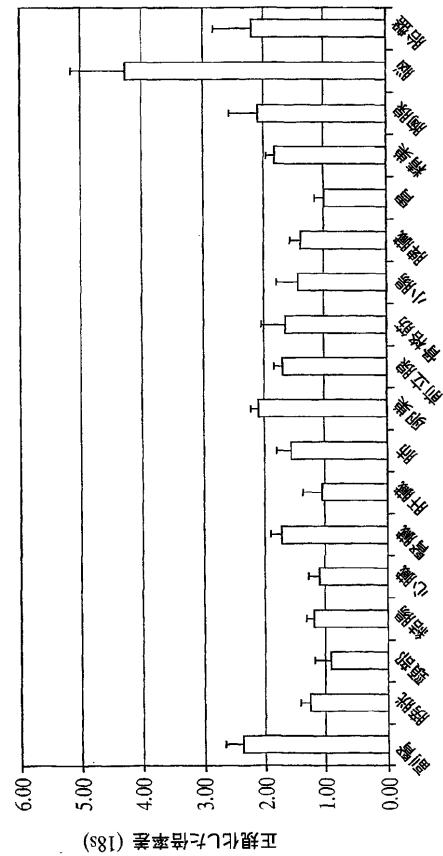


FIG. 17

【図18】



70組織cDNA

【図20-1】

FIG. 20-1

1lbd ( 225 ) sanedMpvverlEAEElavepk-----te  
 1lbd ( 182 ) lspqlezelitkvSH2hqeEPd-slcqlg-----kytn-----  
 1prga ( 207 ) esadlraltkhlygsY1kGp1tkaKs1s1Ltgktjdkspfvlydhs  
 3erta ( 305 ) 1algtadqMvsalidkpp-----ilysSeydp-----  
 1a28a ( 632 ) qlippl1n1lmpiedp-----viyaghKhn-----aaaaaaa  
 1lbd ( 248 ) tyyveannm1nppspn-----dptvhnicAadkq1ftLveWakr1P  
 2lbd ( 214 ) -----ssadhrvq-----Ldlqlwqkfselatkcik1kiveAkr1p  
 1prga ( 255 ) lmmGedk1kfk1h1p1qeqskevaiRifogccfrveavge1teAks1p  
 3erta ( 334 ) -----trpfs-----easmml1t1LadrelvhM1n1h1k1p  
 1a28a ( 706 ) -----tkpdt-----ssg1l1g1n1q1geryL1s1v1k1w1s1p  
 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa  
 1lbd ( 288 ) hFseLp1ldbgv1llr1gawneL11asFsh1S1a-vk-----dc11Latc-lhv  
 2lbd ( 250 ) gF1oLs1abq1t1k1aC1d11M1B1C1R-----ytpqgd1M1f1sd6-L1L  
 1prga ( 305 ) gF1v1ld1n1g1t1l1k1yG1H1l1Y1m1a1s1-----M1k1-dC1v1i1s1e1g1f1m  
 3erta ( 366 ) gF1v1ld1h1g1h1L1e1g1w1E1H1g1V1v1S1e1-hp-gk1l1f1A-----p11L1  
 1a28a ( 738 ) gF1n1L1h1D1g1t1q1s1m1s1L1m1V1F1g1G1v1h1k1v1s1G-----q1m1y1f1A-----p11L1  
 333 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa  
 bbb  
 1lbd ( 333 ) h1n1s1h1g1a-----g1v1s1P1d1v1l1e1v1s1k1m1d1M1k1TEL1G1L1R1a1v1L1p1p1d1s  
 2lbd ( 295 ) m1t1Q1M1A1-----M1t1Q1M1A1-----M1t1Q1M1A1-----y1t1p1q1d1M1f1s1d1s1-L1L  
 1prga ( 349 ) g1F1v1k1p1e1d1g1d1L1q1A1e1-----Q1k1L1h1p1-----e1s1g1f1a1k1L1  
 3erta ( 411 ) d1n1g1k1c1v1-----e1g1v1e1P1d1m1l1a1t1s1r1r1-----m1n1l1g1-e1F1v1L1k1s1i1L1n1g1v1  
 1a28a ( 785 ) neq1M1k1-e1s1f1s1c1f1m1y1P1e1F1-v1k1g1v1s1q1e1F1L1M1v1L1L1N1t1p1  
 baaa aaaaaaaaaaaaaaaaaaaaaaa  
 210 220 230 240 250  
 1lbd ( 381 ) k-----g1s1p1a1v1e1L1e1k1v1y1s1e1y1g1h1k1y1p1e1g1-----p1g1f1a1k1L1  
 2lbd ( 342 ) m-----d1e1e1p1k1d1g1p1l1e1A1r1i1y1A1r1r1p1-----s1g1y1M1p1p1M1  
 1prga ( 398 ) p-----g1L1n1v1k1p1e1d1g1d1L1q1A1e1-----Q1k1L1h1p1-----e1s1g1f1a1k1L1  
 3erta ( 459 ) y1t1f1s1t1k1g1e1k1h1v1L1d1k1L1d1T1i1h1M1a1k1y1t1q1q1h1q1r1a1L1  
 1a28a ( 832 ) 1-----e1g1r1s1t1g1e1r1m1s1s1Y1r1E1L1k1i1g1r1g1k1g1v1s1s1s1g1r1y1L1  
 aaaaaaaaaaaaaaaaaaaaaaa  
 260 270 280 290 300  
 1lbd ( 420 ) l1l1p1s1L1r1s1C1k1C1e1h1F1f1-----k1L1c1d1p1d1f1l1m1e1  
 2lbd ( 381 ) m1k1t1L1r1g1s1t1k1s1e1r1t1L1m1e1p1-----g1p1M1p1l1r1e1m1p1  
 1prga ( 437 ) g1k1t1d1r1q1i1t1e1h1g1l1q1v1k1t1e1d1-----m1s1h1p1L1c1e1y1k1L1  
 3erta ( 509 ) l1l1s1b1r1h1M1s1n1k1h1s1m1k1c1-----kn1v1-----p1y1d1l1e1M1d1-----  
 1a28a ( 875 ) k1l1d1h1l1v1k1q1h1i1y1c1n1f1q1s1r1a1s1v1e1p1m1e1v1s1a1q1p1k1l1g1m1  
 aaaaaaaaaaaaaaaaaaaaaaa  
 1lbd ( 458 ) ph---qmt  
 2lbd ( )  
 1prga ( )  
 3erta ( 546 ) --ahrlha  
 1a28a ( 925 ) V1k1l1l1f1k1

SUBSTITUTE SHEET (RULE 26)

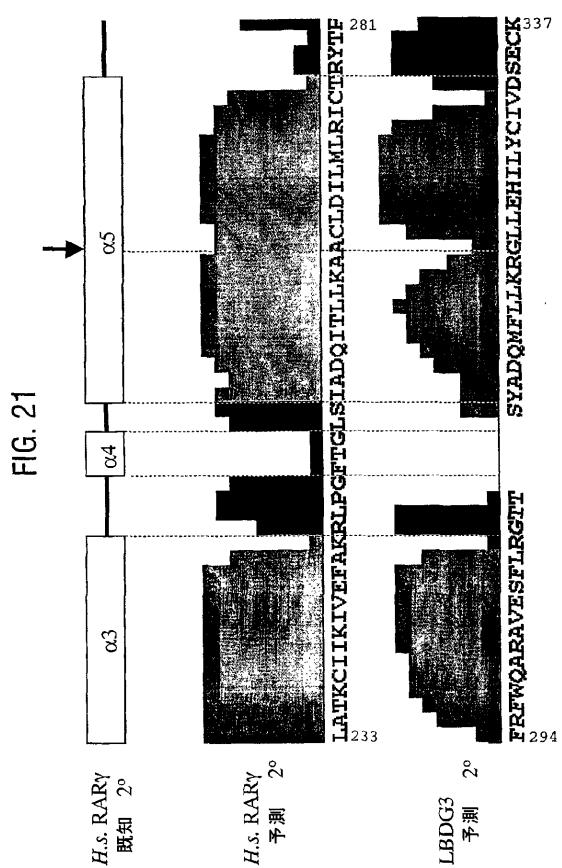
【図20-2】

FIG. 20-2

Key to Homstrand

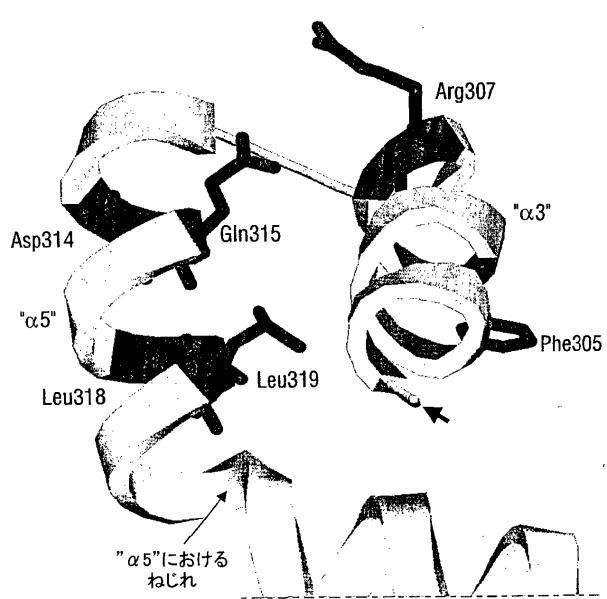
aaa  $\alpha$ ヘリックス  
 bbb  $\beta$ 鎖  
 333  $\beta_{10}$ ヘリックス  
 小文字 溶媒接近性  
 大文字 溶媒非接近性  
 太字 主鎖のアミドに対する水素結合  
 下線 主鎖のカルボニルに対する水素結合  
 イタリック 正のゆねじ角

【 図 2 1 】



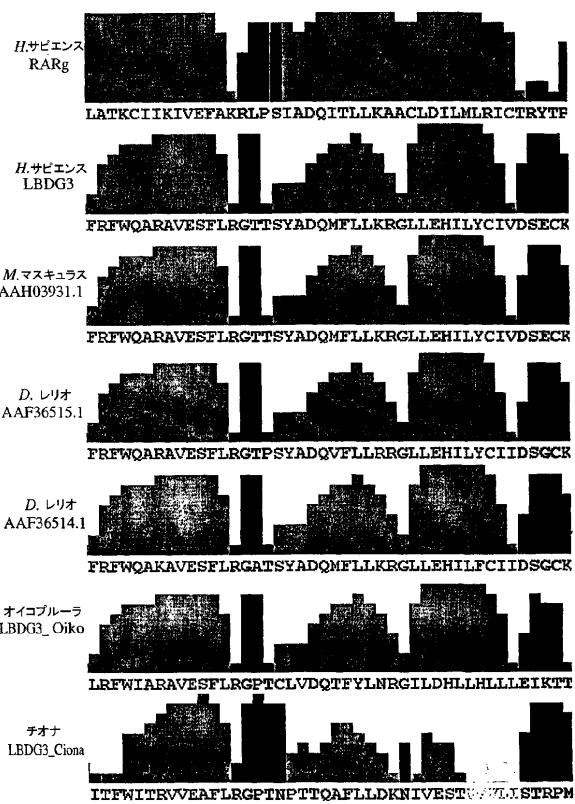
【 図 25 】

FIG. 25



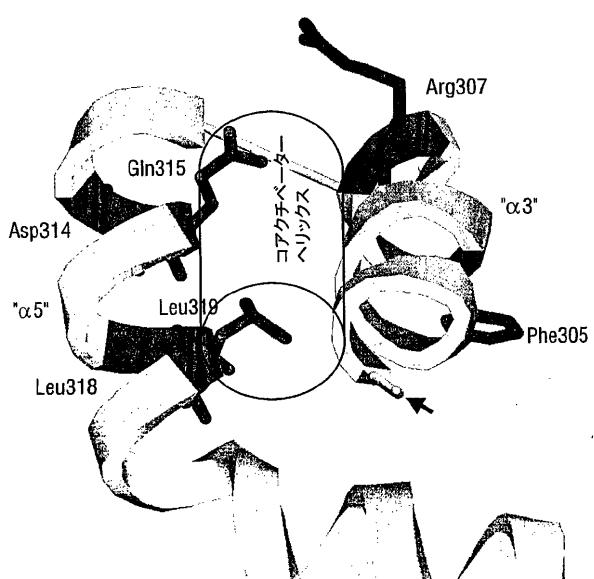
【 図 2 2 】

FIG. 22



【 図 2 6 】

FIG. 26



## 【国際公開パンフレット】

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

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number  
WO 02/070557 A2(51) International Patent Classification<sup>1</sup>: C07K 14/705 2NU (GB). POTTER, Sarah, Jane [GB/GB]; 39 Kneller Road, Brockley, London SE4 2AR (GB).

(21) International Application Number: PCT/GB02/00937

(22) International Filing Date: 5 March 2002 (05.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0105/02.2 5 March 2001 (05.03.2001) GB

(71) Applicant (for all designated States except US): INPHARMATICA LIMITED [GB/GB]; 60 Charlotte Street, London W1T 2NU (GB).

(72) Inventors: and

(75) Inventors/Applicants (for US only): FAGAN, Richard, Joseph [US/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). PHELPS, Christopher, Benjamin [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). PHILLIPS, Tom [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). PIERRON, Valerie, Nathalie [FR/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). ALLEN, Kathryn, Elizabeth [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). ALLEN, Janet, Marjorie [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T

(74) Agents: MERCER, Christopher, Paul et al.; Carpmaels &amp; Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CI, CN, CO, CR, CU, CZ, DL, DK, DM, DZ, EC, EI, ES, FI, GB, GD, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, L, I, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PL, PT, RO, RU, SD, SE, SG, SL, SK, SL, TJ, TM, TN, TR, TI, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SI, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:  
— without international search report and to be republished upon receipt of that 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 02/070557 A2

(54) Title: NUCLEAR HORMONE RECEPTOR LIGAND BINDING DOMAIN

(57) Abstract: This invention relates to a novel protein, termed LBDG3, herein identified as a Nuclear Hormone Receptor Ligand Binding Domain and to the use of this protein and nucleic acid sequence from the encoding genes in the diagnosis, prevention and treatment of disease.

WO 02/070557

PCT/GB02/00937

**NUCLEAR HORMONE RECEPTOR LIGAND BINDING DOMAIN**

This invention relates to a novel protein, termed CAB5593.1 herein identified as a Nuclear Hormone Receptor Ligand Binding Domain and to the use of this protein and nucleic acid sequence from the encoding gene in the diagnosis, prevention and treatment of disease.

All publications, patents and patent applications cited herein are incorporated in full by reference.

**BACKGROUND**

The process of drug discovery is presently undergoing a fundamental revolution as the era of functional genomics comes of age. The term "functional genomics" applies to an approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

Recently, a remarkable tool for the evaluation of sequences of unknown function has been developed by the Applicant for the present invention. This tool is a database system, termed the Biopendium search database, that is the subject of co-pending International Patent Application No. PCT/GB01/01105. This database system consists of an integrated data resource created using proprietary technology and containing information generated from an all-by-all comparison of all available protein or nucleic acid sequences.

The aim behind the integration of these sequence data from separate data resources is to

combine as much data as possible, relating both to the sequences themselves and to information relevant to each sequence, into one integrated resource. All the available data relating to each sequence, including data on the three-dimensional structure of the encoded protein, if this is available, are integrated together to make best use of the 5 information that is known about each sequence and thus to allow the most educated predictions to be made from comparisons of these sequences. The annotation that is generated in the database and which accompanies each sequence entry imparts a biologically relevant context to the sequence information.

This data resource has made possible the accurate prediction of protein function from 10 sequence alone. Using conventional technology, this is only possible for proteins that exhibit a high degree of sequence identity (above about 20%-30% identity) to other proteins in the same functional family. Accurate predictions are not possible for proteins that exhibit a very low degree of sequence homology to other related proteins of known function.

15 In the present case, a protein whose sequence is recorded in a publicly available database as CAB55953.1 (NCBI Genebank nucleotide accession number AL117480 and a Genebank protein accession number CAB55953.1), is implicated as a novel member of the Nuclear Hormone Receptor Ligand Binding Domain family.

20 **I. Introduction to Nuclear Hormone Receptor Ligand Binding Domains**

The Nuclear Hormone Receptor gene superfamily (see Table 1) encodes structurally related proteins that regulate the transcription of target genes. These proteins include receptors for steroid and thyroid hormones, vitamins, and other proteins for which no ligands have been found. Nuclear Receptors are composed of two key domains, a DNA- 25 Binding Domain (DBD) and a Ligand Binding Domain (LBD). The DBD directs the receptors to bind specific DNA sequences as monomers, homodimers, or heterodimers. The DBD is a particular type of zinc-finger, found only in Nuclear Receptors. Nuclear Receptors with DBDs can be readily identified at the sequence level by searching for matches to the PROSITE consensus sequence (PS00031).

The Ligand Binding Domain (LBD) binds and responds to the cognate hormone. Ligand binding to the LBD triggers a conformational change which expels a bound "Nuclear Receptor Co-Repressor". The site previously occupied by the Co-Repressor is then free to recruit a "Nuclear Receptor Co-Activator". This Ligand-triggered swap of a Co-Repressor for a Co-Activator is the mechanism by which Ligand binding leads to the transcriptional activation of target genes. All ligand binding domains contain a consensus sequence, the "LBD motif" (see Table 2) which mediates Co-Repressor and Co-Activator binding. The LBD is the binding site for all Nuclear Hormone Receptor targeted drugs to date and it is thus desirable to identify novel Ligand Binding Domains since these will be attractive drug targets. Ligand Binding Domains share low sequence identity (~15%) but have very similar structures and so present ideal targets for a structure-based relationship tool such as Genome Threader.

Many protein sequences have already been annotated in the public domain as Nuclear Hormone Receptors by their possession of DBDs using basic search tools like PROSITE, and their LBDs inferred on the basis of this. Because of this it is anticipated that any novel LBDs identified by Genome Threader *which are not annotated as nuclear receptors* will lack the DBD entirely. A precedent for a protein which has an LBD but lacks a DBD is provided by DAX1. Thus we annotate these DBD-less hits not as "Nuclear Hormone Receptors" but rather as containing a "Nuclear Hormone Receptor 20 Ligand Binding Domain".

Table 1: Nuclear hormone Receptor Superfamily

<i>Family: Steroid Hormone Receptors</i>	
Subfamilies	Glucocorticoid Receptors
	Progesterone Receptors
	Androgen Receptors
	Estrogen Receptors
<i>Family: Thyroid Hormone Receptor-like Factors</i>	
Subfamilies	Retinoic Acid Receptors (RARs)
	Retinoid X Receptors (RXRs)
	Thyroid Hormone Receptors
	Vitamin D Receptor
	NGFI-B
	FTZ-F1
	Peroxisome Proliferator Activated Receptors (PPARs)
	Ecdysone Receptors
	Retinoid Orphan Receptors (RORs)
	Tailless/COUP
	HNF-4
	CF1
	Knirps
<i>Family: DAX1</i>	
Subfamilies	DAX1

Table 2: The “LBD motif”. Numbers along the top row refer to residue position within the motif. Letters refer to amino acids by the 1-letter code. Letters within one column are all acceptable for that position within the motif. For example L, I, A, V, M, F, Y or W can occupy the first position of the “LBD motif”. Note that there is observed variation in the number of residues found between position 4 and 8, and position 9 and 12. The “LBD motif” was constructed by aligning 681 sequences of Nuclear Hormone Receptor Ligand Binding Domains, and identifying conserved patterns of residues.

1	2	3	4	5	6	7	8	9	10	11	12	13
L			L				D	Q			L	L
I			I				E	N			I	I
A			A					R			A	A
V	Any residues	2	V	Any 3 residues (or 2 residues or 4 residues)			H		Any residues (or 1 or 3 residues)	2	V	V
M			M					K			M	M
F			F					S			F	F
Y			Y					T			Y	Y
W			W								W	W

10

#### **II. Nuclear Hormone Receptors and Disease**

Nuclear Hormone Receptors have been shown to play a role in diverse physiological functions, many of which can play a role in disease processes (see Table 3).

Table 3. Nuclear Hormone Receptors and disease.

Nuclear Hormone Receptor	Disease
Androgen Receptor	Androgen Insensitivity Syndrome (Lubahn <i>et al.</i> 1989 Proc. Natl. Acad. Sci. USA 86, 9534-9538).
	Reifenstein syndrome (Wooster <i>et al.</i> 1992 Nat. Genet. 2, 132-134).
	X-linked recessive spinal and bulbar muscular atrophy (MacLean <i>et al.</i> 1995 Mol. Cell. Endocrinol. 112,133-141).
Glucocorticoid Receptor	Male breast cancer ((Wooster <i>et al.</i> 1992 Nat. Genet. 2, 132-134).
	Nelson's syndrome (Karl <i>et al.</i> 1996 J. Clin. Endocrinol. Metab. 81, 124-129).
Mineralocorticoid Receptor	Glucocorticoid resistant acute T-cell leukemia (Hala <i>et al.</i> 1996 Int. J. Cancer 68, 663-668).
	Pseudohypoaldosteronism (Chung <i>et al.</i> 1995 J. Clin. Endocrinol. Metab. 80, 3341-3345).
Estrogen Receptor alpha	ER alpha expression is elevated in a subset of human breast cancers. The application of Tamoxifen is the major therapy to prevent breast tumour progression. Unfortunately 35% of ER alpha positive breast cancers are Tamoxifen resistant (Petrangeli <i>et al.</i> 1994 J. Steroid Biochem. Mol. Biol. 49, 327-331).
Vitamin D3 Receptor	Mutations in the Vitamin D3 receptor produce a hereditary disorder similar in phenotype to Vitamin D3 deficiency (Rickets) (Hughes <i>et al.</i> 1988 Science 242, 1702-1725).
Retinoic Acid Receptor alpha	Acute Myeloid Leukemia (Lavau and Dejean 1994 Leukemia 8, 9-15).
Thyroid Hormone Receptor beta	"Generalised Resistance to Thyroid Hormones" (GRTH) (Refetoff 1994 Thyroid 4, 345-349).
DAX1	X-linked Adrenal Hypoplasia Congenita (AHC) and Hypogonadism (Ito <i>et al.</i> 1997 Mol. Cell. Biol. 17, 1476-1483).

Alteration of Nuclear Hormone Receptors by ligands which bind to their LBD thus provides a means to alter the disease phenotype. There is thus a great need for the identification of novel Nuclear Hormone Receptor Ligand Binding Domains, as these proteins may play a role in the diseases identified above, as well as in other disease states.

The identification of novel Nuclear Hormone Receptor Ligand Binding Domains is thus highly relevant for the treatment and diagnosis of disease, particularly those identified in Table 3.

#### THE INVENTION

- 5 The invention is based on the discovery that the CAB55953.1 protein functions as a Nuclear Hormone Receptor Ligand Binding Domain.

