(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2004-500111
(P2004-500111A)
(43)公表日　平成16年1月8日(2004.1.8)

(51)Int.Cl.[7]	F I			テーマコード（参考）
C12N 15/09	C12N	15/00	ZNAA	2G045
C07K 14/47	C07K	14/47		4B024
C07K 16/18	C07K	16/18		4B063
C07K 19/00	C07K	19/00		4B064
C12N 1/15	C12N	1/15		4B065

審査請求 未請求　予備審査請求 有　（全 118 頁）　最終頁に続く

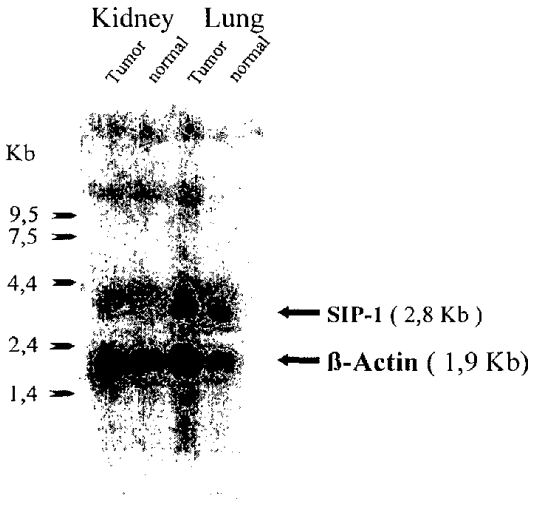
(21)出願番号　特願2001-570731(P2001-570731)
(86)(22)出願日　平成13年3月21日(2001.3.21)
(85)翻訳文提出日　平成14年9月24日(2002.9.24)
(86)国際出願番号　PCT/EP2001/003204
(87)国際公開番号　WO2001/073014
(87)国際公開日　平成13年10月4日(2001.10.4)
(31)優先権主張番号　00106408.8
(32)優先日　平成12年3月24日(2000.3.24)
(33)優先権主張国　欧州特許庁(EP)
(81)指定国　EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),CA,JP,US

(71)出願人　591032596
メルク　パテント　ゲゼルシャフト　ミット　ベシュレンクテル　ハフトング
Ｍｅｒｃｋ　Ｐａｔｅｎｔ　Ｇｅｓｅｌｌｓｃｈａｆｔ　ｍｉｔ　ｂｅｓｃｈｒａｅｎｋｔｅｒ　Ｈａｆｔｕｎｇ
ドイツ連邦共和国　デー－６４２９３　ダルムシュタット　フランクフルター　シュトラーセ　２５０
Ｆｒａｎｋｆｕｒｔｅｒ　Ｓｔｒ．　２５０，Ｄ－６４２９３　Ｄａｒｍｓｔａｄｔ，Ｆｅｄｅｒａｌ　Ｒｅｐｕｂｌｉｃ　ｏｆ　Ｇｅｒｍａｎｙ
(74)代理人　100123788
弁理士　宮崎　昭夫

最終頁に続く

(54)【発明の名称】ヒトサバイビン相互作用タンパク質１（ＳＩＰ－１）

(57)【要約】
サバイビン相互作用タンパク質１およびポリヌクレオチドならびにそのようなポリペプチドを組換え技術によって製造するための方法が開示される。ＳＩＰ－１リガンドのポリペプチドおよびポリヌクレオチドを診断アッセイにおいて用いるための方法もまた開示される。



【特許請求の範囲】
【請求項１】
（ａ）配列番号１の配列を含むポリヌクレオチドによってコードされるポリペプチド、
（ｂ）配列番号２のポリペプチド配列に対して少なくとも９５％の同一性を有するポリペプチド配列を含むポリペプチド、
（ｃ）配列番号２のポリペプチド配列に対して少なくとも９５％の同一性を有するポリペプチド、
（ｄ）配列番号２のポリペプチド配列、および
（ｅ）（ａ）～（ｄ）に記載のかかるポリペプチドのフラグメントおよび変異体
からなる群から選択されるポリペプチド。
【請求項２】
配列番号２のポリペプチド配列を含む、請求項１に記載のポリペプチド。
【請求項３】
配列番号２のポリペプチド配列である、請求項１に記載のポリペプチド。
【請求項４】
（ａ）配列番号