For the CAB55953.1 protein, it has been found that a region including residues 311-452 of this protein sequence adopts an equivalent fold to residues 74 (Ser255) to 216 (Ala297) of the Human Retinoic Acid Receptor gamma (PDB code 1EXA:A). Human Retinoic Acid Receptor gamma is known to function as a Nuclear Hormone Receptor Ligand Binding Domain. Furthermore, the "LBD motif" residues ASP258, GLN259, LEU262 and LEU263 of the Human Retinoic Acid Receptor gamma are conserved as ASP314, GLN315, LEU318 and LEU319 in CAB55953.1, respectively. This relationship is not just to Human Retinoic Acid Receptor gamma, but rather to the Nuclear Hormone Receptor Ligand Binding Domain family as a whole. Thus, by reference to the Genome Threader™ alignment of CAB55953.1 with the Human Retinoic Acid Receptor gamma (1EXA:A) ASP314, GLN315, LEU318 and LEU319 of CAB55953.1 are predicted to form the "LBD motif" residues.

20 The combination of equivalent fold and conservation of "LBD motif" residues allows the functional annotation of this region of CAB55953.1, and therefore proteins that include this region, as possessing Nuclear Hormone Receptor Ligand Binding Domain activity.

In one embodiment of the first aspect of the invention, there is provided a polypeptide, which polypeptide:

- 25 (i) comprises the amino acid sequence as recited in SEQ ID NO:2;
- (ii) is a fragment thereof having Nuclear Hormone Receptor Ligand Binding Domain activity or having an antigenic determinant in common with the polypeptides of (i); or
- (iii) is a functional equivalent of (i) or (ii).

The polypeptide having the sequence recited in SEQ ID NO:2 is referred to hereafter as

"the LBDG3 polypeptide".

The sequence with accession number CAB55953.1 has now been extended at the N terminus. The new sequence is referred to as CAC14946.1 and is 797 amino acids in length. This sequence is presented herein as SEQ ID NO:4. The new sequence does not, 5 of course, alter the annotation of the CAB55953.1 polypeptide sequence as having Nuclear Hormone Receptor Ligand Binding Domain activity, but merely extends the N terminal domain of the protein. The polypeptide having the sequence recited in SEQ ID NO:4 is referred to hereafter as "the LBDG3 full length polypeptide".

In a second embodiment of the first aspect of the invention, there is provided a 10 polypeptide, which polypeptide:

- (i) comprises the amino acid sequence as recited in SEQ ID NO:4;
- (ii) is a fragment thereof having Nuclear Hormone Receptor Ligand Binding Domain activity or having an antigenic determinant in common with the polypeptides of (i); or
- 15 (iii) is a functional equivalent of (i) or (ii).

Preferably, the polypeptide:

- (i) consists of the amino acid sequence as recited in SEQ ID NO:2 or SEQ ID NO:4;
- (ii) is a fragment thereof having Nuclear Hormone Receptor Ligand Binding Domain activity or having an antigenic determinant in common with the polypeptides of 20 (i); or
- (iii) is a functional equivalent of (i) or (ii).

According to this aspect of the invention, a preferred polypeptide fragment according to part ii) above includes the region of the LBDG3 polypeptide that is predicted as that responsible for Nuclear Hormone Receptor Ligand Binding Domain activity (hereafter, 25 the "LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region"), or is a variant thereof that possesses the "LBD motif" (ASP314, GLN315, LEU318 and LEU319, or equivalent residues). As defined herein, the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 311

and residue 452 of the LBDG3 polypeptide sequence.

This aspect of the invention also includes fusion proteins that incorporate polypeptide fragments and variants of these polypeptide fragments as defined above, provided that said fusion proteins possess activity as a Nuclear Hormone Receptor Ligand Binding

5 Domain.

In a second aspect, the invention provides a purified nucleic acid molecule that encodes a polypeptide of the first aspect of the invention. Preferably, the purified nucleic acid molecule has the nucleic acid sequence as recited in SEQ ID NO:1 (encoding the LBDG3 polypeptide) or SEQ ID NO:3 (encoding the LBDG3 full length polypeptide), or is a

10 redundant equivalent or fragment of this sequence. A preferred nucleic acid fragment is one that encodes a polypeptide fragment according to part ii) above, preferably a polypeptide fragment that includes the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region, or that encodes a variant of these fragments as this term is defined above.

15 In a third aspect, the invention provides a purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule of the second aspect of the invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that contains a nucleic acid molecule of the second or third aspect of the invention.

20 In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth aspect of the invention.

In a sixth aspect, the invention provides a ligand which binds specifically to, and which preferably inhibits the Nuclear Hormone Receptor Ligand Binding Domain activity of, a polypeptide of the first aspect of the invention.

25 In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the

polypeptide. Importantly, the identification of the function of the region defined herein as the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide, respectively, allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of diseases in which Nuclear Hormone Receptor Ligand Binding Domains are implicated.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the fifth aspect of the invention, or a compound of the sixth aspect of the invention, for use in therapy or diagnosis.

- 10 The inventors have discovered that the mRNA for LBDG3 is expressed at significant levels in the human brain. This is noteworthy as this provides a potential link to human disease states and development of agonists and antagonists for the ligand binding domain of LBDG3 and offers the potential for therapeutic intervention in various human diseases including, cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hypothyroidism, hyperparathyroidism, hypercalcemia, hypocalcemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.
- 15
- 20
- 25
- 30

The finding of a "non-classical" nuclear hormone receptor such as LBDG3 which contains a ligand binding domain in the absence of a DNA binding domain is consistent with the known literature which has consistently reported widespread effects of steroids in the brain (known as neurosteroids) and that these effects, in general, are mediated not through the known classic steroid hormone nuclear receptors which requires transcriptional activation. For instance, neurosteroids have been shown to influence neurotransmission particularly in the field of receptors such as those for GABA and NMDA and Sigma receptors. Neurosteroids have been shown to play a neuroprotective role. Therapeutic intervention through the development of agonists (or antagonists) to LBDG3 may therefore have a role in treatment of neurodegenerative conditions such as dementia, Parkinson's disease and neurodegeneration following cerebrovascular disease such as infarction or haemorrhage (stroke) and trauma to the central nervous system and spinal cord. In addition, neurosteroids have been shown to influence cognitive processing, spatial learning and memory, anxiety and behaviours such as craving which leads to addictive behaviour patterns. Development of agonists and antagonists to LBDG3 may therefore lead to therapeutic intervention to treat dementias, learning difficulties, anxiety, addictive behaviours such as but not exclusively alcoholism, eating disorders and drug addiction.

The inventors have shown that LBDG3 is not exclusively expressed in the brain. Significant levels of mRNA are also found in the adrenal, ovary, testis and thymus. The adrenal, ovary and testis are significant sites for the biosynthesis of steroids and their activity and is consistent but not exclusive with a role for LBDG3 in steroid actions. The finding of LBDG3 in these tissues supports the development of antagonists and agonists for treatment of diseases including but not exclusive to hypertension, responses to stress including stress of infectious diseases, regulation of salt and water homeostasis, control of fertility through regulation of ovulation (infertility and contraception), regulation of implantation (infertility and contraception) and regulation of spermatogenesis (infertility and contraception). In addition, these agents may be of value in treating steroid responsive tumours such as benign prostatic hypertrophy, prostatic cancer, ovarian cancer and testicular cancer.

- The finding of LBDG3 in the thymus is consistent with a role in T cell development. Agonists and antagonists developed to LBDG3 may therefore play a role in regulating T cells in disease processes such as autoimmune diseases and allergies including type I diabetes mellitus, rheumatoid arthritis, multiple sclerosis, psoriasis, renal failure arising from glomerulopathies, scleroderma, inflammatory bowel disease (both Crohns disease and ulcerative colitis), transplant rejection, asthma, atopic dermatitis, eczema, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.
- 10 In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of the first aspect of the invention or the activity of a polypeptide of the first aspect of the invention in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease. Such a method will preferably be carried out *in vitro*. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of expression or activity of a polypeptide or nucleic acid molecule over the period of time towards a control level is indicative of regression of disease.
- 20 A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.
- 25 A number of different such methods according to the ninth aspect of the invention exist, as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used on a short or long term basis to allow therapeutic treatment of a disease to be monitored in a patient. The invention also provides kits that are useful in these methods for diagnosing disease.
- 30 In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of

the invention as a Nuclear Hormone Receptor Ligand Binding Domain. The invention also provides for the use of a nucleic acid molecule according to the second or third aspects of the invention to express a protein that possesses Nuclear Hormone Receptor Ligand Binding Domain activity. The invention also provides a method for effecting 5 Nuclear Hormone Receptor Ligand Binding Domain activity, said method utilising a polypeptide of the first aspect of the invention.

In an eleventh aspect, the invention provides a pharmaceutical composition comprising a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a 10 ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the 15 invention, or a compound of the seventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or treatment of a disease, such as cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's 20 sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's 25 disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hypothyroidism, hyperparathyroidism, hypercalcemia, hypocalcemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological 30 disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic

infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an agonist. Conversely, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules, ribozymes and ligands, such as antibodies.

In a fourteenth aspect, the invention provides transgenic or knockout non-human animals that have been transformed to express higher, lower or absent levels of a polypeptide of the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be used in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.

A summary of standard techniques and procedures which may be employed in order to utilise the invention is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and it is not intended that this terminology should limit the scope of the present invention. The extent of the invention is limited only by the terms of the appended claims.

30 Standard abbreviations for nucleotides and amino acids are used in this specification.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of the those working in the art.

Such techniques are explained fully in the literature. Examples of particularly suitable 5 texts for consultation include the following: Sambrook Molecular Cloning: A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and 10 Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 & 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) 15 Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986).

As used herein, the term "polypeptide" includes any peptide or protein comprising two or 20 more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e. peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to longer chains (proteins).

The polypeptide of the present invention may be in the form of a mature protein or may 25 be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that is employed for purification of the mature polypeptide sequence.

The polypeptide of the first aspect of the invention may form part of a fusion protein. For 30 example, it is often advantageous to include one or more additional amino acid sequences which may contain secretory or leader sequences, pro-sequences, sequences which aid in purification, or sequences that confer higher protein stability, for example during

recombinant production. Alternatively or additionally, the mature polypeptide may be fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, 5 modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known modifications which may commonly be present in polypeptides of the present invention are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADP-ribosylation. Other potential 10 modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of 15 pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the 20 amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in naturally-occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention.

The modifications that occur in a polypeptide often will be a function of how the 25 polypeptide is made. For polypeptides that are made recombinantly, the nature and extent of the modifications in large part will be determined by the post-translational modification capacity of the particular host cell and the modification signals that are present in the amino acid sequence of the polypeptide in question. For instance, glycosylation patterns vary between different types of host cell.

30 The polypeptides of the present invention can be prepared in any suitable manner. Such

polypeptides include isolated naturally-occurring polypeptides (for example purified from cell culture), recombinantly-produced polypeptides (including fusion proteins), synthetically-produced polypeptides or polypeptides that are produced by a combination of these methods.

- 5 The functionally-equivalent polypeptides of the first aspect of the invention may be polypeptides that are homologous to the LBDG3 polypeptide or the LBDG3 full length polypeptide. Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity" indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome 10 Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heijne, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).
- 15 20 Homologous polypeptides therefore include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the LBDG3 polypeptide or the LBDG3 full length polypeptide. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. 25 30 Particularly preferred are variants in which several, i.e. between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination.

Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions.

Such mutants also include polypeptides in which one or more of the amino acid residues 5 includes a substituent group.

Typically, greater than 80% identity between two polypeptides (preferably, over a specified region) is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the LBDG3 polypeptide or the LBDG3 full length polypeptide, or 10 with active fragments thereof, of greater than 80%. More preferred polypeptides have degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, respectively with the LBDG3 polypeptide or the LBDG3 full length polypeptide, or with active fragments thereof.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 15 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

In the present case, preferred active fragments of the LBDG3 polypeptide or the LBDG3 full length polypeptide are those that include the LBDG3 Nuclear Hormone Receptor 20 Ligand Binding Domain region and which possess the "LBD motif" of residues ASP314, GLN315, LEU318 and LEU319, or equivalent residues (in the LBDG3 full length polypeptide, the relevant residues are ASP551, GLN552, LEU555 and LEU556). By "equivalent residues" is meant residues that are equivalent to the "LBD motif" residues, provided that the Nuclear Hormone Receptor Ligand Binding Domain region retains 25 activity as a Nuclear Hormone Receptor Ligand Binding Domain. For example ASP314 may be replaced by GLU. For example GLN315 may be replaced by ASN, ARG, HIS, LYS, SER or THR. For example LEU318 may be replaced by ILE, ALA, VAL, MET, PHE, TYR, or TRP. For example LEU319 may be replaced by ILE, ALA, VAL, MET, PHE, TYR, or TRP. Accordingly, this aspect of the invention includes polypeptides that 30 have degrees of identity of greater than 80%, preferably, greater than 85%, 90%, 95%,

98% or 99%, respectively, with the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide and which possess the "LBD motif" of ASP314, GLN315, LEU318 and LEU319, or equivalent residues. As discussed above, the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend 5 between residue 311 and residue 452 of the LBDG3 polypeptide sequence. In the LBDG3 full length polypeptide, the relevant boundaries are amino acid residues 548 and 689.

The functionally-equivalent polypeptides of the first aspect of the invention may also be polypeptides which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome Threader<sup>TM</sup> technology that forms one 10 aspect of the search tools used to generate the Biopendium search database may be used (see co-pending International patent application PCT/GB01/01105) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the LBDG3 polypeptide, are predicted to have Nuclear Hormone Receptor Ligand Binding Domain activity, by virtue of sharing significant structural 15 homology with the LBDG3 polypeptide sequence.

By "significant structural homology" is meant that the Inpharmatica Genome Threader<sup>TM</sup> predicts two proteins, or protein regions, to share structural homology with a certainty of at least 10% more preferably, at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and above. The certainty value of the Inpharmatica Genome Threader<sup>TM</sup> is calculated as 20 follows. A set of comparisons was initially performed using the Inpharmatica Genome Threader<sup>TM</sup> exclusively using sequences of known structure. Some of the comparisons were between proteins that were known to be related (on the basis of structure). A neural network was then trained on the basis that it needed to best distinguish between the known relationships and known not-relationships taken from the CATH structure 25 classification ([www.biochem.ncl.ac.uk/bsm/cath](http://www.biochem.ncl.ac.uk/bsm/cath)). This resulted in a neural network score between 0 and 1. However, again as the number of proteins that are related and the number that are unrelated were known, it was possible to partition the neural network results into packets and calculate empirically the percentage of the results that were correct. In this manner, any genuine prediction in the Biopendium search database has an 30 attached neural network score and the percentage confidence is a reflection of how successful the Inpharmatica Genome Threader<sup>TM</sup> was in the training/testing set.

Structural homologues of LBDG3 should share structural homology with the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region and possess the "LBD motif" residues ASP314, GLN315, LEU318 and LEU319, or equivalent residues (ASP551, GLN552, LEU555 and LEU556 in the LBDG3 full length polypeptide). Such structural homologues are predicted to have Nuclear Hormone Receptor Ligand Binding Domain activity by virtue of sharing significant structural homology with this polypeptide sequence and possessing the "LBD motif" residues.

The polypeptides of the first aspect of the invention also include fragments of the LBDG3 polypeptide, functional equivalents of the fragments of the LBDG3 polypeptide, and fragments of the functional equivalents of the LBDG3 polypeptides, provided that those functional equivalents and fragments retain Nuclear Hormone Receptor Ligand Binding Domain activity or have an antigenic determinant in common with the LBDG3 polypeptide or the LBDG3 full length polypeptide.

As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the LBDG3 polypeptides or one of its functional equivalents. The fragments should comprise at least n consecutive amino acids from the sequence and, depending on the particular sequence, n preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments may form an antigenic determinant.

Preferred polypeptide fragments according to this aspect of the invention are fragments that include a region defined herein as the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptides, respectively. These regions are the regions that have been annotated as Nuclear Hormone Receptor Ligand Binding Domain.

For the LBDG3 polypeptide, this region is considered to extend between residue 311 and residue 452 (residues 548-689 in the LBDG3 full length polypeptide).

Variants of this fragment are included as embodiments of this aspect of the invention, provided that these variants possess activity as a Nuclear Hormone Receptor Ligand Binding Domain.

In one respect, the term "variant" is meant to include extended or truncated versions of this polypeptide fragment.

For extended variants, it is considered highly likely that the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide will fold correctly and show 5 Nuclear Hormone Receptor Ligand Binding Domain activity if additional residues C terminal and/or N terminal of these boundaries in the LBDG3 polypeptide sequence are included in the polypeptide fragment. For example, an additional 5, 10, 20, 30, 40 or even 10 50 or more amino acid residues from the LBDG3 polypeptide sequence, or from a homologous sequence, may be included at either or both the C terminal and/or N terminal of the boundaries of the Nuclear Hormone Receptor Ligand Binding Domain regions of the LBDG3 polypeptide, without prejudicing the ability of the polypeptide fragment to fold correctly and exhibit Nuclear Hormone Receptor Ligand Binding Domain activity.

For truncated variants of the LBDG3 polypeptide, one or more amino acid residues may be deleted at either or both the C terminus or the N terminus of the Nuclear Hormone 15 Receptor Ligand Binding Domain region of the LBDG3 polypeptide, although the "LBD motif" residues (ASP314, GLN315, LEU318 and LEU319; ASP551, GLN552, LEU555 and LEU556 in the LBDG3 full length polypeptide), or equivalent residues should be maintained intact; deletions should not extend so far into the polypeptide sequence that any of these residues are deleted.

20 In a second respect, the term "variant" includes homologues of the polypeptide fragments described above, that possess significant sequence homology with the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide or the LBDG3 full length polypeptide and which possess the "LBD motif" residues (ASP314, GLN315, LEU318 and LEU319 in SEQ ID NO:2), or equivalent residues, provided that said 25 variants retain activity as an Nuclear Hormone Receptor Ligand Binding Domain.

Homologues include those polypeptide molecules that possess greater than 80% identity with the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain regions, of the LBDG3 polypeptides, respectively. Percentage identity is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for 30 Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open

penalty=11 and gap extension penalty=1]. Preferably, variant homologues of polypeptide fragments of this aspect of the invention have a degree of sequence identity with the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain regions, of the LBDG3 polypeptides, respectively, of greater than 80%. More preferred variant polypeptides have 5 degrees of identity of greater than 85%, 90%, 95%, 98% or 99%, respectively with the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain regions of the LBDG3, polypeptides, provided that said variants retain activity as a Nuclear Hormone Receptor Ligand Binding Domain. Variant polypeptides also include homologues of the truncated forms of the polypeptide fragments discussed above, provided that said variants retain 10 activity as a Nuclear Hormone Receptor Ligand Binding Domain.

The polypeptide fragments of the first aspect of the invention may be polypeptide fragments that exhibit significant structural homology with the structure of the polypeptide fragment defined by the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain regions, of the LBDG3 polypeptide sequence, for example, as identified by the 15 Inpharmatica Genome Threader™. Accordingly, polypeptide fragments that are structural homologues of the polypeptide fragments defined by the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain regions of the LBDG3 polypeptide sequence should adopt the same fold as that adopted by this polypeptide fragment, as this fold is defined above.

20 Structural homologues of the polypeptide fragment defined by the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region should also retain the "LBD motif" residues ASP314, GLN315, LEU318 and LEU319, or equivalent residues (ASP551, GLN552, LEU555 and LEU556 in the LBDG3 full length polypeptide).

Such fragments may be "free-standing", i.e. not part of or fused to other amino acids or 25 polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre- and/or pro- polypeptide region fused to the amino terminus of the fragment and/or an additional region fused to the carboxyl

terminus of the fragment. However, several fragments may be comprised within a single larger polypeptide.

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or 5 monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

10 The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')2 and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the 15 first aspect of the invention.

If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA 20 technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

25 Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, 77-96 in *Monoclonal Antibodies and Cancer Therapy*, 30 Alan R. Liss, Inc. (1985).

Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, i.e., for isotype, epitope, affinity, etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the 5 monoclonal antibodies of interest may be isolated from hybridomas, for instance by PCR techniques known in the art, and cloned and expressed in appropriate vectors.

Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, for example, Liu *et al.*, Proc. Natl. Acad. Sci. USA, 84, 3439 (1987)), may also be of use.

10 The antibody may be modified to make it less immunogenic in an individual, for example by humanisation (see Jones *et al.*, Nature, 321, 522 (1986); Verhoeven *et al.*, Science, 239: 1534 (1988); Kabat *et al.*, J. Immunol., 147: 1709 (1991); Queen *et al.*, Proc. Natl Acad. Sci. USA, 86, 10029 (1989); Gorman *et al.*, Proc. Natl Acad. Sci. USA, 88: 34181 (1991); and Hodgson *et al.*, Bio/Technology 9: 421 (1991)). The term "humanised 15 antibody", as used herein, refers to antibody molecules in which the CDR amino acids and selected other amino acids in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody. The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

20 In a further alternative, the antibody may be a "bispecific" antibody, that is an antibody having two different antigen binding domains, each domain being directed against a different epitope.

Phage display technology may be utilised to select genes which encode antibodies with 25 binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. *et al.*, (1990), Nature 348, 552-554; Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) Nature 352, 624-628).

30 Antibodies generated by the above techniques, whether polyclonal or monoclonal, have

additional utility in that they may be employed as reagents in immunoassays, radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these applications, the antibodies can be labelled with an analytically-detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

- 5 Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode the polypeptide sequences recited in SEQ ID NO:2 or SEQ ID NO:4, and functionally equivalent polypeptides, including active fragments of the LBDG3 polypeptide, such as a fragment including the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide sequence, or a homologue 10 thereof.

Nucleic acid molecules encompassing these stretches of sequence form a preferred embodiment of this aspect of the invention.

- These nucleic acid molecules may be used in the methods and applications described herein. The nucleic acid molecules of the invention preferably comprise at least n 15 consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

- The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or 20 probing purposes).

- Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic DNA. Such nucleic acid molecules may be obtained by cloning, by chemical 25 synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for example, by chemical synthesis using techniques such as solid phase phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation from an organism. RNA molecules may generally be generated by the *in vitro* or *in vivo* transcription of DNA sequences.

The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded

DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", 5 as used herein, refers to an antisense molecule or an anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they preferentially bind complementary single stranded DNA and RNA and stop 10 transcript elongation (Nielsen, P.E. *et al.* (1993) *Anticancer Drug Des.* 8:53-63).

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO:2 or SEQ ID NO:4, or an active fragment thereof, may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO:1 or SEQ ID NO:3, respectively. These molecules also may have a different sequence which, as a result of the degeneracy of the 15 genetic code, encodes the polypeptide SEQ ID NO:2 or SEQ ID NO:4, or an active fragment of the LBDG3 polypeptide, such as a fragment including the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region is considered to extend between residue 311 and residue 452 of the LBDG3 polypeptide sequence. In the LBDG3 20 full length polypeptide, the relevant boundaries are residues 548 and 689. In SEQ ID NO:1 the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region is thus encoded by a nucleic acid molecule including nucleotide 932 to 1357. In SEQ ID NO:3, these boundaries are 1642 and 2067. Nucleic acid molecules encompassing this stretch of sequence, and homologues of this sequence, form a preferred embodiment of this aspect 25 of the invention.

Such nucleic acid molecules that encode the polypeptide of SEQ ID NO:2 or SEQ ID NO:4 may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pro-, pre- or 30 prepro- polypeptide sequence; the coding sequence of the mature polypeptide, with or

without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules may also include 5 additional sequences which encode additional amino acids, such as those which provide additional functionalities.

The nucleic acid molecules of the second and third aspects of the invention may also encode the fragments or the functional equivalents of the polypeptides and fragments of the first aspect of the invention.

- 10 As discussed above, a preferred fragment of the LBDG3 polypeptide is a fragment including the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region, or a homologue thereof. The Nuclear Hormone Receptor Ligand Binding Domain region is encoded by a nucleic acid molecule including nucleotides 932 to 1357 of SEQ ID NO:1 (in SEQ ID NO:3, these boundaries are 1642 and 2067).
- 15 Functionally equivalent nucleic acid molecules according to the invention may be naturally-occurring variants such as a naturally-occurring allelic variant, or the molecules may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or organisms.
- 20 Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or insertions.
- 25 The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides are included as techniques which may be used to engineer the nucleotide sequences. Site-directed mutagenesis may be used to insert new restriction sites, alter 30

glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

Nucleic acid molecules which encode a polypeptide of the first aspect of the invention may be ligated to a heterologous sequence so that the combined nucleic acid molecule 5 encodes a fusion protein. Such combined nucleic acid molecules are included within the second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the activity of the polypeptide, it may be useful to express, using such a combined nucleic acid molecule, a fusion protein that can be recognised by a commercially-available antibody. A fusion protein may also be engineered to contain a 10 cleavage site located between the sequence of the polypeptide of the invention and the sequence of a heterologous protein so that the polypeptide may be cleaved and purified away from the heterologous protein.

The nucleic acid molecules of the invention also include antisense molecules that are 15 partially complementary to nucleic acid molecules encoding polypeptides of the present invention and that therefore hybridize to the encoding nucleic acid molecules (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a polypeptide of the invention, as will be known by those of ordinary skill in the art (see, for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 20 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee *et al.*, Nucleic Acids Res 6, 3073 (1979); Cooney *et al.*, Science 241, 456 (1988); Dervan *et al.*, Science 251, 1360 (1991).

The term "hybridization" as used here refers to the association of two nucleic acid 25 molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment 30 of the liquid phase molecule to the solid support (Denhardt's reagent or BLOTTO); the concentration of the molecules; use of compounds to increase the rate of association of

molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook *et al.* [*supra*]).

The inhibition of hybridization of a completely complementary molecule to a target molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook *et al* [*supra*]). A substantially homologous molecule will then compete for and inhibit the binding of a completely homologous molecule to the target molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; *Methods Enzymol.* 152:399-407) and Kimmel, A.R. (1987; *Methods Enzymol.* 152:507-511).

"Stringency" refers to conditions in a hybridization reaction that favour the association of very similar molecules over association of molecules that differ. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook *et al.* [*supra*]). Preferably, the conditions used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 80% identical over their entire length to a nucleic acid molecule encoding the LBDG3 polypeptide (SEQ ID NO:2) or the LBDG3 full length polypeptide (SEQ ID NO:4), and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. A preferred active fragment is a fragment that includes an LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide sequences, respectively. Accordingly, preferred nucleic acid molecules include those that are at least 80% identical over their entire length to a nucleic acid molecule encoding the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide sequence.

Percentage identity, as referred to herein, is as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for

Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>).

Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO:1, to a region including nucleotides 932-1357

5 of this sequence, or a nucleic acid molecule that is complementary to any one of these regions of nucleic acid. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides which retain substantially the same biological

10 function or activity as the LBDG3 polypeptide.

The invention also provides a process for detecting a nucleic acid molecule of the invention, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting any such duplexes that are formed.

15 As discussed additionally below in connection with assays that may be utilised according to the invention, a nucleic acid molecule as described above may be used as a hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the LBDG3 polypeptide or the LBDG3 full length polypeptide and to isolate cDNA and genomic clones of homologous or orthologous 20 genes that have a high sequence similarity to the gene encoding this polypeptide.

In this regard, the following techniques, among others known in the art, may be utilised and are discussed below for purposes of illustration. Methods for DNA sequencing and analysis are well known and are generally available in the art and may, indeed, be used to practice many of the embodiments of the invention discussed herein. Such methods may 25 employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated 30 using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier

Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an equivalent function to that of the LBDG3 polypeptide, particularly with an equivalent 5 function to the LBDG3 Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide, is to probe a genomic or cDNA library with a natural or artificially-designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel *et al.* (eds). Greene Publishing Association and John Wiley Interscience, New York, 1989,1992). Probes comprising at 10 least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO:1), particularly a region from nucleotides 932-1357 of SEQ ID NO:1, are particularly useful probes.

Such probes may be labelled with an analytically-detectable reagent to facilitate their 15 identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation of a detectable product. Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding 20 proteins of interest from human, mammalian or other animal sources and screening such sources for related sequences, for example, for additional members of the family, type and/or subtype.

In many cases, isolated cDNA sequences will be incomplete, in that the region encoding the polypeptide will be cut short, normally at the 5' end. Several methods are available to 25 obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended utilising a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed is based on the method of Rapid Amplification of cDNA Ends (RACE; see, for example, Frohman *et al.*, Proc. Natl. Acad. Sci. USA (1988) 85: 8998-9002). Recent modifications of this technique, exemplified by the Marathon<sup>TM</sup> 30 technology (Clontech Laboratories Inc.), for example, have significantly simplified the

search for longer cDNAs. A slightly different technique, termed "restriction-site" PCR, uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T., *et al.* (1988) Nucleic Acids Res. 16:8186). Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. *et al.* (1991) PCR Methods Applic. 1: 111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. *et al.* (1991); Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important step in the confirmatory correlation of those sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene

discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to 5 translocation, inversion, etc. among normal, carrier, or affected individuals.

The nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the polypeptide in tissues by detection of the mRNAs that encode them. These techniques include *in situ* hybridization techniques and nucleotide amplification techniques, such as 10 PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

15 The vectors of the present invention comprise nucleic acid molecules of the invention and may be cloning or expression vectors. The host cells of the invention, which may be transformed, transfected or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

The polypeptides of the invention may be prepared in recombinant form by expression of 20 their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in detail by Sambrook *et al. (supra)* and Fernandez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto).

25 Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those described in Sambrook *et al.*, (supra). Generally, the encoding gene can be placed under the control of 30 a control element such as a promoter, ribosome binding site (for bacterial expression)

and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide is transcribed into RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial 5 plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may 10 also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

Particularly suitable expression systems include microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected 15 with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (for example, Ti or pBR322 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

20 Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory manuals, such as Davis *et al.*, *Basic Methods in Molecular Biology* (1986) and Sambrook *et al.*, (*supra*). Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-25 mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook *et al.*, 1989 [*supra*]; Ausubel *et al.*, 1991 [*supra*]; Spector, Goldman & Leinwald, 1998). In eukaryotic cells, expression systems may either be transient (for example, episomal) or permanent (chromosomal integration) according to the needs of the system.

30 The encoding nucleic acid molecule may or may not include a sequence encoding a

control sequence, such as a signal peptide or leader sequence, as desired, for example, for secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences 5 can be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory sequences are those which cause the expression of a gene to be increased or decreased in response to a chemical or physical stimulus, including the 10 presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to carry out transcription and translation. Such regulatory sequences may vary in their 15 strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSport<sup>TM</sup> plasmid (Gibco BRL) and the like may be used. The 20 baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (for example, heat shock, RUBISCO and storage protein genes) or from plant viruses (for example, viral promoters or leader sequences) 25 may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, 30 i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the

sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can 5 be cloned directly into an expression vector that already contains the control sequences and an appropriate restriction site.

For long-term, high-yield production of a recombinant polypeptide, stable expression is preferred. For example, cell lines which stably express the polypeptide of interest may be transformed using expression vectors which may contain viral origins of replication 10 and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. 15 Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby 20 hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, inter alia, Invitrogen, San Diego CA (the 25 "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells.

There are many plant cell culture and whole plant genetic expression systems known in 30 the art. Examples of suitable plant cellular genetic expression systems include those

described in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, (1991) *Phytochemistry* 30, 3861-3863.

In particular, all plants from which protoplasts can be isolated and cultured to give whole 5 regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include streptococci, 10 staphylococci, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells.

Examples of particularly suitable host cells for fungal expression include yeast cells (for example, *S. cerevisiae*) and *Aspergillus* cells.

Any number of selection systems are known in the art that may be used to recover 15 transformed cell lines. Examples include the herpes simplex virus thymidine kinase (Wigler, M. *et al.* (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. *et al.* (1980) *Cell* 22:817-23) genes that can be employed in tk- or aprt $\pm$  cells, respectively.

Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dihydrofolate reductase (DHFR) that confers resistance to 20 methotrexate (Wigler, M. *et al.* (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al.* (1981) *J. Mol. Biol.* 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

25 Although the presence or absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the relevant sequence is inserted within a marker gene sequence, transformed cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence

encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. *et al.* (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.* (1983) *J. Exp. Med.*, 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridization or PCR probes for detecting sequences related to nucleic acid molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesise RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. 25 Biochemical Corp., Cleveland, OH)).

Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create 30 transgenic animals, particularly rodent animals. Such transgenic animals form a further

aspect of the present invention. This may be done locally by modification of somatic cells, or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

- 5 The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is
- 10 particularly useful for purification. Well known techniques for refolding proteins may be employed to regenerate an active conformation when the polypeptide is denatured during isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Examples of such purification-facilitating domains include metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the polypeptide of the invention may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. *et al.* (1992) *Prot. Exp. Purif.* 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. *et al.* (*DNA Cell Biol.* 25: 199312:441-453).

If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity 5 techniques. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a 10 variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

15 Agonist or antagonist compounds may be isolated from, for example, cells, cell-free preparations, chemical libraries or natural product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991).

20 Compounds that are most likely to be good antagonists are molecules that bind to the polypeptide of the invention without inducing the biological effects of the polypeptide upon binding to it. Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby 25 inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be inhibited, such that the normal biological activity of the polypeptide is prevented.

The polypeptide of the invention that is employed in such a screening technique may be 30 free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to

observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

Alternatively, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest (see International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide can also be

coated directly onto plates for use in the aforementioned drug screening techniques.

The polypeptide of the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, that are described above.

The invention includes the agonists, antagonists, ligands, receptors, substrates and enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound [X] is "substantially free of" impurities [herein, Y] when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent needed to treat, ameliorate, or prevent a targetted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for 5 example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the 10 disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 15 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other 20 polypeptides, genes and other therapeutic agents such as liposomes, provided that the carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers and inactive virus particles.

25 Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

30 Pharmaceutically acceptable carriers in therapeutic compositions may additionally

contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, 5 slurries, suspensions, and the like, for ingestion by the patient.

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

The pharmaceutical compositions utilised in this invention may be administered by any 10 number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, 15 the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, 20 subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state, 25 several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

30 In another approach, soluble forms of the polypeptide that retain binding affinity for the

ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered. Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. *et al.* (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated *in situ* from expression *in vivo*.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, *et al.*, *Curr. Opin. Struct. Biol.* (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5'

and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, 5 methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine which are not as easily recognised by endogenous endonucleases.

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound that activates 10 the polypeptide, i.e., an agonist as described above, to alleviate the abnormal condition. Alternatively, a therapeutic amount of the polypeptide in combination with a suitable pharmaceutical carrier may be administered to restore the relevant physiological balance of polypeptide.

Gene therapy may be employed to effect the endogenous production of the polypeptide 15 by the relevant cells in the subject. Gene therapy is used to treat permanently the inappropriate production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur *in vivo* or *ex vivo*. *Ex vivo* gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic 20 gene and introduction of the genetically altered cells back into the patient. In contrast, *in vivo* gene therapy does not require isolation and purification of a patient's cells.

The therapeutic gene is typically "packaged" for administration to a patient. Gene 25 delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in *Curr. Top. Microbiol. Immunol.*, 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in *Curr. Top. Microbiol. Immunol.*, 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention 30 may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide,

such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo* (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in 5 Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are 10 disease-causing agents, the invention provides that they can be used in vaccines to raise antibodies against the disease causing agent.

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in 15 combination with pharmaceutically-acceptable carriers as described above, which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H.* 20 *pylori*, and other pathogens.

Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may 25 contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose 30 containers. For example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use.

The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

- This invention also relates to the use of nucleic acid molecules according to the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.
- 10 Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki *et al.*, *Nature*, 324, 163-166 (1986); Bej, *et al.*, *Crit. Rev. Biochem. Molec. Biol.*, 26, 301-334 (1991); Birkenmeyer *et al.*, *J. Virol. Meth.*, 35, 117-126 (1991); Van Brunt, J., *Bio/Technology*, 8, 291-294 (1990)) prior to analysis.

In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of disease. The method may comprise the steps of:

- a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;
- 25 b) contacting a control sample with said probe under the same conditions used in step a);
- c) and detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

- a) obtaining a tissue sample from a patient being tested for disease;
- b) isolating a nucleic acid molecule according to the invention from said tissue sample;  
and,
- c) diagnosing the patient for disease by detecting the presence of a mutation in the  
5 nucleic acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, for example using PCR, may be included.

Deletions and insertions can be detected by a change in the size of the amplified product, in comparison to the normal genotype. Point mutations can be identified by hybridizing 10 amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly-matched sequences can be distinguished from mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent 15 conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

20 Such diagnostics are particularly useful for prenatal and even neonatal testing.

Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita *et al.*, *Genomics*, 5, 874-879 25 (1989)). For example, a sequencing primer may be used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. Further, point mutations and

other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

- DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (for example, Myers *et al.*, *Science* (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, *Proc. Natl. Acad. Sci. USA* (1985) 85: 4397-4401).
- 10 In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be detected by *in situ* analysis (see, for example, Keller *et al.*, *DNA Probes*, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane.
- 15 Fluorescence *in situ* hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck *et al.*, *Science*, 250: 559-562 (1990), and Trask *et al.*, *Trends, Genet.* 7:149-154 (1991)).
- In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient 20 screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee *et al.*, *Science* (1996) 274: 610-613).
- In one embodiment, the array is prepared and used according to the methods described in 25 PCT application WO95/11995 (Chee *et al.*); Lockhart, D. J. *et al.* (1996) *Nat. Biotech.* 14: 1675-1680; and Schena, M. *et al.* (1996) *Proc. Natl. Acad. Sci.* 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or 30 any other suitable solid support. In another aspect, an oligonucleotide may be synthesized

on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler *et al.*). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a 5 vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends 10 itself to the efficient use of commercially-available instrumentation.

In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the 15 quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present 20 invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

25 Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex 30 formation. The amount of standard complex formation may be quantified by various

methods, such as by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

- 20 (a) a nucleic acid molecule of the present invention;
- (b) a polypeptide of the present invention; or
- (c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of

nucleic acid molecules, at least one of which may be a nucleic acid molecule according to the invention.

To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent 5 useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to disease, particularly cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis 10 disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system 15 disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hypothyroidism, hyperparathyroidism, hypercalcemia, hypocalcemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular 20 hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

Various aspects and embodiments of the present invention will now be described in more 25 detail by way of example, with particular reference to the LBDG3 polypeptide and LBDG3 full length polypeptide.

It will be appreciated that modification of detail may be made without departing from the scope of the invention.

**Brief description of the Figures**

Figure 1: This is the front end of the Biopendium Target Mining Interface. A search of the database is initiated using the PDB code "1EXA:A".

Figure 2A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1EXA:A. The arrow indicates the *Homo sapiens* Retinoic Acid Receptor Gamma-1, which has a typical Nuclear Hormone Receptor Ligand Binding Domain.

Figure 2B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1EXA:A. The arrow indicates CAB55953.1 (LBDG3).

Figure 2C: Full list of forward PSI-BLAST results for the search using 1EXA:A. 10 CAB55953.1 (LBDG3) is not identified.

Figure 3: The Redundant Sequence Display results page for CAB55953.1 (LBDG3).

Figure 4: InterPro PFAM search results for CAB55953.1 (LBDG3).

Figure 5: NCBI Protein Report for CAB55953.1 (LBDG3).

Figure 6A: This is the front end of the Biopendium database. A search of the database is initiated using CAB55953.1 (LBDG3), as the query sequence.

Figure 6B: A selection of the Inpharmatica Genome Threader results of search using CAB55953.1 (LBDG3), as the query sequence. The arrow points to 1EXA:A.

Figure 6C: The reverse-maximised PSI-BLAST results obtained using CAB55953.1 (LBDG3), as the query sequence.

20 Figure 7: AlEye sequence alignment of CAB55953.1 (LBDG3) and 1EXA:A.

Figure 8A: LigEye for 1EXA:A that illustrates the sites of interaction of R-3-fluoro-4-[2-hydroxy-2-(5,5,8,8-tetramethyl-5,6,7,8-tetrahydro-naphthalen-2-yl)-acetylamino]-benzoic acid [ abbreviated to: Bns270394 ] (394450) with the Ligand Binding Domain of *Homo sapiens* Retinoic Acid Receptor Gamma-2, 1EXA:A.

25 Figure 8B: iRasMol view of 1EXA:A, the Ligand Binding Domain of *Homo sapiens* Retinoic Acid Receptor Gamma-2.

Figure 9: AlEye sequence alignment of CAB55953.1 (LBDG3), AAF36513.1,

AAF36514.1 and AAF36515.1.

Figure 10: NCBI protein report for DKFZP727M231.

Figure 11: NCBI sequence listing for DKFZP727M231.

Figure 12: Electronic PCR results showing that AL117480 can be found in the genomic  
5 fragment AL13285.35.

Figure 13: NCBI nucleotide report showing that AL13285.35 can be found at 20q11.21-  
12.

Figure 14: PubMed search result showing that a susceptibility gene for the autoimmune  
thyroid disease (AITD) Graves Disease maps to 20q11.2.

10 Figure 15: PubMed search result showing that amplification of genomic regions between  
20q11.2-q12 have been correlated to several cancers.

Figure 16: The linear dynamic range for target CAC14946.1 (LBDG3) reactions on colon  
cDNA.

15 Figure 17: The linear dynamic range for internal control 18s rRNA reactions on colon  
cDNA.

Figure 18: Normalised expression of CAC14946.1 (LBDG3) in 18 normal human tissues.

Figure 19: Genome Threader alignment of *Homo sapiens* Retinoic Acid Receptor gamma  
structure 1EXA:A with CAB55953.1 (LBDG3). Included in the alignment are the five  
homologues of CAB55953.1 (LBDG3); *Mus musculus* AAH03931.1, *Danio rerio*

20 AAF36515.1 and AAF36514.1, *Ciona* LBDG3\_Ciona and *Oikopleura* LBDG3\_Oiko.  
Characteristic Ligand Binding Domain residues are marked by black boxes labeled a-i.

Figure 20: Homstrand report for a selection of known Nuclear Hormone Receptor Ligand  
Binding Domain structures. 1LBD=*Homo sapiens* Retinoic X Receptor alpha,  
2LBD=*Homo sapiens* Retinoic Acid Receptor gamma, 1PRG:A=*Homo sapiens*

25 Peroxisome Proliferator Activated Receptor gamma, 3ERT:A=*Homo sapiens* Estrogen  
Receptor alpha and 1A28:A=*Homo sapiens* Progesterone Receptor. The structural role of  
each residue is described by case, bold, underline or italic as described in the Key to  
Homstrand.

Figure 21: Results of the PSI-PRED secondary structure prediction program for the *Homo sapiens* Retinoic Acid Receptor gamma sequence (*H.s.* RAR $\gamma$ predicted 2 $\circ$ ) and the CAB55953.1 (LBDG3) sequence (LBDG3predicted 2 $\circ$ ), in the region of the "LBD motif". (The two PSI-PRED outputs have been aligned from PHE305 onwards by reference to the Genome Threader alignment of CAB55953.1 (LBDG3) with *Homo sapiens* Retinoic Acid Receptor gamma). Grey shading indicates prediction of an  $\alpha$ -helix (the higher the column, the more confident the helical prediction), black shading indicates no secondary structure prediction has been made. The top of the figure (*H.s.* RAR $\gamma$ known 2 $\circ$ ) depicts the actual known secondary structure derived from the crystal structure of *Homo sapiens* Retinoic Acid Receptor gamma (IEXA:A). An experimentally observed kink in helix  $\alpha$ 5 is marked by an arrow.

Figure 22: Results of the PSI-PRED secondary structure prediction program for *Homo sapiens* Retinoic Acid Receptor gamma sequence (*H.sapiens* RAR $\gamma$ ), the CAB55953.1 (LBDG3) sequence (*H.sapiens* LBDG3), and the five CAB55953.1 (LBDG3) homologues (*M.musculus* AAH03931.1, *D.rerio* AAF36515.1, *D.rerio* AAF36514.1, *Oikopleura* LBDG3\_Oiko and *Ciona* LBDG3\_Ciona). (As in Figure 21, the PSI-PRED outputs have been aligned from PHE305 onwards by reference to the Genome Threader alignment of CAB55953.1 (LBDG3) with *Homo sapiens* Retinoic Acid Receptor gamma. In addition the RAR $\gamma$  residues GLY250 - LEU254 have been cropped from the RAR $\gamma$  output due to space limitations, this is marked as a white line between PRO249 and SER255).

Figure 23: Overall view of the homology model of CAB55953.1 (LBDG3) adopting the Ligand Binding Domain fold. Residues of particular interest are marked in black.

Figure 24: View of the homology model of CAB55953.1 (LBDG3) adopting the Ligand Binding Domain fold, showing only the region encompassing the predicted helices " $\alpha$ 3" and " $\alpha$ 5". Grey arrows mark the direction of the polypeptide chain running N-terminal to C-terminal.

Figure 25: Close-up of the predicted Co-activator binding site of the homology model of CAB55953.1 (LBDG3) adopting the Ligand Binding Domain fold. Residues of particular interest are shown in black.

Figure 26: Close-up of the predicted Co-activator binding site of the homology model of CAB55953.1 (LBDG3) adopting the Ligand Binding Domain fold. Residues of particular interest are shown in black. A cartoon of a Co-activator helix has been added to illustrate the predicted mode of binding to the homology model.

**Example: 1 CAB55953.1 (LBDG3)**

In order to initiate a search for novel, distantly related Nuclear Hormone Receptor Ligand Binding Domains, an archetypal family member is chosen, the Ligand Binding Domain of *Homo sapiens* Retinoic Acid Receptor Gamma-2. More specifically, the search is 5 initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the Ligand Binding Domain of *Homo sapiens* Retinoic Acid Receptor Gamma-2 (PDB code 1EXA:A; see Figure 1).

A search of the Biopendium (using the Target Mining Interface) for relatives of 1EXA:A 10 takes place and returns 4462 Genome Threader results. The 4462 Genome Threader results include examples of typical Nuclear Hormone Receptor Ligand Binding Domains, such as that found between residues 182-417 of the *Homo sapiens* Retinoic Acid Receptor Gamma-1 (see arrow in Figure 2A).

Among the proteins known to contain a Nuclear Hormone Receptor Ligand Binding 15 Domain appears a protein which is not annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain, CAB55953.1 (LBDG3; see arrow in Figure 2B). The Inpharmatica Genome Threader has identified a region of the sequence CAB55953.1 (LBDG3), between residues 311-452, as having a structure similar to the Ligand Binding Domain of *Homo sapiens* Retinoic Acid Receptor Gamma-2. The possession of a 20 structure similar to the Ligand Binding Domain of *Homo sapiens* Retinoic Acid Receptor Gamma-2 suggests that residues 311-452 of CAB55953.1 (LBDG3) function as a Nuclear Hormone Receptor Ligand Binding Domain. The Genome Threader identifies this with 94% confidence.

The search of the Biopendium (using the Target Mining Interface) for relatives of 1EXA:A 25 also returns 850 Forward PSI-Blast results. Forward PSI-Blast (see Figure 2C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

In order to assess what is known in the public domain databases about CAB55953.1 (LBDG3) the Redundant Sequence Display Page (Figure 3) is viewed. There are no PROSITE or PRINTS hits which identify CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain. PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no Nuclear Hormone Receptor Ligand Binding Domain hits from both databases means that CAB55953.1 (LBDG3) is unidentifiable as containing a Nuclear Hormone Receptor Ligand Binding Domain using PROSITE or PRINTS.

In order to identify if any other public domain annotation vehicle is able to annotate CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain, the CAB55953.1 (LBDG3) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) at the InterPro website (see Figure 4). There are no PFAM-A matches annotating CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain. Thus PFAM does not identify CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

The National Center for Biotechnology Information (NCBI) Genebank protein database is then viewed to examine if there is any further information that is known in the public domain relating to CAB55953.1 (LBDG3). This is the US public domain database for protein and gene sequence deposition (Figure 5). CAB55953.1 (LBDG3) is a *Homo sapiens* sequence, its Genebank protein ID is CAB55953.1 and it is 560 amino acids in length. CAB55953.1 (LBDG3) was cloned by a group of scientists at MIPS, Am Klopferspitz 18a, D-82152, Martinsried, Germany. The public domain information for this gene does not annotate it as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Therefore, it can be concluded that using all public domain annotation tools, CAB55953.1 (LBDG3) may not be annotated as containing a Nuclear Hormone Receptor Ligand Binding Domain. Only the Inpharmatica Genome Threader is able to annotate this protein as containing a Nuclear Hormone Receptor Ligand Binding Domain.

The reverse search is now carried out. CAB55953.1 (LBDG3) is now used as the query sequence in the Biopendium (see Figure 6A). The Inpharmatica Genome Threader identifies residues 311-452 of CAB55953.1 (LBDG3) as having a structure that is the same as the Ligand Binding Domain of *Homo sapiens* Retinoic Acid Receptor Gamma-2 with 94% confidence (see arrow in Figure 6B). The Ligand Binding Domain of *Homo sapiens* Retinoic Acid Receptor Gamma-2 (1EXA:A) was the original query sequence. Positive iterations of PSI-Blast do not return this result (Figure 6C). It is only the Inpharmatica Genome Threader that is able to identify this relationship.

The sequence of the *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain, 1EXA:A is chosen against which to view the sequence alignment of CAB55953.1 (LBDG3). Viewing the AlEye alignment (Figure 7) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

The *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain contains an "LBD motif" which has been found in all annotated Nuclear Hormone Receptor Ligand Binding Domains to date. The "LBD motif" is involved in recruiting Nuclear Hormone Receptor Co-Activators and Co-Repressors. The 6 residues PHE251, LEU254, ASP258, GLN259, LEU262 and LEU263 constitute this motif in the *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain (see square boxes Figure 7). However, only 4 (ASP258, GLN259, LEU262 and LEU263) of these 6 residues lie within the Genome Threader alignment. Thus only these 4 residues can be used to consolidate the Genome Threader annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain. 4 residues (ASP314, GLN315, LEU318 and LEU319) in CAB55953.1 (LBDG3) precisely match all 4 (ASP258, GLN259, LEU262 and LEU263) of the "LBD motif" residues in the *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain (which are within the region of genome Threader alignment). This indicates that CAB55953.1 (LBDG3) contains a Nuclear Hormone Receptor Ligand Binding Domain similar to the *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs "LigEye" (Figure 8A) and "iRasmol" (Figure 8B) are used. These visualization tools identify the active site of known protein structures by indicating the amino acids with which known small molecule inhibitors interact at the active site. These 5 interactions are either through a direct hydrogen bond or through hydrophobic interactions. In this manner, one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of known structure. The LigEye view of the *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain reveals 14 residues which bind R-3-fluoro-4-[2-hydroxy-2-(5,5,8,8-tetramethyl-5,6,7,8,-10 tetrahydro-naphthalen-2-yl)-acetylamino]-benzoic acid [abbreviated to: Bms270394] (circled in Figure 7). However, only 9 (LEU271, MET272, ILE275, ARG278, PHE288, SER289, GLY303, PHE304, and LEU307) of these 14 residues lie within the Genome Threader alignment. Thus only these 9 residues can be used to consolidate the Genome Threader annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone 15 Receptor Ligand Binding Domain.

Of these 9 residues, 6 (LEU271, MET272, ILE275, PHE288, PHE304, and LEU307) are hydrophobic and form a hydrophobic pocket in the *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain (Bms270394 binds this hydrophobic cavity). 2 of these hydrophobic residues (ILE275 and PHE304) are perfectly conserved in 20 CAB55953.1 (LBDG3; ILE331 and PHE363; see Figure 7). Furthermore, LEU271, MET272, PHE288 and LEU307 of the *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain are substituted by the hydrophobic residues ILE327, LEU328, LEU349 and TYR366 (respectively) in CAB55953.1 (LBDG3): (broken circle in Figure 7). This conservation of hydrophobicity in 6 out of the 6 hydrophobic residues (within the 25 region of Genome Threader alignment) which line the binding pocket indicates that CAB55953.1 (LBDG3) will bind a hydrophobic steroid-like ligand. This indicates that indeed as predicted by the Inpharmatica Genome Threader, CAB55953.1 (LBDG3) folds in a similar manner to the *Homo sapiens* Retinoic Acid Receptor Gamma-2 Ligand Binding Domain and as such is identified as containing a Nuclear Hormone Receptor 30 Ligand Binding Domain.

Reverse-maximised PSI-BLAST of CAB55953.1 (LBDG3) identifies homologues in

- Mus musculus* (AAF36513.1, see Figure 6C arrow ①) and *Danio rerio* (AAF36514.1 Figure 6C, arrow ② and AAF36515.1 Figure 6C, arrow ③).
- CAB55953.1 (LBDG3), AAF36513.1 (*Mus musculus* homologue), AAF36514.1 (*Danio rerio* homologue) and AAF36515.1 (second *Danio rerio* homologue) are aligned and viewed in AlEye (Figure 9).
- AlEye reveals that the 6 hydrophobic residues of CAB55953.1 (LBDG3) predicted to line the ligand binding pocket (ILE327, LEU328, ILE331, LEU349, PHE363 and TYR366) are all precisely conserved in AAF36513.1 (*Mus musculus* homologue), AAF36514.1 (*Danio rerio* homologue) and AAF36515.1 (second *Danio rerio* homologue).
- Furthermore, the 4 residues (within the genome threader alignment) matching the "LBD motif" in CAB55953.1 (LBDG3), (ASP314, GLN315, LEU318 and LEU319) are all precisely conserved in AAF36513.1 (*Mus musculus* homologue), AAF36514.1 (*Danio rerio* homologue) and AAF36515.1 (second *Danio rerio* homologue). Residues which are essential for the function of a protein will be conserved in homologues of that protein.
- Thus the precise conservation of "LBD motif" and hydrophobic residues which would be essential for the function of the predicted CAB55953.1 (LBDG3) Nuclear Hormone Receptor Ligand Binding Domain in the *Mus musculus* and 2 *Danio rerio* homologues strongly supports the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain.
- Figure 10 is a report generated from the NCBI UniGene database. This database is a collection of expressed sequence tags (ESTs) from various human tissues, it can be used to give a general tissue distribution for a protein provided that its sequence is present in the database. CAB55953.1 (LBDG3) is present in the database and is shown to be expressed in the tissues described as Aorta, Blood, Brain, Breast, CNS, Colon, Eye, Germ Cell, Heart, Kidney, Lung, Muscle, Nose, Ovary, Pancreas, Parathyroid, Peripheral nervous system, Placenta, Prostate, Stomach, Testis, Uterus, Whole embryo, amnion\_normal, bladder\_tumor, brain, breast, breast\_normal, cervix, colon, colon\_ins, colon\_normal, denis\_drasch, epid\_tumor, eye, genitourinary tract, head\_neck, head\_normal, kidney, kidney\_tumor, lung, lung\_tumor, lymph, muscle, nervous\_tumor, ovary, pancreas, placenta, pnet, prostate\_normal, skin, thymus, pooled, thyroid, uterus, uterus\_tumor, whole blood.

Using public domain information, CAB55953.1 (LBDG3) can be mapped *in silico* to a cytogenetic locus. CAB55953.1 (LBDG3) is encoded by AL117480 (Figure 11), which by ePCR at the NCBI site can be found within the genomic fragment AL13285.35 (Figure 12, arrow). AL13285.35 has been mapped, and lies between 20q11.21-12 (Figure 5 13, arrow). Thus the gene which encodes CAB55953.1 (LBDG3) must also lie between 20q11.21-12.

A susceptibility gene for the autoimmune thyroid disease (AITD) Graves Disease maps to 20q11.2 (Figure 14). This is particularly interesting because thyroid hormone (and related molecules) are known to bind Nuclear Hormone Receptor Ligand Binding Domains. Also 10 the UniGene report indicates that CAB55953.1 (LBDG3) is expressed in the thyroid. Amplification of genomic regions between 20q11.2-q12 have been correlated with several cancers for example gastric carcinoma (Figure 15). It may thus be inferred that the CAB55953.1 (LBDG3) gene may be implicated in these conditions, so paving the way for 15 the development of agents that target the CAB55953.1 (LBDG3) gene or its encoded protein, to diagnose and/or treat these conditions. In particular, the identification of this gene as containing a Nuclear Hormone Receptor Ligand Binding Domain facilitates the development of agents that are specific to the CAB55953.1 (LBDG3) protein, for example, through the use of Nuclear Hormone Receptor Ligand Binding Domain agonists or antagonists.

20

#### Example 2

In order to determine the tissue expression of the proposed LBD, Taqman RT-PCR quantitation was used. The TaqMan 3'- 5' exonuclease assay signals the formation of PCR amplicons by a process involving the nucleolytic degradation of a double-labeled 25 fluorogenic probe that hybridises to the target template at a site between the two primer recognition sequences (cf. U. S. Patent 5,876,930). The ABI Prism 7000 automates the detection and quantitative measurement of these signals, which are stoichiometrically related to the quantities of amplicons produced, during each cycle of amplification. In addition to providing substantial reductions in the time and labour requirements for PCR

analyses, this technology permits simplified and potentially highly accurate quantification of target sequences in the reactions.

Human RNA prepared from non-diseased organs were purchased from either Ambion Europe (Huntingdon, UK) or Clontech. Oligonucleotide primers and probes were designed 5 using Primer Express software (PE Applied Biosystems, Foster City CA) with a GC-content of 40-60%, no G-nucleotide at the 5'-end, and no more than 4 contiguous Gs. Each primer and probe was analysed using BLAST<sup>®</sup> (Basic Local Alignment Search Tool, Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.; J Mol Biol 1990 Oct 5;215(3):403-10). Results confirmed that each oligonucleotide recognised the target sequence with a specificity >3 bp 10 when compared to other known cDNA's or genomic sequence represented in the Unigene and GoldenPath publicly available databases.

The sequence of the primers and probes were:

CAC14946.1 (LBDG3) Forward primer: AAC ATC CCA GAG GTG GAA GCT

CAC14946.1 (LBDG3) Reverse primer: CAG ACG AGT TAA TAA GCC CCT CTT

15 CAC14946.1 (LBDG3) Probe: CAT CAC ACA CCA AAC TCC CGG TGT T

18s pre-optimised primers and probe were purchased from Applied Biosystems, Foster City, CA. Probes were covalently conjugated with a fluorescent reporter dye (e.g. 6carboxy-fluorescein [FAM]; Xem = 518nm) and a fluorescent quencher dye (6carboxytetramethylrhodamine [TAMRA]; Mem = 582nm) at the most 5' and most 3' base, respectively. All 20 primers and probes were obtained from Perkin Elmer, Germany. Primer/probe concentrations were titrated in the range of 50nM to 900nM and optimal concentrations for efficient PCR reactions were determined. Optimal primer and probe concentrations varied in between 100nM and 900nM depending on the target gene that was amplified. cDNA was prepared using components from PE Applied Biosystems, Foster City CA. 50 $\mu$ l reactions 25 were prepared in 0.5ml RNase free tubes. Reactions contained 500ng total RNA; 1x reverse transcriptase buffer; 5.5mM MgCl<sub>2</sub>; 1mM dNTP's; 2.5 $\mu$ l random hexamers; 20U RNase inhibitor; and 62.5U reverse transcriptase.

25 $\mu$ l reactions were prepared in 0.5 ml thin-walled, optical grade PCR 96 well plates (PE Applied Biosystems, Foster City CA). Reactions contained: 1x final concentration of 30 TaqMan Universal Master Mix (a proprietary mixture of AmpliTaq Gold DNA polymerase,

AmpEraseX UNG, dNTPs with UTP, passive reference dye and optimised buffer components, PE Applied Biosystems, Foster City CA); 100nM Taqman probe; 300nM forward primer; 900nM reverse primer and 15ng of cDNA template.

Standard procedures for the operation of the ABI Prism 7000 or similar detection system  
5 were used. This included, for example with the ABI Prism 7000, use of all default program settings with the exception of reaction volume which was changed from 50 to 25  $\mu$ l. Thermal cycling conditions consisted of two min at 50 C, 10 min at 95 C, followed by 40 cycles of 15 sec at 95 C and 1 min at 60 C. Cycle threshold (Ct) determinations, (i.e. non-integer calculations of the number of cycles required for reporter dye fluorescence resulting  
10 from the synthesis of PCR products to become significantly higher than background fluorescence levels), were automatically performed by the instrument for each reaction using default parameters. Assays for target sequences and ribosomal 18s (reference) sequences in the same cDNA samples were performed in separate reaction tubes. Within  
15 each experiment, a standard curve was carried out of a typical tissue sample, from 50ng to 0.78ng of cDNA template. From this standard curve, the amount of actual starting target or 18s cDNA in each test sample is determined.

The levels of target cDNA in each sample were normalised to the level of expression of target in stomach. The levels of 18s cDNA in each sample were normalised to the level of expression of 18s in stomach. The data was then represented as fold expression of  
20 normalised target sequence relative to the level of expression in stomach cDNA, which was set arbitrarily to 1. For each experiment, the transcript is quantitated 2-4 times. The data shows the mean  $\pm$  SEM for the set of experiments.

Taqman RT-PCR was carried out on 2-fold dilutions of colon cDNA using primers/probes specific for CAC14946.1 (LBDG3) as described in the detailed description. Figure 16 shows  
25 the Ct values plotted vs. the log input cDNA value, and illustrates that a linear relationship was seen over this range of input cDNA concentrations. Linear regression analysis of the standard curve was used to calculate the starting amount of cDNA from test Ct values.

Taqman RT-PCR was carried out on 2-fold dilutions of colon cDNA using primers/probes specific for 18s rRNA as described in the detailed description. Figure 17 shows the Ct  
30 values plotted vs. the log input cDNA value, and illustrates that a linear relationship was

seen over this range of input cDNA concentrations. Linear regression analysis of the standard curve was used to calculate the starting amount of cDNA from test Ct values.

Taqman RT-PCR was carried out using 15ng of the indicated cDNA using primers/probes specific for CAC14946.1 (LBDG3) and 18s rRNA as described in the detailed description.

- 5 A standard curve for target and internal control was also carried out, using between 50ng to 0.78ng of cDNA template of a typical tissue sample. Using linear regression analysis of the standard curves, the Ct values were used to calculate the amount of actual starting target or 18s cDNA in each test sample.

The levels of target cDNA in each sample were normalised to the level of expression of 10 target in a comparative sample, in this case, stomach. The levels of 18s cDNA in each sample were also normalised to the level of expression of 18s in stomach. The expression levels of CAC14946.1 (LBDG3) were then normalised to the expression levels of 18s. Figure 18 represents the fold expression of normalised target sequence relative to the level of expression in stomach cDNA, which is set arbitrarily to 1. Each sample was quantitated 15 in 3 individual experiments. Figure 18 shows the mean  $\pm$  SEM for the multiple experiments.

The finding that the mRNA for LBDG3 is expressed at significant levels in the human brain is noteworthy as this provides a potential link to human disease states and development of agonists and antagonists for the ligand binding domain of LBDG3 and offers the potential for therapeutic intervention in various human diseases including, cell proliferative disorders, 20 including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, 25 cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, 30 hyperthyroidism, hypothyroidism, hyperparathyroidism, hypercalcemia, hypocalcemia,

hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

The finding of a "non-classical" nuclear hormone receptor such as LBDG3 which contains a ligand binding domain in the absence of a DNA binding domain, is consistent with the known literature which has consistently reported widespread effects of steroids in the brain (known as neurosteroids) and that these effects, in general, are mediated not through the known classic steroid hormone nuclear receptors which require transcriptional activation. For instance, neurosteroids have been shown to influence neurotransmission particularly in the field of receptors such as those for GABA and NMDA and Sigma receptors. Neurosteroids have been shown to play a neuroprotective role. Therapeutic intervention through the development of agonists (or antagonists) to LBDG3 may therefore have a role in treatment of neurodegenerative conditions such as dementia, Parkinson's disease and neurodegeneration following cerebrovascular disease such as infarction or haemorrhage (stroke) and trauma to the central nervous system and spinal cord. In addition, neurosteroids have been shown to influence cognitive processing, spatial learning and memory, anxiety and behaviours such as craving which leads to addictive behaviour patterns. Development of agonists and antagonists to LBDG3 may therefore lead to therapeutic intervention to treat dementias, learning difficulties, anxiety, addictive behaviours such as but not exclusively alcoholism, eating disorders and drug addiction.

LBDG3 is not exclusively expressed in the brain. Significant levels of mRNA are also found in the adrenal, ovary, testis and thymus. The adrenal, ovary and testis are significant sites for the biosynthesis of steroids and their activity and is consistent but not exclusive with a role for LBDG3 in steroid actions. The finding of LBDG3 in these tissues supports the development of antagonists and agonists for treatment of diseases including but not exclusive to hypertension, responses to stress including stress of infectious diseases, regulation of salt and water homeostasis, control of fertility through regulation of ovulation (infertility and contraception), regulation of implantation (infertility and contraception) and regulation of spermatogenesis (infertility and contraception). In addition, these agents may

be of value in treating steroid responsive tumours such as benign prostatic hypertrophy, prostatic cancer, ovarian cancer and testicular cancer.

The finding of LBDG3 in the thymus is consistent with a role in T cell development. Agonists and antagonists developed to LBDG3 may therefore play a role in regulating T 5 cells in disease processes such as autoimmune diseases and allergies including type I diabetes mellitus, rheumatoid arthritis, multiple sclerosis, psoriasis, renal failure arising from glomerulopathies, scleroderma, inflammatory bowel disease (both Crohns disease and ulcerative colitis), transplant rejection, asthma, atopic dermatitis, eczema, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and 10 other pathological conditions, particularly those in which nuclear hormone receptors are implicated.

#### Example 3

Three homologues of CAB55953.1 (LBDG3) are herein identified and aligned with 15 CAB55953.1 (LBDG3) to demonstrate the conservation of key functional residues across the homologues (such as those of the predicted "LBD motif"; see Figure 6C). These homologues are *Mus musculus* AAH03931.1, *Danio rerio* AAF36515.1 and *Danio rerio* AAF36514.1. The conservation of such residues across these homologues indicates that 20 evolutionary pressure is being specifically applied to these particular residues, supporting their annotation as "LBD motif" residues. This provides support for the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

The *Mus musculus* homologue (AAH03931.1) has 98% sequence identity to 25 CAB55953.1 (LBDG3) at the amino acid level. The *Danio rerio* homologue (AAF36514.1) has 74% sequence identity to CAB55953.1 (LBDG3) at the amino acid level. The *Danio rerio* homologue (AAF36515.1) has 69% sequence identity to 30 CAB55953.1 (LBDG3) at the amino acid level. The identification of other, more divergent, homologues of CAB55953.1 (LBDG3) would serve as a further test of the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain: it would be predicted that even when overall sequence identity

descends towards 30%, residues which are functionally critical (such as those of the "LBD motif") would still be conserved.

Two new, more divergent, homologues of CAB55953.1 (LBDG3) have also been identified in the NCBI-month database. A homologue of CAB55953.1 (LBDG3) was 5 identified from a genomic BAC sequence AF374376.1 of the marine chordate *Oikopleura*. This *Oikopleura* homologue shares 44% sequence identity with CAB55953.1 (LBDG3), and will be referred to as LBDG3\_Oiko.

Another homologue of CAB55953.1 (LBDG3) was identified from 5' (AL666329.1) and 10 3' (AL664923.1) EST sequences of a transcript derived from the chordate *Ciona*. The EST sequences do not overlap, and the gap is indicated by "X"s in the sequence. This 15 *Ciona* homologue shares 37% sequence identity with CAB55953.1 (LBDG3), and will be referred to as LBDG3\_Ciona. Figure 19 shows an alignment of CAB55953.1 (LBDG3) with the *Homo sapiens* Retinoic Acid Receptor gamma Ligand Binding Domain (1EXA:A) and the CAB55953.1 (LBDG3) homologues from *Mus musculus* (AAH03931.1), *Danio rerio* (AAF36515.1 and AAF36514.1), *Ciona* (LBDG3\_Ciona) and *Oikopleura* (LBDG3\_Oiko).

The *Homo sapiens* Retinoic Acid Receptor gamma Ligand Binding Domain contains an "LBD motif" which has been found in all annotated Nuclear Hormone Receptor Ligand Binding Domains to date. The "LBD motif" is involved in recruiting Nuclear Hormone 20 Receptor Co-Activators and Co-Repressors. The 6 residues PHE251, LEU254, ASP258, GLN259, LEU262 and LEU263 constitute this motif in the *Homo sapiens* Retinoic Acid Receptor gamma Ligand Binding Domain (see black boxes marked c, d, e, f, g, and h Figure 19). As discussed above, four of these "LBD motif" residues are conserved in CAB55953.1 (LBDG3). 1EXA:A ASP258 is conserved as ASP314 in CAB55953.1 25 (LBDG3). 1EXA:A GLN259 is conserved as GLN315 in CAB55953.1 (LBDG3). 1EXA:A LEU262 is conserved as LEU318 in CAB55953.1 (LBDG3). 1EXA:A LEU263 is conserved as LEU319 in CAB55953.1 (LBDG3). Strikingly, these four residues are all 30 conserved in the five CAB55953.1 (LBDG3) homologues (with the exception of a conservative substitution of a TYR for LEU318 in LBDG3\_Oiko), see Figure 19, black boxes marked c, f, g, and h. The specific conservation of these four residues, particularly

when overall sequence identity is as low as 37% for LBDG3\_Ciona, provides strong evidence for these four residues playing a critical role in CAB55953.1 (LBDG3) and the five CAB55953.1 (LBDG3) homologues. This agrees with ASP314, GLN315, LEU318 and LEU319 functioning as "LBD motif" residues and supports the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

In addition to the "LBD motif" residues, there are other residues which are well conserved in known Nuclear Hormone Receptor Ligand Binding Domains. A further test of the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain is to analyse whether important residues outside of the "LBD motif" are conserved in CAB55953.1 (LBDG3) and the five CAB55953.1 (LBDG3) homologues.

One useful source of data on important conserved residues is the Homstrad database (Mizuguchi K, Deane CM, Blundell TL, Overington JP. (1998) HOMSTRAD: a database of protein structure alignments for homologous families. Protein Science 7:2469-2471). Homstrad (HOMologous STRucture Alignment Database) documents which residues are conserved across a particular protein family and details their structural role. Figure 20 shows the Homstrad report for a selection of known Nuclear Hormone Receptor Ligand Binding Domain structures.

The Homstrad report contains an entry for the *Homo sapiens* Retinoic Acid Receptor gamma structure 2LBD. 2LBD and 1EXA:A are two equivalent structures of the *Homo sapiens* Retinoic Acid Receptor gamma Ligand binding domain, differing only in their bound ligand (2LBD contains all-trans retinoic acid, 1EXA:A contains bms270394). Referring to Figure 20, arrow a, it can be seen that PHE244 of *Homo sapiens* Retinoic Acid Receptor gamma is a solvent inaccessible residue that is conserved as a solvent inaccessible aromatic residue in the other Nuclear Hormone Receptor Ligand Binding Domains (1LBD *Homo sapiens* Retinoid X Receptor alpha, 1PRG:A *Homo sapiens* Peroxisome Proliferator Activated Receptor gamma, 3ERT:A *Homo sapiens* Estrogen Receptor alpha and 1A28:A *Homo sapiens* Progesterone Receptor). The conservation of this solvent inaccessible aromatic residue across the known Nuclear Hormone Receptor

Ligand Binding Domain family indicates that it plays an important structural role in the Ligand Binding Domain fold. In the Genome Threader alignment, PHE244 of *Homo sapiens* Retinoic Acid Receptor gamma is conserved as PHE305 of CAB55953.1 (LBDG3), (Figure 19 box a). The conservation of this key solvent inaccessible aromatic residue supports the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain. The significance of this conservation is enhanced by the observation that PHE305 is conserved as a PHE in all five CAB55953.1 (LBDG3) homologues (Figure 19 box a). This further supports the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Referring to Figure 20, arrow b it can be seen that LYS246 of *Homo sapiens* Retinoic Acid Receptor gamma is a solvent accessible residue that is conserved as a solvent accessible positively charged residue in the other Nuclear Hormone Receptor Ligand Binding Domains (1LBD *Homo sapiens* Retinoid X Receptor alpha, 1PRG:A *Homo sapiens* Peroxisome Proliferator Activated Receptor gamma, 3ERT:A *Homo sapiens* Estrogen Receptor alpha and 1A28:A *Homo sapiens* Progesterone Receptor). The conservation of this solvent accessible positively charged residue across the known Nuclear Hormone Receptor Ligand Binding Domain family indicates that it plays an important role in Ligand Binding Domain function. Indeed, experimental data indicates that this solvent exposed positively charged residue functions as a "charge clamp" in the recruitment of Nuclear Hormone Receptor Co-Activators (A. K. Shiu, D. Barstad, P. M. Loria, Lin Cheng, P. J. Kushner, D. A. Agard, and G. L. Greene, The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen Cell 1998 95: 927). In the Genome Threader alignment, LYS246 of *Homo sapiens* Retinoic Acid Receptor gamma is conservatively substituted by the positively charged ARG307 of CAB55953.1 (LBDG3), (Figure 19 box b). The conservation of this key positively charged residue supports the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain. The significance of this conservation is enhanced by the observation that ARG307 is conserved as an ARG in all five CAB55953.1 (LBDG3) homologues (Figure 19 box b). This further supports the

annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Referring to Figure 20, arrow i, it can be seen that ASP269 of *Homo sapiens* Retinoic Acid Receptor gamma is conserved as an acidic residue in three out of the four other

5 Nuclear Hormone Receptor Ligand Binding Domains (1LBD *Homo sapiens* Retinoid X Receptor alpha, 1PRG:A *Homo sapiens* Peroxisome Proliferator Activated Receptor gamma, and 3ERT:A *Homo sapiens* Estrogen Receptor alpha). The conservation of this acidic residue in many of the known Nuclear Hormone Receptor Ligand Binding Domains indicates that it plays an important role in the Ligand Binding Domain fold. In  
10 the Genome Threader alignment, ASP269 of *Homo sapiens* Retinoic Acid Receptor gamma is conservatively substituted as GLU325 of CAB55953.1 (LBDG3), (Figure 19 box i). The conservation of this acidic residue supports the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain. The significance of this conservation is enhanced by the observation that GLU325 is  
15 conserved as a acidic (GLU or ASP) in all five CAB55953.1 (LBDG3) homologues (Figure 19 box i). This further supports the annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain.

Genome Threader has annotated CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain. The composite secondary structure  
20 prediction program PSI-PRED (Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292: 195-202) provides an independent method to test this annotation. Figure 21 shows the PSI-PRED output for the *Homo sapiens* Retinoic Acid Receptor gamma sequence (in the region of the "LBD motif") compared to the actual known secondary structure derived from the  
25 crystal structure of *Homo sapiens* Retinoic Acid Receptor gamma (1EXA:A). It can be seen that PSI-PRED correctly predicts the alpha helices  $\alpha_3$  and  $\alpha_5$  from the sequence of *Homo sapiens* Retinoic Acid Receptor gamma. PSI-PRED has been documented to have an accuracy of 70%. If CAB55953.1 (LBDG3) does indeed adopt a Ligand Binding Domain fold, then it should have an identical secondary structure to that of the *Homo  
30 sapiens* Retinoic Acid Receptor gamma

- Ligand Binding Domain. Figure 21 shows the PSI-PRED output for CAB55953.1 (LBDG3) (in the region of the "LBD motif"). When this is compared to the known secondary structure of *Homo sapiens* Retinoic Acid Receptor gamma (1EXA:A), it can be clearly seen that CAB55953.1 (LBDG3) is predicted to adopt 2  $\alpha$ -helices that
- 5 correspond to  $\alpha$ 3 and  $\alpha$ 5 of 1EXA:A. This is an independent validation of the Genome Threader annotation of CAB55953.1 (LBDG3) as containing a Nuclear Receptor Ligand Binding Domain. A distinctive feature of  $\alpha$ 5 in *Homo sapiens* Retinoic Acid Receptor gamma 1EXA:A structure is that it has a "kink" half way along the helix. This  $\alpha$ 5 helix kink is also observed in other known Ligand Binding Domains. The  $\alpha$ 5 helix kink is
- 10 critical in forming the ligand binding pocket. The  $\alpha$ 5 helix kink is marked in Figure 21 as an arrow, and one can see that this is reflected in the *Homo sapiens* Retinoic Acid Receptor gamma PSI-PRED output as a small dip in the probability of the helix prediction. Interestingly, CAB55953.1 (LBDG3) also shows a dip in the probability of the helix prediction at precisely this position.
- 15 The known structure of *Homo sapiens* Retinoic Acid Receptor gamma 1EXA:A contains a three residue long  $3_{10}$  helix ( $\alpha$ 4) that connects  $\alpha$ 3 and  $\alpha$ 5 (from the PSI-PRED output for *Homo sapiens* Retinoic Acid Receptor gamma sequence it can be seen that PSI-PRED makes no prediction of this  $3_{10}$  helix as would be expected with this program). The Genome Threader alignment of CAB55953.1 (LBDG3) with 1EXA:A, inserts gaps in the
- 20 CAB55953.1 (LBDG3) sequence in the region aligned with the 1EXA:A  $3_{10}$  helix. This indicates that the helices which correspond to  $\alpha$ 3 and  $\alpha$ 5 are not connected by a  $3_{10}$  helix but are instead connected by a short loop region with the sequence (GLY308-THR309-THR310). Indeed the predicted absence of this  $3_{10}$  helix accounts for the absence of the first two residues from the "LBD motif" (PHE251 and LEU254 of 1EXA:A) in
- 25 CAB55953.1 (LBDG3).

The PSI-PRED outputs for the five CAB55953.1 (LBDG3) homologues are shown aligned with the outputs for CAB55953.1 (LBDG3) and *Homo sapiens* Retinoic Acid Receptor gamma in Figure 22. Strikingly, all five homologues are also predicted to adopt alpha helices which correspond to  $\alpha$ 3 and  $\alpha$ 5 of *Homo sapiens* Retinoic Acid Receptor

30 gamma (LBDG3\_Ciona shows a slightly weaker prediction for the C-terminus of the

“ $\alpha$ 5” helix, but this can be accounted for by the shorter length of the LBDG3\_Ciona sequence failing to pick up as many PSI-BLAST hits during the PSI-PRED calculation). The observation that all five homologues have equivalent PSI-PRED predictions to that of CAB55953.1 (LBDG3) adds significance to the prediction that CAB55953.1 (LBDG3) will adopt 2  $\alpha$ -helices that correspond to  $\alpha$ 3 and  $\alpha$ 5 of the *Homo sapiens* Retinoic Acid Receptor gamma Ligand Binding Domain. This further strengthens the CAB55953.1 (LBDG3) PSI-PRED output as an independent validation of the Genome Threader annotation that CAB55953.1 (LBDG3) contains a Nuclear Hormone Receptor Ligand Binding Domain.

10 A further test of the Genome Threader annotation of CAB55953.1 (LBDG3) as containing a Nuclear Hormone Receptor Ligand Binding Domain is to attempt to construct a homology model of CAB55953.1 (LBDG3) adopting the LBD fold. A homology model of CAB55953.1 (LBDG3) was constructed by submitting the Genome Threader alignment of CAB55953.1 (LBDG3) with the *Homo sapiens* Retinoic X Receptor gamma structure 1LBD to Modeller version 5 (Sali, A and Blundell T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234,779-815, 1993). An overall view of the homology model is presented in Figure 23. It can be seen that CAB55953.1 (LBDG3) has been successfully modelled onto the 1LBD structure and exhibits the typical organization of cross-latticed  $\alpha$ -helices which make up

15 the LBD fold. Figure 24 focuses on the two key alpha helices “ $\alpha$ 3” and “ $\alpha$ 5” (which correspond to  $\alpha$ 3 and  $\alpha$ 5 in known Nuclear Hormone Receptor Ligand Binding Domain structures). (The “ $\alpha$ 3” and “ $\alpha$ 5” helices in this homology model are the two helices which were predicted for CAB55953.1 (LBDG3) by PSI-PRED in Figure 21). Figure 25 zooms in on the predicted Co-Activator binding site. In known LBD structures the Co-

20 Activator binding site is a groove formed between  $\alpha$ 3 and  $\alpha$ 5 on the Ligand Binding Domain surface. Figure 26 shows the same view, but with a cartoon of the interacting Co-Activator helix added to illustrate the predicted mode of Co-Activator binding to the groove). The four conserved residues of the predicted “LBD motif” in CAB55953.1 (LBDG3), ASP314, GLN315, LEU318 and LEU319 are marked in dark grey, and are all

25 in the appropriate orientations to fulfill their roles without any clashes or steric hindrance.

30

For example, GLN315 is in a suitable position to interact with the C-terminus of the Co-Activator helix, just as is observed for the equivalent residue GLN375 of Estrogen receptor alpha, when it binds the Co-activator helix of SRC-1 (A. K. Shiu, D. Barstad, P. M. Loria, Lin Cheng, P. J. Kushner, D. A. Agard, and G. L. Greene, The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen Cell 1998 95: 927). Similarly, LEU319 projects into the groove and could make hydrophobic contacts with several of the Leucines present in the LXXLL Co-Activator helix, just as is observed for the equivalent residue LEU379 of Estrogen receptor alpha, when it binds the Co-activator helix of SRC-1 (A. K. Shiu, D. Barstad, P. M. Loria, Lin Cheng, P. J. Kushner, D. A. Agard, and G. L. Greene, The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen Cell 1998 95: 927). Residues outside of the "LBD motif" are also positioned in appropriate orientations to fulfill their roles. For example, ARG307 projects into the groove and could operate as a "charge clamp" on the C-terminus of the Co-Activator helix, just as is observed for the equivalent residue LYS362 of Estrogen receptor alpha, when it binds the Co-activator helix of SRC-1 (A. K. Shiu, D. Barstad, P. M. Loria, Lin Cheng, P. J. Kushner, D. A. Agard, and G. L. Greene, The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen Cell 1998 95: 927). The homology modelling of CAB55953.1 (LBDG3) as a Ligand Binding Domain supports the Genome Threader annotation of CAB55953.1 (LBDG3) as containing a Ligand Binding Domain.

## CLAIMS

1. A polypeptide, which polypeptide:
  - (i) comprises the amino acid sequence as recited in SEQ ID NO:2 or SEQ ID NO:4;
  - (ii) is a fragment thereof having Nuclear Hormone Receptor Ligand Binding Domain activity or having an antigenic determinant in common with the polypeptide of (i); or
  - (iii) is a functional equivalent of (i) or (ii).
- 10 2. A polypeptide according to claim 1, which consists of the amino acid sequence as recited in SEQ ID NO:2 or SEQ ID NO:4.
3. A polypeptide which is a fragment according to claim 1(ii), which includes the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide, said Nuclear Hormone Receptor Ligand Binding Domain region being defined as including residues 311 to 452 inclusive, of the amino acid sequence recited in SEQ ID NO:2, wherein said fragment possesses the "LBD motif" residues ASP314, GLN315, LEU318 and LEU319, or equivalent residues, of the amino acid sequence recited in SEQ ID NO:2, and possesses Nuclear Hormone Receptor Ligand Binding Domain activity.
- 15 4. A polypeptide which is a functional equivalent according to claim 1(iii), is homologous to the amino acid sequence as recited in SEQ ID NO:2, possesses the "LBD motif" residues ASP314, GLN315, LEU318 and LEU319, or equivalent residues, and has Nuclear Hormone Receptor Ligand Binding Domain activity.
5. A polypeptide according to claim 4, wherein said functional equivalent is homologous to the Nuclear Hormone Receptor Ligand Binding Domain region of the LBDG3 polypeptide.
- 20 6. A fragment or functional equivalent according to any one of claims 1-5, which has greater than 80% sequence identity with an amino acid sequence as recited in SEQ ID

SUBSTITUTE SHEET (RULE 26)

- NO:2, or with a fragment thereof that possesses Nuclear Hormone Receptor Ligand Binding Domain activity, preferably greater than 85%, 90%, 95%, 98% or 99% sequence identity, as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].
5. 7. A functional equivalent according to any one of claims 1-6, which exhibits significant structural homology with a polypeptide having the amino acid sequence given in any one of SEQ ID NO:2, or with a fragment thereof that possesses Nuclear Hormone Receptor Ligand Binding Domain activity.
10. 8. A fragment as recited in claim 1-3, or 6, having an antigenic determinant in common with the polypeptide of claim 1(i), which consists of 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more) amino acid residues from the sequence of SEQ ID NO:2.
15. 9. A purified nucleic acid molecule which encodes a polypeptide according to any one of the preceding claims.
10. 10. A purified nucleic acid molecule according to claim 9, which has the nucleic acid sequence as recited in SEQ ID NO:1 or SEQ ID NO:3, or is a redundant equivalent or fragment thereof.
20. 11. A fragment of a purified nucleic acid molecule according to claim 10, which comprises nucleotides 932 to 1357 of SEQ ID NO:1, or is a redundant equivalent thereof.
12. A purified nucleic acid molecule which hybridizes under high stringency conditions with a nucleic acid molecule according to any one of claims 9-11.
13. A vector comprising a nucleic acid molecule as recited in any one of claims 9-12.
25. 14. A host cell transformed with a vector according to claim 13.
15. A ligand which binds specifically to, and which preferably inhibits the Nuclear Hormone Receptor Ligand Binding Domain activity of, a polypeptide according to any one of claims 1-8.

**SUBSTITUTE SHEET (RULE 26)**

16. A ligand according to claim 15, which is an antibody.
17. A compound that either increases or decreases the level of expression or activity of a polypeptide according to any one of claims 1-8.
18. A compound according to claim 17 that binds to a polypeptide according to any one of claims 1-8 without inducing any of the biological effects of the polypeptide.
19. A compound according to claim 17 or claim 18, which is a natural or modified substrate, ligand, enzyme, receptor or structural or functional mimetic.
20. A polypeptide according to any one of claims 1-8, a nucleic acid molecule according to any one of claims 9-12, a vector according to claim 13, a ligand according to claim 15 or 16, or a compound according to any one of claims 17-19, for use in therapy or diagnosis of disease.
21. A method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to any one of claims 1-8, or assessing the activity of a polypeptide according to any one of claims 1-8, in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of disease.
22. A method according to claim 21 that is carried out *in vitro*.
23. A method according to claim 21 or claim 22, which comprises the steps of: (a) contacting a ligand according to claim 15 or claim 16 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.
24. A method according to claim 21 or claim 22, comprising the steps of:
  - a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 9-12 and the probe;
  - b) contacting a control sample with said probe under the same conditions used in step a); and
  - c) detecting the presence of hybrid complexes in said samples;

SUBSTITUTE SHEET (RULE 26)

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

25. A method according to claim 21 or claim 22, comprising:

- 5 a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 9-12 and the primer;
  - b) contacting a control sample with said primer under the same conditions used in step a); and
  - c) amplifying the sampled nucleic acid; and
  - 10 d) detecting the level of amplified nucleic acid from both patient and control samples;
- wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of disease.

15 26. A method according to claim 21 or claim 22 comprising:

- a) obtaining a tissue sample from a patient being tested for disease;
- b) isolating a nucleic acid molecule according to any one of claims 9-12 from said tissue sample; and
- c) diagnosing the patient for disease by detecting the presence of a mutation which is 20 associated with disease in the nucleic acid molecule as an indication of the disease.

27. The method of claim 26, further comprising amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the amplified product.

28. The method of either claim 26 or 27, wherein the presence or absence of the mutation 25 in the patient is detected by contacting said nucleic acid molecule with a nucleic acid probe that hybridises to said nucleic acid molecule under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an

**SUBSTITUTE SHEET (RULE 26)**

unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation.

- 5 29. A method according to any one of claims 21-28, wherein said disease is selected from cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hypothyroidism, hyperparathyroidism, hypercalcemia, hypocalcemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated..
- 10 30. Use of a polypeptide according to any one of claims 1-8 as a Nuclear Hormone Receptor Ligand Binding Domain.
- 15 31. Use of a nucleic acid molecule according to any one of claims 9-12 to express a protein that possesses Nuclear Hormone Receptor Ligand Binding Domain activity.
- 20 32. A method for effecting cell-cell adhesion, utilising a polypeptide according to any one of claims 1-8.

33. A pharmaceutical composition comprising a polypeptide according to any one of claims 1-8, a nucleic acid molecule according to any one of claims 9-12, a vector according to claim 13, a ligand according to claim 15 or 16, or a compound according to any one of claims 17-19.
- 5 34. A vaccine composition comprising a polypeptide according to any one of claims 1-8 or a nucleic acid molecule according to any one of claims 9-12.
35. A polypeptide according to any one of claims 1-8, a nucleic acid molecule according to any one of claims 9-12, a vector according to claim 13, a ligand according to claim 15 or 16, a compound according to any one of claims 17-19, or a pharmaceutical composition according to claim 33 for use in the manufacture of a medicament for the treatment of cell proliferative disorders, including neoplasm, melanoma, lung, colorectal, breast, pancreas, head and neck and other solid tumours, myeloproliferative disorders, such as leukemia, non-Hodgkin lymphoma, leukopenia, thrombocytopenia, angiogenesis disorder, Kaposi's sarcoma, 10 autoimmune/inflammatory disorders, including allergy, inflammatory bowel disease, arthritis, psoriasis and respiratory tract inflammation, asthma, and organ transplant rejection, cardiovascular disorders, including hypertension, oedema, angina, atherosclerosis, thrombosis, sepsis, shock, reperfusion injury, heart arrhythmia, and ischemia, neurological disorders including, central nervous system disease, 15 Alzheimer's disease, brain injury, stroke, amyotrophic lateral sclerosis, anxiety, depression, and pain, developmental disorders, metabolic disorders including diabetes mellitus, osteoporosis, lipid metabolism disorder, hyperthyroidism, hypothyroidism, hyperparathyroidism, hypercalcemia, hypocalcemia, hypercholesterolemia, hyperlipidemia, and obesity, renal disorders, including glomerulonephritis, 20 renovascular hypertension, dermatological disorders, including, acne, eczema, and wound healing, negative effects of aging, AIDS, infections including viral infection, bacterial infection, fungal infection and parasitic infection and other pathological conditions, particularly those in which nuclear hormone receptors are implicated..
- 25 36. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1-8, a nucleic acid molecule according to

any one of claims 9-12, a vector according to claim 13, a ligand according to claim 15 or 16, a compound according to any one of claims 17-19, or a pharmaceutical composition according to claim 33.

37. A method according to claim 36, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist.
38. A method according to claim 36, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist.
39. A method of monitoring the therapeutic treatment of disease in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide according to any one of claims 1-8, or the level of expression of a nucleic acid molecule according to any one of claims 9-12 in tissue from said patient, wherein altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease.
40. A method for the identification of a compound that is effective in the treatment and/or diagnosis of disease, comprising contacting a polypeptide according to any one of claims 1-8, a nucleic acid molecule according to any one of claims 9-12, or a host cell according to claim 13 with one or more compounds suspected of possessing binding affinity for said polypeptide or nucleic acid molecule, and selecting a compound that binds specifically to said nucleic acid molecule or polypeptide.
41. A kit useful for diagnosing disease comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to any one of claims 9-12; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease.

42. The kit of claim 41, further comprising a third container holding an agent for digesting unhybridised RNA.
43. A kit comprising an array of nucleic acid molecules, at least one of which is a nucleic acid molecule according to any one of claims 9-12.
- 5 44. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1-8; and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.
45. A transgenic or knockout non-human animal that has been transformed to express higher, lower or absent levels of a polypeptide according to any one of claims 1-8.
- 10 46. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 45 with a candidate compound and determining the effect of the compound on the disease of the animal.

SUBSTITUTE SHEET (RULE 26)

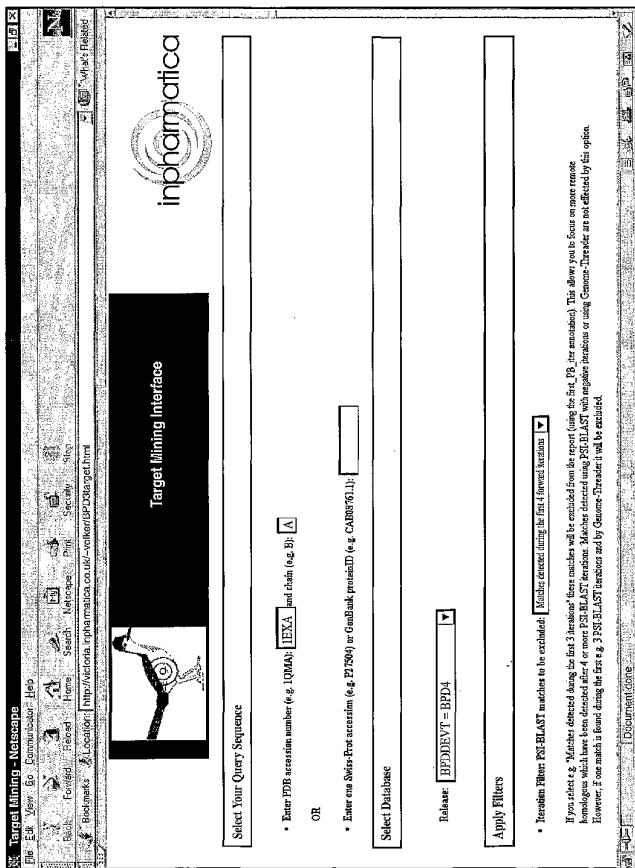


FIG. 1

**SUBSTITUTE SHEET (RULE 26)**

FIG. 2A

**SUBSTITUTE SHEET (RULE 26)**

Target Mining Results - Netscape									
Algorithm name	EDD file	WWW link	Title	Organization	EDD	WWW	Category	Count	Count
CA_AFFINITY			Hypothetical protein	Human experts	PR	1618	14.2%	52,323	52,323
CA_CENTRAL									
CA_DENSE									
CA_DENSEST									
CA_DENSEST2									
CA_DENSEST3									
CA_DENSEST4									
CA_DENSEST5									
CA_DENSEST6									
CA_DENSEST7									
CA_DENSEST8									
CA_DENSEST9									
CA_DENSEST10									
CA_DENSEST11									
CA_DENSEST12									
CA_DENSEST13									
CA_DENSEST14									
CA_DENSEST15									
CA_DENSEST16									
CA_DENSEST17									
CA_DENSEST18									
CA_DENSEST19									
CA_DENSEST20									
CA_DENSEST21									
CA_DENSEST22									
CA_DENSEST23									
CA_DENSEST24									
CA_DENSEST25									
CA_DENSEST26									
CA_DENSEST27									
CA_DENSEST28									
CA_DENSEST29									
CA_DENSEST30									
CA_DENSEST31									
CA_DENSEST32									
CA_DENSEST33									
CA_DENSEST34									
CA_DENSEST35									
CA_DENSEST36									
CA_DENSEST37									
CA_DENSEST38									
CA_DENSEST39									
CA_DENSEST40									
CA_DENSEST41									
CA_DENSEST42									
CA_DENSEST43									
CA_DENSEST44									
CA_DENSEST45									
CA_DENSEST46									
CA_DENSEST47									
CA_DENSEST48									
CA_DENSEST49									
CA_DENSEST50									
CA_DENSEST51									
CA_DENSEST52									
CA_DENSEST53									
CA_DENSEST54									
CA_DENSEST55									
CA_DENSEST56									
CA_DENSEST57									
CA_DENSEST58									
CA_DENSEST59									
CA_DENSEST60									
CA_DENSEST61									
CA_DENSEST62									
CA_DENSEST63									
CA_DENSEST64									
CA_DENSEST65									
CA_DENSEST66									
CA_DENSEST67									
CA_DENSEST68									
CA_DENSEST69									
CA_DENSEST70									
CA_DENSEST71									
CA_DENSEST72									
CA_DENSEST73									
CA_DENSEST74									
CA_DENSEST75									
CA_DENSEST76									
CA_DENSEST77									
CA_DENSEST78									
CA_DENSEST79									
CA_DENSEST80									
CA_DENSEST81									
CA_DENSEST82									
CA_DENSEST83									
CA_DENSEST84									
CA_DENSEST85									
CA_DENSEST86									
CA_DENSEST87									
CA_DENSEST88									
CA_DENSEST89									
CA_DENSEST90									
CA_DENSEST91									
CA_DENSEST92									
CA_DENSEST93									
CA_DENSEST94									
CA_DENSEST95									
CA_DENSEST96									
CA_DENSEST97									
CA_DENSEST98									
CA_DENSEST99									
CA_DENSEST100									
CA_DENSEST101									
CA_DENSEST102									
CA_DENSEST103									
CA_DENSEST104									
CA_DENSEST105									
CA_DENSEST106									
CA_DENSEST107									
CA_DENSEST108									
CA_DENSEST109									
CA_DENSEST110									
CA_DENSEST111									
CA_DENSEST112									
CA_DENSEST113									
CA_DENSEST114									
CA_DENSEST115									
CA_DENSEST116									
CA_DENSEST117									
CA_DENSEST118									
CA_DENSEST119									
CA_DENSEST120									
CA_DENSEST121									
CA_DENSEST122									
CA_DENSEST123									
CA_DENSEST124									
CA_DENSEST125									
CA_DENSEST126									
CA_DENSEST127									
CA_DENSEST128									
CA_DENSEST129									
CA_DENSEST130									
CA_DENSEST131									
CA_DENSEST132									
CA_DENSEST133									
CA_DENSEST134									
CA_DENSEST135									
CA_DENSEST136									
CA_DENSEST137									
CA_DENSEST138									
CA_DENSEST139									
CA_DENSEST140									
CA_DENSEST141									
CA_DENSEST142									
CA_DENSEST143									
CA_DENSEST144									
CA_DENSEST145									
CA_DENSEST146									
CA_DENSEST147									
CA_DENSEST148									
CA_DENSEST149									
CA_DENSEST150									
CA_DENSEST151									
CA_DENSEST152									
CA_DENSEST153									
CA_DENSEST154									
CA_DENSEST155									
CA_DENSEST156									
CA_DENSEST157									
CA_DENSEST158									
CA_DENSEST159									
CA_DENSEST160									
CA_DENSEST161									
CA_DENSEST162									
CA_DENSEST163									
CA_DENSEST164									
CA_DENSEST165									
CA_DENSEST166									
CA_DENSEST167									
CA_DENSEST168									
CA_DENSEST169									
CA_DENSEST170									
CA_DENSEST171									
CA_DENSEST172									
CA_DENSEST173									
CA_DENSEST174									
CA_DENSEST175									
CA_DENSEST176									
CA_DENSEST177									
CA_DENSEST178									
CA_DENSEST179									
CA_DENSEST180									
CA_DENSEST181									
CA_DENSEST182									
CA_DENSEST183									
CA_DENSEST184									
CA_DENSEST185									
CA_DENSEST186									
CA_DENSEST187									
CA_DENSEST188									
CA_DENSEST189									
CA_DENSEST190									
CA_DENSEST191									
CA_DENSEST192									
CA_DENSEST193									
CA_DENSEST194									
CA_DENSEST195									
CA_DENSEST196									
CA_DENSEST197									
CA_DENSEST198									
CA_DENSEST199									
CA_DENSEST200									
CA_DENSEST201									



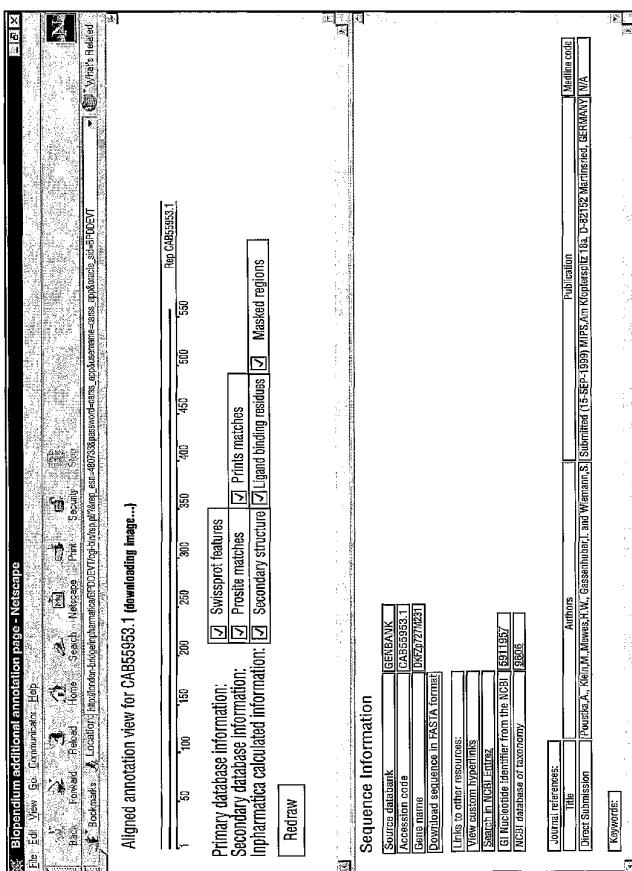


FIG. 3

SUBSTITUTE SHEET (RULE 26)

FIG. 4

InterPro search Results.				
1 Query Sequence <u>gi</u> Length 560 aa.				
InterPro	Results of PPsearch against <u>PROSITE</u>	Results of PFScan against <u>PROSITE</u>	Results of FingerPRINTScan against <u>PRINTS</u>	Results of HMMDecoder against <u>PFAM-A</u>
No significant hits reported; check program's raw output.				
XML / TXT formatted.				
Applications finished normally.				

SUBSTITUTE SHEET (RULE 26)

FIG. 5

**SUBSTITUTE SHEET (RULE 26)**

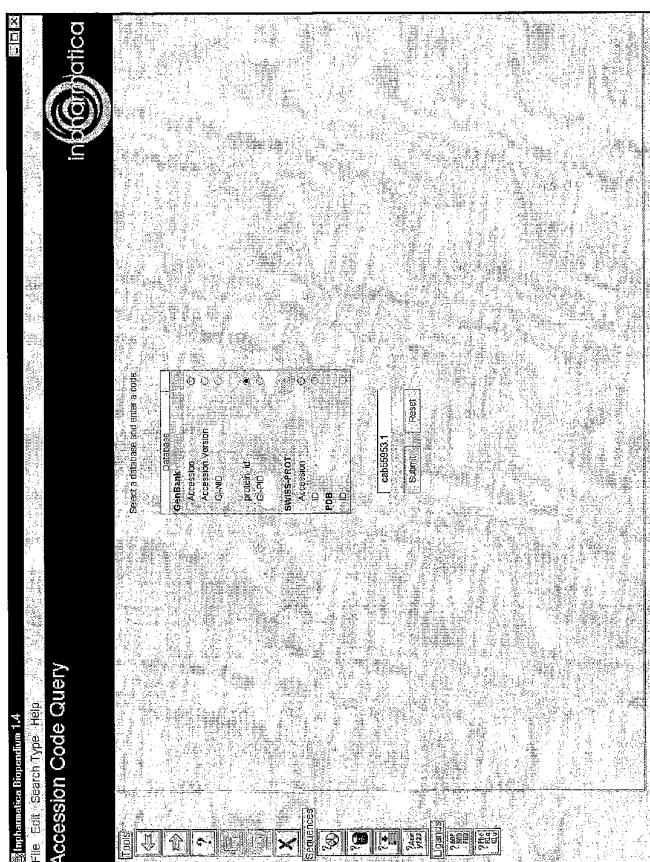


FIG. 6A  
Accession Code Query

WO 02/070557

PCT/GB02/00937

9/34

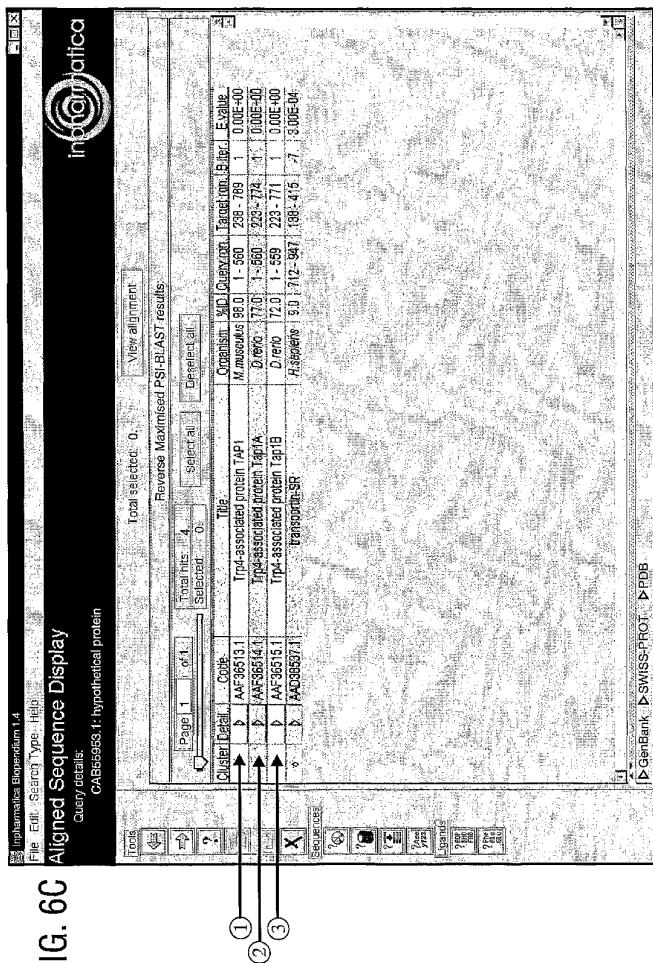
**SUBSTITUTE SHEET (RULE 26)**

FIG. 6C Aligned Sequence Display

File Edit Search Type Help

Query details:  
GAB65653, 1: hypothetical protein

**SUBSTITUTE SHEET (RULE 26)**



WO 02/070557

PCT/GB02/00937

AIEye output (January 15, 2001 2:18 PM)

FIG. 7

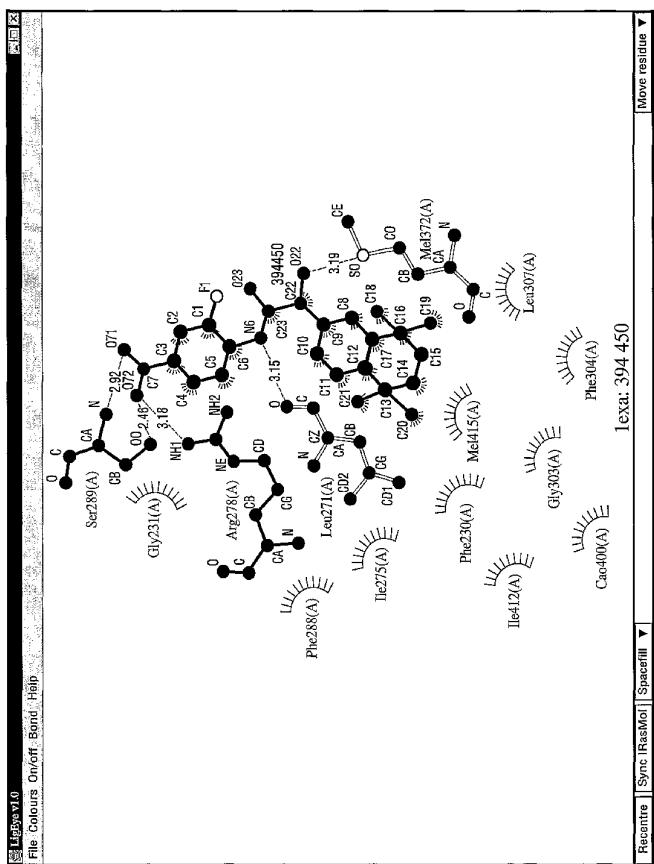


FIG. 8A

SUBSTITUTE SHEET (RULE 26)

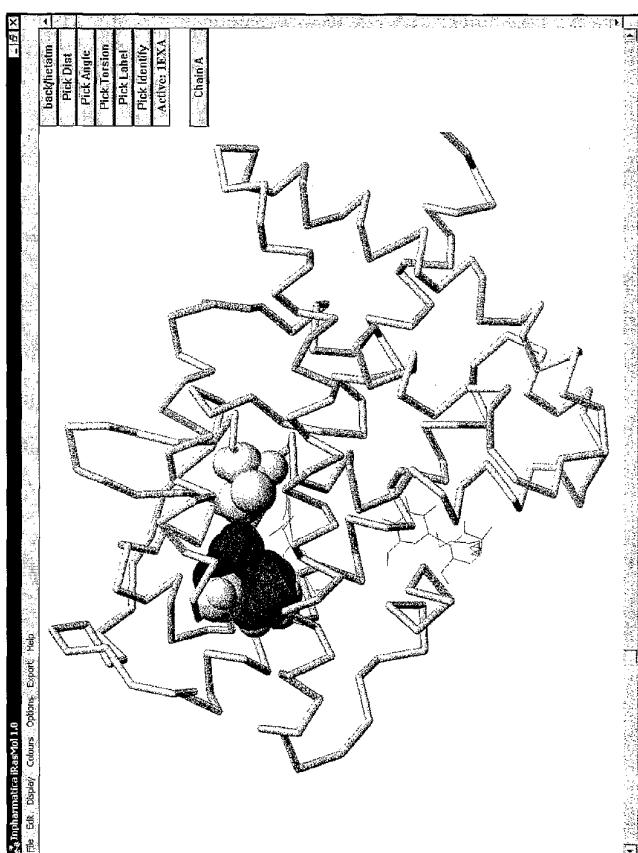


FIG. 8B

SUBSTITUTE SHEET (RULE 26)

WO 02/070557

PCT/GB02/00937

14/34

FIG. 3.1 AlEye output (February 26, 2001 5:04 PM)

FIG. 1

**SUBSTITUTE SHEET (RULE 26)**

WO 02/070557

PCT/GB02/00937

15/34

Figure 2. AIEye output (February 26, 2001 5:04 PM)

FIG. 9-2

**SUBSTITUTE SHEET (RULE 26)**

## FIG. 10

 **UniGene**  
**Homo sapiens**  
**DKFZP727M231 protein**

PubMed	Entrez	BLAST	OMIM	Taxonomy	Structure
Search <input type="text" value="Human"/> for <input type="text"/>			display as html <input type="checkbox"/> <input type="button" value="Go"/>		

switch to text mode

UniGene Cluster Hs.168073 DKFZP727M231

UniGene **DKFZP727M231 protein**

**SEE ALSO**  
**LocusLink:** 26133  
**HomoloGene:** Hs.168073

**SELECTED MODEL ORGANISM PROTEIN SIMILARITIES**  
 organism, protein and percent identity and length of aligned region

*H. sapiens*: PIR:T17263 - T17263 hypothetical protein 100 % / 559 aa  
 DKFZP727M231.I - human

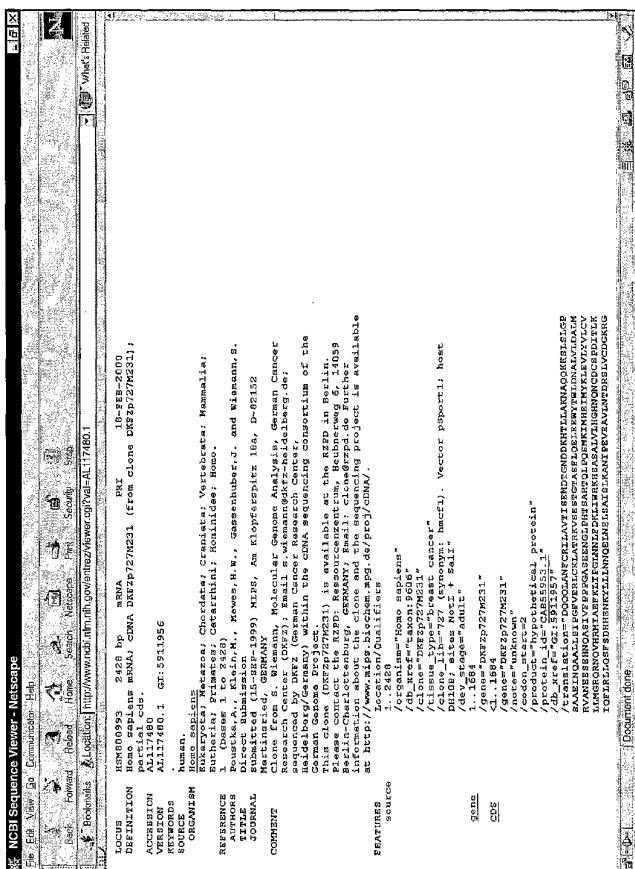
*M. musculus*: PID:g7108682 - AF130458\_1 Trp4-associated 98 % / 559 aa  
 protein TAP1

**MAPPING INFORMATION**  
**Chromosome:** 20  
**UniSTS entries:** 1923  
**UniSTS entries:** stSG4612 Genomic Context: Map View  
**UniSTS entries:** stSG3028

**EXPRESSION INFORMATION**  
**UniGene Organisms**  
**cDNA sources:** Aorta, Blood, Brain, Breast, CNS, Colon, Eye, Germ Cell, Heart, Kidney, Lung, Muscle, Nose, Ovary, Pancreas, Parathyroid, Peripheral nervous system, Placenta, Prostate, Stomach, Testis, Uterus, Whole embryo, amnion\_normal, brain, breast, breast\_normal, cervix, colon, colon\_normal, denis, drash, eye, genitourinary tract, head\_neck, kidney, lung, lung\_tumor, lymph, muscle, nervous\_tumor, ovary, pancreas, placenta, pnt, prostate\_normal, skin, thymus, pooled, thyroid, uterus, uterus\_tumor, whole blood

**SAGE :** Gene to Tag mapping

**mRNA/GENE SEQUENCES (3)**  
**SUBSTITUTE SHEET (RULE 26)**



11

**SUBSTITUTE SHEET (RULE 26)**

WO 02/070557

PCT/GB02/00937

18/34

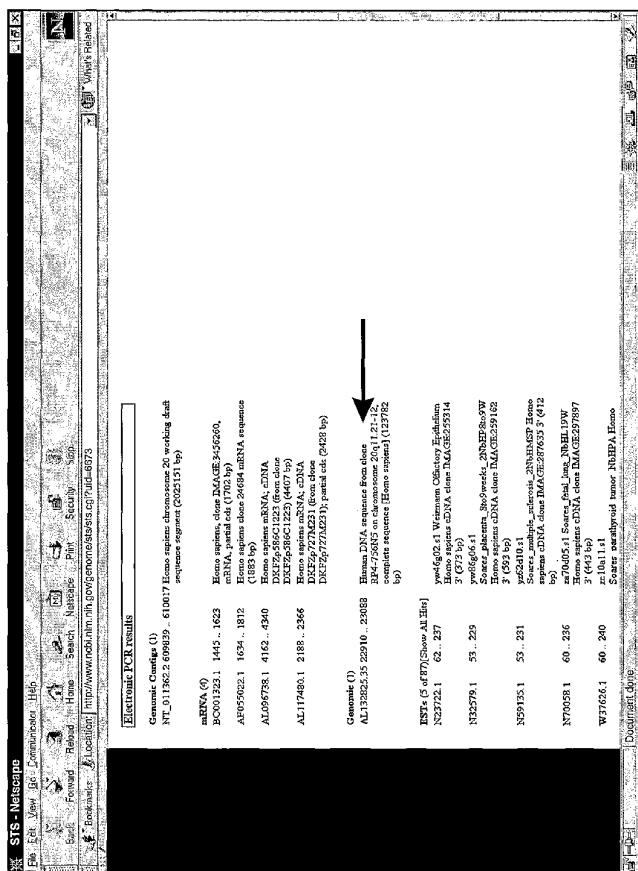


FIG. 12

**SUBSTITUTE SHEET (RULE 26)**

Entrez-Nucleotide - Ntsepe

File Edit View Go Communicator Help

Search Reference For All 1325355

Submit

Pubmed BioStudies Locations <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=AL> 32/25 3:24pm 8/26/07 MDoping

Protein Nucleotide Structure PeptSet PeptTax Protein Related Searches Taxonomy Link/DOI

What's Related

Protein History Alignment Cigar

Protein Description ID Name [Text] [Email] [Print] [Clipboard]

□ 1: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=AL&id=1325355> [Get] [Help]

LOCUS [HAD756N](#) 123723 Nucleic acid sequence

DEFINITION Human RNA virus, clone HAD-756N, on chromosome 20q11.21-12. Contains a novel gene (KA1A122) encoding a myelin heavy chain protein (h), a novel gene encoding a protein similar to Tp4-associated Protein 1A1 (hacB2), 2 protein similar to Tp4-associated protein sequences.

ACCESSION [M13328](#)

VERSION [M13328.35](#)

KEYWORDS Hvg, AscA2, Cpe, island, KHA1A122, myelin, human, virus, single-stranded, RNA, virus, Cpe, AscA2, Cpe, island, KHA1A122, primates, Ceballosi, Verhaertda, Eubalostomii; [M13328.35](#)

REFERENCE 1 (bases to 123723)

AUTHORS Moore, M.

JOURNAL Direct Submission

COMMENT On Aug 22, 2007 this sequence version replaced [G184973.9](#).

Where differences are found, they are annotated as variations together with a note of the overlapping clone names. Note that the sequence version replaced is the one in the original submission and not the one in the cloning clone. The sequence submitted with this version is the one in the cloning clone.

This sequence has been finished according to sequence map criteria, as attempts is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known sequence. The sequence is submitted in FASTA format. The sequence is submitted in FASTA format.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Document Home

13

**SUBSTITUTE SHEET (RULE 26)**

## FIG. 14

The University of  
Chicago press

**A new Graves disease-susceptibility locus maps to chromosome 20q11.2. International Consortium for the Genetics of Autoimmune Thyroid Disease.**

**Tomer Y, Barbescino G, Greenberg DA, Concepcion E, Davies TF**

Division of Endocrinology and Metabolism, Department of Medicine,  
Mount Sinai School of Medicine, New York, USA.  
ytomer@smplink.mssm.edu

The autoimmune thyroid diseases (AITDs) include two related disorders, Graves disease (GD) and Hashimoto thyroiditis, in which perturbations of immune regulation result in an immune attack on the thyroid gland. The AITDs are multifactorial and develop in genetically susceptible individuals. However, the genes responsible for this susceptibility remain unknown. Recently, we initiated a whole-genome linkage study of patients with AITD, in order to identify their susceptibility genes. We studied a data set of 53 multiplex, multigenerational AITD families (323 individuals), using highly polymorphic and densely spaced microsatellite markers (intermarker distance <10 cM). Linkage analysis was performed by use of two-point and multipoint parametric methods (classic LOD-score analysis). While studying chromosome 20, we found a locus on chromosome 20q11.2 that was strongly linked to GD. A maximum two-point LOD score of 3.2 was obtained at marker D20S195, assuming a recessive mode of inheritance and a penetrance of .3. The maximum nonparametric LOD score was 2.4 (P=.00043); this score also was obtained at marker D20S195. Multipoint linkage analysis yielded a maximum LOD score of 3.5 for a 6-cM interval between markers D20S195 and D20S107. There was no evidence for heterogeneity in our sample. In our view, these results indicate strong evidence for linkage and suggest the presence of a major GD-susceptibility gene on chromosome 20q11.2.

PMID: 9837828, UI: 99057513

12: *Cancer Res* 1998 Oct 1;58(19):4260-3      Related Articles, Books, LinkOut

**Chromosomal amplification is associated with cisplatin resistance of human male germ cell tumors.**

**Rao PH, Houldsworth J, Palanisamy N, Murty VV, Reuter VE, Motzer RJ, Bost GJ, Chaganti RS**

Laboratory of Cancer Genetics, Sloan-Kettering Institute for Cancer Research, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

Chemotherapy resistance of tumors is an important biological and clinical problem. Studies from many tumor types have indicated that resistance may be based on multiple genetic pathways. Human male germa cell

SUBSTITUTE SHEET (RULE 26)

FIG. 15

  National Library of Medicine

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM

Search PubMed for 20q11+ [Go]

Clear Limits Preview/Index History Clipboard

Entrez PubMed Show 20 Items 1-20 of 26 Page 1 of 2 Select page: 1 2

PubMed Services Nov;123(1):27-34

□ 1: *Cancer Genet Cytogenet* 2000 Related Articles, Books, LinkOut

Nov;123(1):27-34

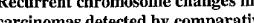
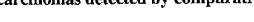
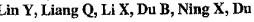
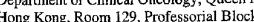
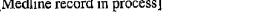
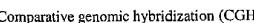
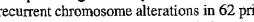
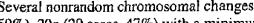
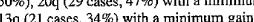
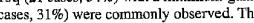
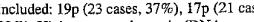
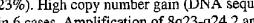
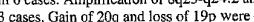
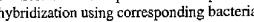
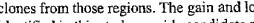
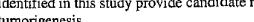
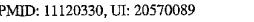
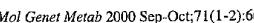
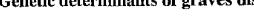
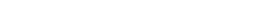
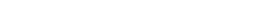
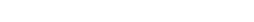
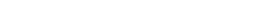
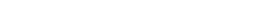
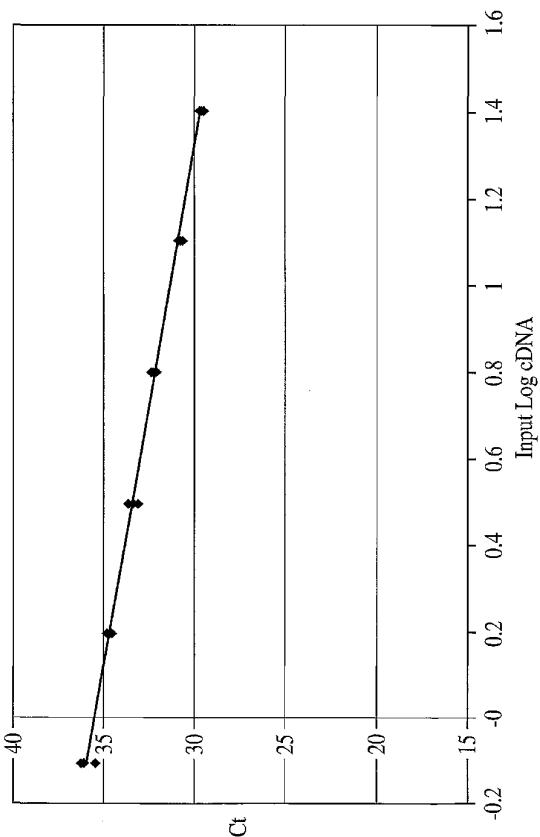
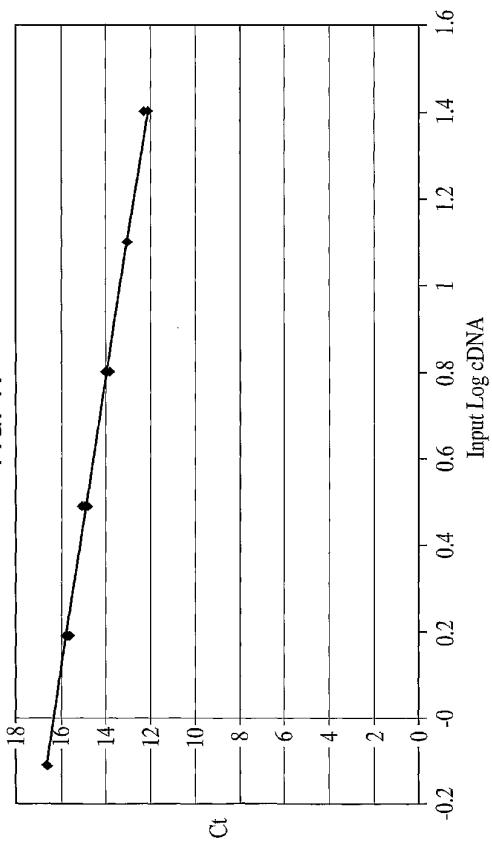
                                                                           <img alt="Elsevier Science logo" data-bbox="278 3751 438 37

FIG. 16



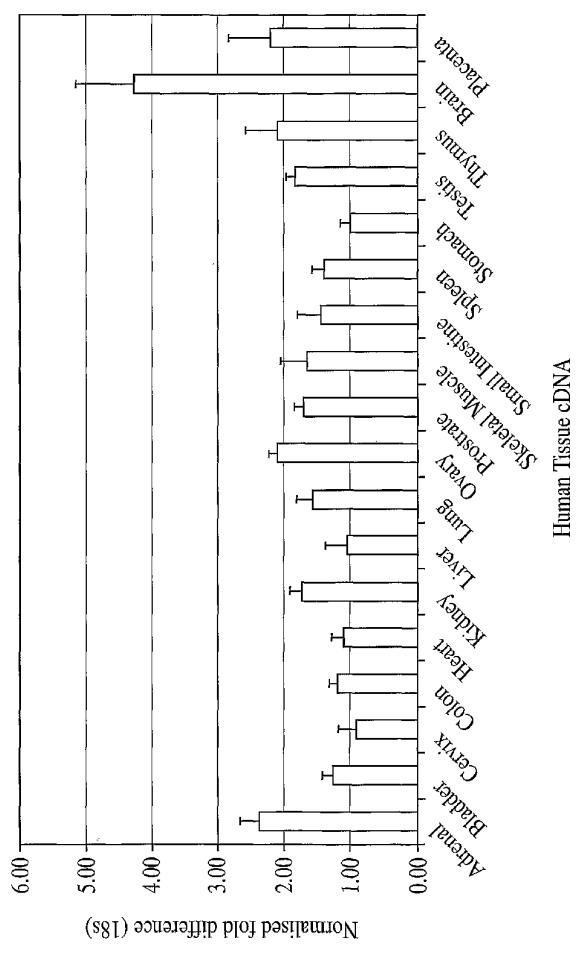
SUBSTITUTE SHEET (RULE 26)

FIG. 17



SUBSTITUTE SHEET (RULE 26)

FIG. 18



SUBSTITUTE SHEET (RULE 26)

## FIG. 19-1

1 LBDG3\_Oiko  
LBDG3\_Ciona  
AAF36514.1  
AAF36515.1  
AAH03931.1  
CAB55933.1  
LEXA\_A

60 LKAN1PE----VEAVLNTDRSLVCDGKKGLIT-RVITVMKEPPDSSEFRWQKAVES-  
LNAN1FE----MEALNNTDRNLVCDGKKGLIT-RLTSVMKKEPIDSSEFRWQKAVES-  
LKAN1PE----VEAVLNTDESLSVCDGKRGGLIT-RLQVMKKEPAESSFRWQKAVES-  
LKAN1PE----VEAVLNTDESLSVCDGKRGGLIT-RLQVMKKEPAESSFRWQKAVES-  
LSPOLEELTTKVSKAHQETFSSLQLGKTYTNSADHRVQDGLWDRESELATKCIKIKI 241

61 FEGPT----CLWDQTFINRGGLDHLILLEIKTTSEVS-QGHEDLAELMKENE 72  
--FERTGPT----NPTTOAFLIDKNTVESTYGVLIISTRPMHELVROGCFDMLSTVMKVNV 69  
--FERTGAT----SYADOMELIKRGGLIEHILFCIIDSCKSRDVQSYFDLGEILMKENI 570  
--FERTGTP----SYADOMELIKRGGLIEHILFCIIDSCKSRDVQSYFDLGEILMKENI 568  
--FERTGTT----SYADOMELIKRGGLIEHILFCIIDSCKSRDVQSYFDLGEILMKENI 265  
--FERTGT----SYADOMELIKRGGLIEHILFCIIDSCKSRDVQSYFDLGEILMKENI 357  
VEPbKRLPSEfdgLIStADQITLIIKAACDILMIRCTRYTPEDDTM---TFSDGLTNR 296

121 a b c d e f g h i 180  
QAEQ--FEKAIGSDAREKRFMAlAENSIVDSNMFRCATLYHKEPLRGY--DEN--- 124  
LBDG3\_Oiko  
LBDG3\_Ciona  
AAF36514.1  
AAF36515.1  
AAH03931.1  
CAB55933.1  
LEXA\_A

122 QAEQ--FEKAIGSDAREKRFMAlAENSIVDSNMFRCATLYHKEPLRGY--DEN--- 124  
DAFTR--XXXXXXXXXXXXAEDSLSDNMFRCVLTHYIOTSOT--ENKEVILV 125  
DSFKR--ENKYVNTDEKFQVFLTQINSSLVDSNMLYRCIVLSDRFENETNDYKVVEFLS 628  
DAFKR--ENKYVNTDEKFQVFLQINSSLVDSNMLYRCIVLSDRFENQV--DMRVAEVILS 626  
DAFKR--ENKYVNTDAKEQFLQINSSLVDSNMLYRCIVLSDRFENQV--DMRVAEVILS 322  
TQMHNAGFPLTDIVFATAGQLPLEMDDETGLSAICLICGDRMDDLE--PERVKDKLOE 355

SUBSTITUTE SHEET (RULE 26)

181	LBDG3_Oiko LBDG3_Ciona AAF36514.1 AAF36515.1 AAH03931.1 CAB55933.1 LEXA_A	K--SKLLHHYMSK--QVRARLVA\$LIQI--TPETLNQENVSCLNTSLVEMITARQ1PNGLG 180 T--NRLLEYYNTN--KRKNKYCKTCNPSR--C\$TLYQENVSCLNTSLVEMITARQ1PNGLG 181 E--CCILISYMARV-ENRNLLEFLRNLINII-INVQTLQENVSCLNTSLVEMITARQ1PNGLG 182 E--CCILISYMAQ-ENRNLLEFLRNLISI--INVQTLQENVSCLNTSLVEMITARQ1PNGLG 183 E--CCILAYISOV-PTQNSFELUNIIT--HVQTLQENVSCLNTSLVEMITARQ1PNGLG 184 E--CCILAYISOV-PTQNSFELUNIIT--HVQTLQENVSCLNTSLVEMITARQ1PNGLG 185 PLLEAFLYARRRSRPSQPMFPRMLK1TDLRG1STKGERAITKME1FG---PMPP 410
241	LBDG3_Oiko LBDG3_Ciona AAF36514.1 AAF36515.1 AAH03931.1 CAB55933.1 LEXA_A	YLSDLS 187 SYLOCY 186 FYNALR 690 LYLSALR 688 LYLRLQ 384 LYLRLQ 476 LIREMEL 417

**SUBSTITUTE SHEET (RULE 26)**

FIG. 19-2

## FIG. 20-1

11bd	( 225 )	sanedMoverIleAElavepk-----te	10	20	30	40	50
21bd	( 182 )	lspqleglitvskahqefp-slcqlg-----kytn-----					
1prga	( 207 )	esadralakLydsvikFpItkakAraItgktidkspfvYdmns					
3erta	( 306 )	lalsitadqfwsaLdAepp-----ilysyedp-----					
1a28a	( 682 )	glippiInLmsiepd-----vivAqRdn-----					
		aaaaaaaaaa					
11bd	( 248 )	tyvseanmglpsspn-----dpvtnicgAadkqftLveWakpIP	60	70	80	90	100
21bd	( 214 )	-----ssadhrvg-----LdlglwdkgselatciklveFakrlp					
1prga	( 255 )	lmmGedikfkhitlpqeqskevAikifqccfrgveavgleiteAksalp					
3erta	( 334 )	trpfs-----easmmgillynLadzelvhminwkrvp					
1a28a	( 706 )	-----tkpd-----ssgulLtslqngLsryvkvwksLp					
		aaaaaaaaaaaaaaaaaaaaaa					
11bd	( 288 )	hFseLplcdDgvilrdragwneLLIAstFshrsia-vk---dG11LatG-lhv	110	120	130	140	150
21bd	( 250 )	gftghsiaDgitlkqAcldlMUR1Ck-----ytpqdtMTfsd0-LtL					
1prga	( 305 )	gFvldinDgvLlkGvHE1IVTmlas1---Mnk---dovLiseCggFm					
3erta	( 366 )	gFvdltlhdqVhLieAv1E1Mig1VwSme---hp-gk11FA---pnLlL					
1a28a	( 738 )	gFrnLhiddqitLidYsWmsLmVFGLGwvSykhvsg-qmLyfa--pdLlL					
		333 aaaaaaaaaaaaaaaaaaaaaaa			bbb	bb	
11bd	( 333 )	hrnsahsaa--gvgaIPdrVlteLvsKmrdMqMdkTELGCLRA1vLFnpds	160	170	180	190	200
21bd	( 295 )	nrtQmHA---G1Gp1TdlVPFAgg11---pL0BdtETG1LSA1C1LcGdr					
1prga	( 349 )	tgeFLkglrkpfqdfmepKFeFavkFp---aleldsD1A1FIAV1ILsgdR					
3erta	( 411 )	drndGkcv-egMve1FdmLlatSsrFr-mmmlLqgeEFVCLKS1ILLnSGV					
1a28a	( 785 )	neqrmk-e-ssFys1C1mugIPqefVv-klqVsqeEPLCmKVLLLNTIp					
		baaaa aaaaaaaaaaaaaaa			aaaaaaa	aaaaaa	
11bd	( 381 )	k-----gLsmpaqVsaLekVyaSLeayGkhkypeq---pgRfaLl	210	220	230	240	250
21bd	( 342 )	m-----dLeepkVdklgepllaLrlyarrrrp---sqpyMfpzM					
1prga	( 398 )	p-----gLnvkipIedigdnLlqLeLqklnhp---essqlfaLl					
3erta	( 459 )	ytFlsst1kgleekdhIhrVldkItdThihMakag1t1qgqghq1agLl					
1a28a	( 832 )	1-----egLrsqtgFemrssYirElkaIgrqkqvssqrifygLT					
		aaaaaaaaaaaaaaaaaaaaaa			aaaaaa	aaaaaa	
11bd	( 420 )	lrlpaLrs1G1kClehlFFF-----kLiGdtpidflmemlea	260	270	280	290	300
21bd	( 381 )	mk1tdLrg1stkgaeraitlkme1p---gpMppliremlenp					
1prga	( 437 )	gkmtdLrq1vtehvqllygv1ktetd---ms1hplLqeykd1					
3erta	( 509 )	l1lsh1ghMSnkghmghlysmkc---knvv---plydlLien1d-----					
1a28a	( 875 )	k1ldmLhd1vkgLh1y1c1nf1qsralsVefPemmgsv1asqLpkilagm					
		a aaaaaaaaaaaaaaaaaaaaaaa			aaaaaa	aaaaaa	
11bd	( 458 )	ph---qmt					
21bd	( )						
1prga	( )	--ahrlha					
3erta	( 546 )						
1a28a	( 925 )	vkpllfhk					

SUBSTITUTE SHEET (RULE 26)

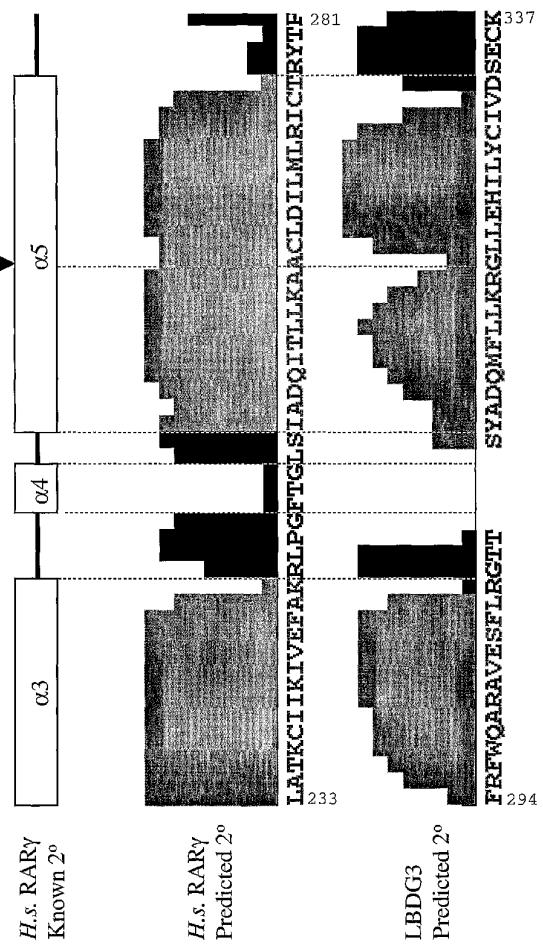
## FIG. 20-2

### Key to Homstrad

aaa	alpha helix
bbb	beta strand
333	$3_{10}$ helix
lower case	solvent accessible
UPPER CASE	solvent inaccessible
<b>Bold</b>	hydrogen bond to main-chain amide
<u>Underline</u>	hydrogen bond to main-chain carbonyl
<i>Italic</i>	positive phi torsion angle

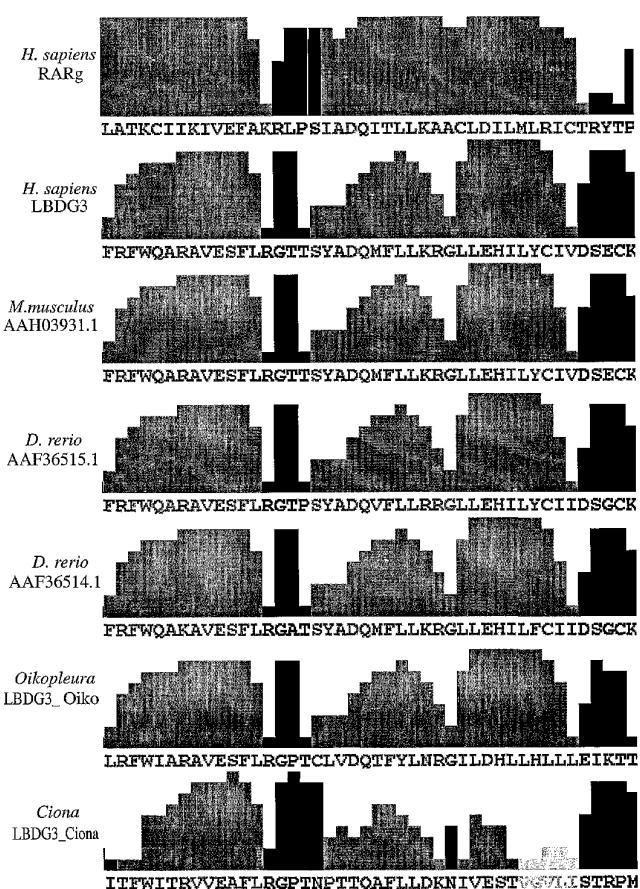
SUBSTITUTE SHEET (RULE 26)

FIG. 21



SUBSTITUTE SHEET (RULE 26)

FIG. 22



SUBSTITUTE SHEET (RULE 26)

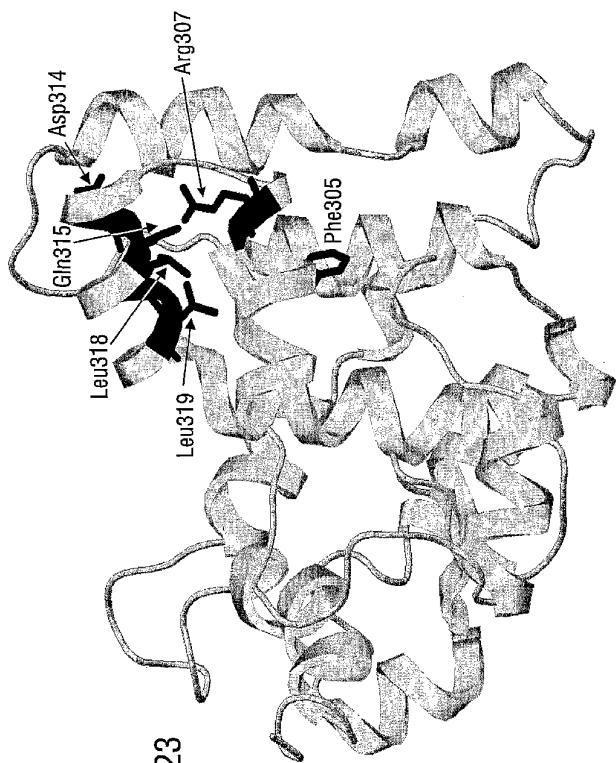
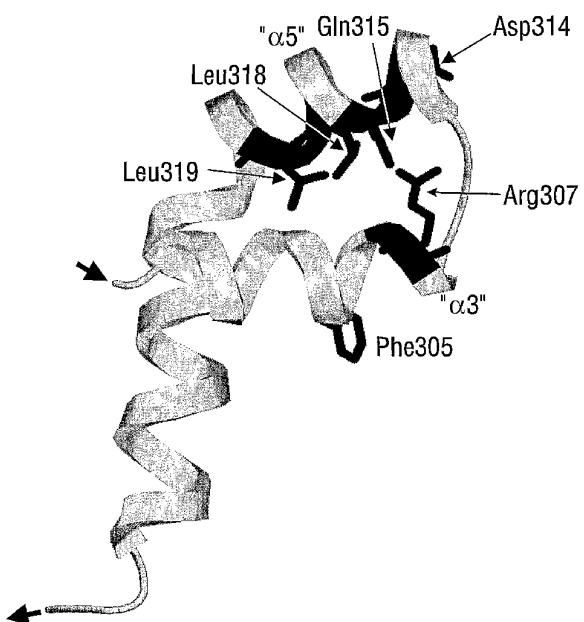


FIG. 23

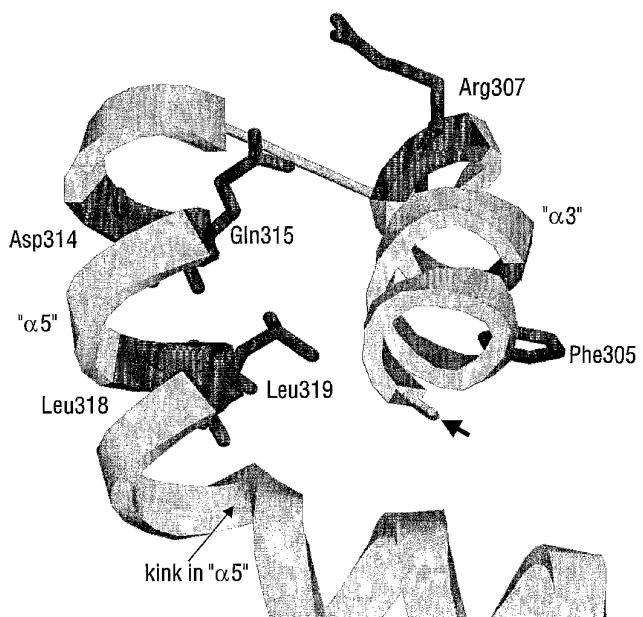
SUBSTITUTE SHEET (RULE 26)

FIG. 24



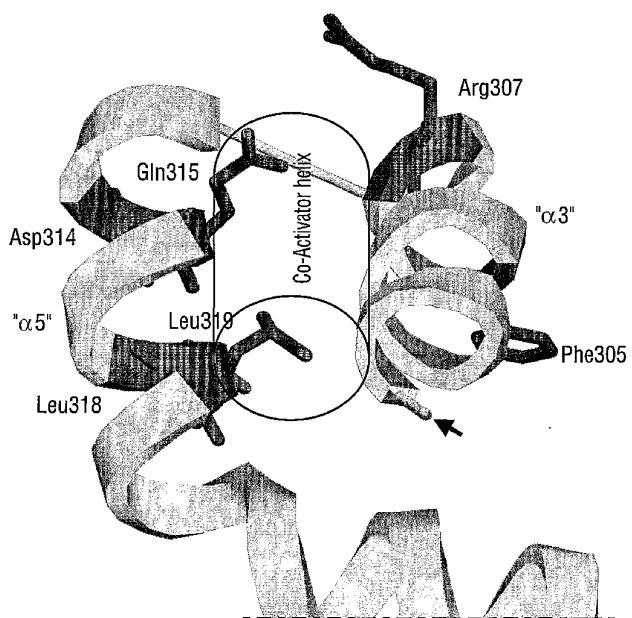
SUBSTITUTE SHEET (RULE 26)

FIG. 25



SUBSTITUTE SHEET (RULE 26)

FIG. 26



SUBSTITUTE SHEET (RULE 26)

WO 02/070557

PCT/GB02/00937

1

## Sequence Listing

SEQ ID NO: 1 Nucleotide coding sequence for CAB55953.1 (LBDG3) protein

```

1 cgtatcgac cagtcgtca atttcgtcc gtttcgtgt gtaccattt cagagatgg
5 61 tacaggaaat gatgacaaacg acacgtttct tgccaaataat gtcacacaga agaaaggatc
121 gatgttgggg ctttcgtcg ctgaaatcaa tcaagggcc cttcgtcgtca ttctgggtt
181 tggttgggggg ctttgcacaa tggcactcg aaagggtca gaggtaacagg gcacaggccag
241 ctttcgtcgtca gatgttgggg agtggtaac acatggctagac aatgtctttgg tgctatgtc
301 ctttcgtcgtca gtggccaaatc aggaggatcaga gcacaaatcaa gtcctccatgg tgtttccctc
10 361 tccaggggct ttttggggata atggcttcc tcacacgtca gacacaaacc agtgtccccca
421 gtcacatggaa attatgtatc agatcatgtca caactggaa gtgttctatgg ttcttcgtgt
481 getgtcgat gggcgctcgc gaaaccaggat tcacagatgg atttgcaggat tcacagatgt
541 ctttcgtcgtca aataatggat ttgtacaaatc gatgttgggg aagcatttcgt catgtccctt
601 tgccatccat ggttcacaaacc agaaatgtca ctgtacccgg gacatcaccc tggaaatata
15 661 gtttttgggg ttcttcgtca gtttcgtca ccacccaggaa aacaggatctt ttgttactcaa
721 caacccaggag ctgtacaaatc tcagtgcacat ctcttcgtca gccaacatccc ctgggggttgg
781 agctgtccat aacaccggaca ggagtttgggt tggtgtatggg aagaggggctt tattaaactcg
841 tctgtgtcgac gtcatggaaaggagccac agatgtgttcc ttccagggtttt ggcacactcg
901 ggttggggat agttttccatc gaggggccacat cccttcgtca gaccatgtat ttctgtgtca
20 961 ggcggggctt ttggggacata ttcttactcg cttttgtggac agccgtgtca agtcaaggaa
1021 tgggtctccat agttttttttccatccctggggatgtatgg aatgttcacatc ttgtatgttccat
1081 caaggatccat aataatataatc tcaacacccgtca tgccaaatgtcc caggatcttgg tggaaaggatc
1141 caacatgtccat cttggggactt ccacccatgtt ggttgcgtgtt cttttatgttccatccggac
1201 atttgaaatccat cttggggat tggaaatgtcc cttttatgttccatccggac
25 1261 ctacatcatcc cttttatgttccatccggac cttttatgttccatccggac
1321 cttttatgttccatccggac cttttatgttccatccggac
1381 cttttatgttccatccggac cttttatgttccatccggac
1441 cttttatgttccatccggac cttttatgttccatccggac
1501 cttttatgttccatccggac cttttatgttccatccggac
30 1561 cttttatgttccatccggac cttttatgttccatccggac
1621 cttttatgttccatccggac cttttatgttccatccggac
1681 cttttatgttccatccggac cttttatgttccatccggac
1741 cttttatgttccatccggac cttttatgttccatccggac
1801 cttttatgttccatccggac cttttatgttccatccggac
35 1861 cttttatgttccatccggac cttttatgttccatccggac

```

**SUBSTITUTE SHEET (RULE 26)**

WO 02/070557

PCT/GB02/00937

2

1921 ttcagaaccc tggggcccg ggcacccagg aggttaagagg acattttagca ttatgtctgt  
1981 gtgagctctt gccgggtttc tggctgtca tgatgtcccg agtggggagg agatatatgg  
2041 tgaccccccggc cccccatctg tgagccaage ctccctgtc cctggccctt ggacccaggc  
2101 aaaggcttct gaggccctggg caagggttggt ggttacccaga gaatgtctgc tttcccccaga  
5 2161 ctcgccttc tgcctcattt tccctgttgtc ctctgggtt tggttgtctca ttgggtgtgt  
2221 tggccatccca agggggtttt cccagaatgt agggggccctt ctccctccatcc ctggggccac  
2281 ggggcagctt tgcctgtccct ggctctgtct gaggccagccg ctctgtctgg aacccggaca  
2341 tggggccctt ctctgttggt ccaattttt aacccaaat aaaccaattt aatggagact  
2401 attaaataac ttatattttt aaaaaaaaa

10

SEQ ID NO: 2 Protein sequence for CAB55953.1 (LBDG3)

1 dgqqlanfler ilavtisemad tgnndkhtll akmaqqkksl slgspaaein qaallslippgf  
61 verlcklatr kvsestgtas flqelsewyt wlndnalvldas lmrvaneese hngasivfpp  
121 pgasenengl htsartqlpg smkimheimy klevlyvclv llmqrgrqnv hrmiaeckli  
15 181 pglnmlfdkl iwrkhnasal vlhghngnqd csgdilktiq flrlqgsfd hhenkylln  
241 ngelnelsai slkanipeva avlntdrslv cdgkrglltr llqvmkcepq essfrfwqar  
301 avesfrlrgt syadgmlflk rgllehhilc ivdseckard vlgysyfdllg elmkfnvdaf  
361 krfnkyintd akfqvflkq nsslvdsmnl vrcvtlsldr fengvdmkva evlsecrllla  
421 yisqvpqtmgs filfrlinihih vgltqenvs clntslvilm larrkerpl ylrlnlqrmeh  
20 481 slkkygfln nfhnllrflwg qhylhdkdks tclensscis fsywketvsi llnpdrgsp  
541 alsvsiepy mdidrftee

SEQ ID NO: 3; Nucleotide accession number AL132825. Note that SEQ ID NO:1 contains 3'UTR sequence not present in SEQ ID NO:3.

**SUBSTITUTE SHEET (RULE 26)**

WO 02/070557

PCT/GB02/00937

3

**SUBSTITUTE SHEET (RULE 26)**

SEQ ID NO: 4; Protein accession number CAC14946.1.

1 maaapvaags gagrgrrsaa tvaawggwggrg rprpgnillq lrqgqltqrg lvravqftet  
61 filterdkgsk wsgipqlllk lhttshlhsd fvecqnilke isplismean afvteerklt  
121 gettypntyi fdlfgggvdll veilmrptis irgqklkisd emskdclsil yntcovctegv  
5 181 tkrlaekndf viiflilmk kktflqatf iedilgvkke mzrldevpn1 ss1vsnfdqg  
241 qlanfrila vtisemdgr addkhtllakn aqgkkslslg psaeninqaa llsipgfver  
301 lcklatrkv sestgtafsflg eleewytwld nalvldalmr vanseehng asivfpppg  
361 seenglphts artqlpqsmk imheimykle vlyvlcvilm grqrnqvhrm iaeefklipgl  
421 nnlfdfkliwr khsasalv1h ghnqncdcsp ditlkicqflr llqsfssdhhe nkyllnnqe  
10 481 lnelsaislk anipeveavl ntdrs1vcg krglltrllq vnkkepaess frfwqarave  
541 sflrgttsya dqmfilkrgl lehilycivd secksrdv1q syfdllgelm kfnvdafkrf  
601 nkyintdakf qvflkginss lvdsnmlvrc vtlsldrfen qvdmkvaevl seerllayis  
661 qvptgmsflf rliniinhvqt ltgenyscln ts1vilmr rkerlplylx llqrmehskk  
721 ypgfilnnfh nllrfwqghy lhkdkdstcl ensscisfsy wketsvilln pdrgspsalv  
15 781 syisepymdi drdftee

SUBSTITUTE SHEET (RULE 26)

## 【国際公開パンフレット（コレクトバージョン）】

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

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number  
WO 02/070557 A3

(51) International Patent Classification: C07K 14/705, 14/72, A61K 38/17

(74) Agents: MERCER, Christopher, Paul et al.; Carpmaels &amp; Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States (national): AE, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CI, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, H, GB, GD, GE, GI, GM, HR, HU, ID, IL, IN, IS, JP, KU, KG, KP, KR, KZ, L, I, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TI, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, L, S, MW, MZ, SD, SI, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TI, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PL, SE, TR), OAPI patent (BJ, BI, CH, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report: 5 December 2002

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.

(72) Inventors; and

(75) Inventors/Applicants (for US only): FAGAN, Richard, Joseph [US/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). PHELPS, Christopher, Benjamin [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). PHILLIPS, Tom [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). PIERRON, Valerie, Nathalie [FR/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). ALLEN, Kathryn, Elizabeth [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). ALLEN, Janet, Marjorie [GB/GB]; Inpharmatica Limited, 60 Charlotte Street, London W1T 2NU (GB). POTTER, Sarah, Jane [GB/GB]; 39 Kneller Road, Brockley, London SE4 2AR (GB).



A3

WO 02/070557

(54) Title: NUCLEAR HORMONE RECEPTOR LIGAND BINDING DOMAIN

(57) Abstract: This invention relates to a novel protein, termed LBDG3, herein identified as a Nuclear Hormone Receptor Ligand Binding Domain and to the use of this protein and nucleic acid sequence from the encoding genes in the diagnosis, prevention and treatment of disease.

## 【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No PCT/GB 02/00937
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 C07K14/705 C07K14/72 A61K38/17		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 C07K		
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) BIOSIS, EPO-Internal, EMBL, SEQUENCE SEARCH, WPI Data, PAJ, MEDLINE		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 712 381 A (SCHIEVELLA ANDREA R ET AL) 27 January 1998 (1998-01-27) Sequences with 99.9% identity with SEQ ID Nos:1 and 3 over 1750 and 2394 nucleotides respectively and 99.8% identity with SEQ ID Nos:2 and 4 over 560 and 797 amino acids respectively column 83, line 1 -column 90, line 14 ---- ----	1-14,16, 20-46
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		<input checked="" type="checkbox"/> Patent family members are listed in annex.
* Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubt 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 award made prior to the international filing date *P* document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 23 September 2002		Date of mailing of the international search report 10/10/2002
Name and mailing address of the ISA European Patent Office, P.B. 5018 Patenttaan 2 NL - 2200 HJ Haarlem Tel: (+31-70) 340-2040, Tx: 31 651 epo nl Fax: (+31-70) 340-3016		Authorized officer Nichogianopoulos, A

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

INTERNATIONAL SEARCH REPORT		International Application No PCT/GB 02/00937
C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Creation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE 'Online! 15 September 1999 (1999-09-15) POUSTKA A ET AL: "Homo sapiens mRNA; cDNA DKFZp727M231" retrieved from EMBL Database accession no. AL117480 XP002214288 cited in the application Sequences with 100% identity with SEQ ID Nos:1 and 2 over 2428 nucleotides and 560 amino acids respectively. the whole document</p> <p>-----</p> <p>DATABASE 'Online! 7 November 1999 (1999-11-07) MOORE M: "Human DNA sequence from clone RP4-756N5 on chromosome 20q11.21-12. Contains gene KIAA1512 for a novel protein similar to myosin heavy chain, the gene encoding a novel protein similar to Trp4-associated protein TAP1 (ABCB2)" retrieved from EMBL Database accession no. AL132825 XP002214289 cited in the application Sequences with 100% identity with SEQ ID Nos:3 and 4 over 2394 nucleotides and 797 amino acids respectively the whole document</p> <p>-----</p> <p>LAMOUR FRANCOIS P Y ET AL: "Analysis of the ligand-binding domain of human retinoic acid receptor alpha by site-directed mutagenesis." MOLECULAR AND CELLULAR BIOLOGY, vol. 16, no. 10, 1996, pages 5386-5392, XP002214287 ISSN: 0270-7306 figure 1</p> <p>-----</p>	1-14
A		1

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

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/GB 02/00937

## 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:  
see FURTHER INFORMATION sheet PCT/ISA/210
2.  Claims Nos.: 15, 17-19 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:

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.:

## 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/GB 02 00937

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

## Continuation of Box I.1

Although claims 21, 23-29 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claims 32, 36-39 and 46 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

## Continuation of Box I.2

Claims Nos.: 15, 17-19

Present claims 15 and 17-19 relate to compounds defined by reference to a desirable characteristic or property, namely their ability to bind to a claimed polypeptide and their ability to increase or decrease the level of expression or activity of a claimed polypeptide.

The claims cover all compounds having this characteristic or property, whereas the application provides no support within the meaning of Article 6 PCT or disclosure within the meaning of Article 5 PCT for any such compounds. 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 compounds 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 said claims.

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

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/GB 02/00937

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5712381	A 27-01-1998	US 5843675 A US 5852173 A US 5849501 A AU 2268697 A WO 9730084 A1 US 6322972 B1 US 5891675 A AU 698575 B2 AU 3826195 A EP 0787145 A1 JP 10507068 T WO 9612735 A1 US 5948638 A US 5847099 A	01-12-1998 22-12-1998 15-12-1998 02-09-1997 21-08-1997 27-11-2001 06-04-1999 05-11-1998 15-05-1996 06-08-1997 14-07-1998 02-05-1996 07-09-1999 08-12-1998

Form PCT/ISA/210 (patent family annex) (July 1992)

## フロントページの続き

(51)Int.Cl. <sup>7</sup>	F I	テーマコード(参考)
A 6 1 K 45/00	A 6 1 K 45/00	4 C 0 8 4
A 6 1 K 48/00	A 6 1 K 48/00	4 C 0 8 5
A 6 1 P 3/00	A 6 1 P 3/00	4 C 0 8 6
A 6 1 P 7/00	A 6 1 P 7/00	4 H 0 4 5
A 6 1 P 9/00	A 6 1 P 9/00	
A 6 1 P 9/10	A 6 1 P 9/10	
A 6 1 P 13/12	A 6 1 P 13/12	
A 6 1 P 19/00	A 6 1 P 19/00	
A 6 1 P 25/00	A 6 1 P 25/00	
A 6 1 P 29/00	A 6 1 P 29/00	
A 6 1 P 31/00	A 6 1 P 31/00	
A 6 1 P 31/18	A 6 1 P 31/18	
A 6 1 P 35/00	A 6 1 P 35/00	
A 6 1 P 37/02	A 6 1 P 37/02	
A 6 1 P 43/00	A 6 1 P 43/00	1 0 5
C 0 7 K 14/72	C 0 7 K 14/72	
C 0 7 K 16/28	C 0 7 K 16/28	
C 1 2 M 1/00	C 1 2 M 1/00	A
C 1 2 N 1/15	C 1 2 N 1/15	
C 1 2 N 1/19	C 1 2 N 1/19	
C 1 2 N 1/21	C 1 2 N 1/21	
C 1 2 N 5/10	C 1 2 Q 1/68	A
C 1 2 Q 1/68	G 0 1 N 33/15	Z
G 0 1 N 33/15	G 0 1 N 33/50	Z
G 0 1 N 33/50	C 1 2 N 15/00	F
	C 1 2 N 5/00	A
	A 6 1 K 37/02	

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

(74)代理人 100084009

弁理士 小川 信夫

(74)代理人 100082821

弁理士 村社 厚夫

(74)代理人 100086771

弁理士 西島 孝喜

(74)代理人 100084663

弁理士 箱田 篤

(72)発明者 ファガン リチャード ジョセフ

イギリス ロンドン ダブリュ1ティー 2エヌユー チャーロット ストリート 60 インフ  
アーマティカ リミテッド

(72)発明者 フェルプス クリストファー ベンジャミン

イギリス ロンドン ダブリュ1ティー 2エヌユー チャーロット ストリート 60 インフ

アーマティカ リミテッド

(72)発明者 フィリップス トム

イギリス ロンドン ダブリュ1ティー 2エヌユー チャーロット ストリート 60 インフ  
アーマティカ リミテッド

(72)発明者 ピアロン ヴァレリー ナサリー

イギリス ロンドン ダブリュ1ティー 2エヌユー チャーロット ストリート 60 インフ  
アーマティカ リミテッド

(72)発明者 アレン キャスリン エリザベス

イギリス ロンドン ダブリュ1ティー 2エヌユー チャーロット ストリート 60 インフ  
アーマティカ リミテッド

(72)発明者 アレン ジャネット マジョリー

イギリス ロンドン ダブリュ1ティー 2エヌユー チャーロット ストリート 60 インフ  
アーマティカ リミテッド

(72)発明者 ポッター サラ ジエイン

イギリス ロンドン エスイー4 2エイアール プロックリー ネラー ロード 39

F ターム(参考) 2G045 AA34 DA36 FB03

4B024 AA01 AA11 BA63 CA04 DA02 DA05 DA06 DA11 DA12 EA04  
GA11 HA12  
4B029 AA07 AA23 BB20 CC03 FA15  
4B063 QA01 QA18 QA19 QQ03 QQ08 QQ43 QR08 QR42 QR56 QS25  
QS34 QX02  
4B065 AA01X AA57X AA72X AA90X AA90Y AB01 BA02 CA24 CA44 CA46  
4C084 AA02 AA06 AA07 AA13 AA17 BA01 BA08 BA22 BA23 CA17  
MA01 NA14 ZA022 ZA032 ZA052 ZA082 ZA122 ZA152 ZA162 ZA362  
ZA372 ZA402 ZA422 ZA452 ZA512 ZA542 ZA592 ZA662 ZA702 ZA812  
ZA832 ZA892 ZA942 ZA962 ZA972 ZB072 ZB112 ZB132 ZB212 ZB262  
ZB322 ZB332 ZB352 ZB372 ZC062 ZC212 ZC332 ZC352 ZC412  
4C085 AA13 AA14 CC32 DD88 EE01 GG01  
4C086 AA01 AA02 AA03 AA04 EA16 MA01 MA04 NA14 ZA02 ZA03  
ZA05 ZA08 ZA12 ZA15 ZA16 ZA36 ZA37 ZA40 ZA42 ZA45  
ZA51 ZA54 ZA59 ZA66 ZA81 ZA83 ZA89 ZA94 ZA96 ZB07  
ZB11 ZB13 ZB21 ZB26 ZC21 ZC35  
4H045 AA10 AA11 AA20 AA30 BA10 CA40 DA50 DA75 EA20 EA50  
FA72 FA74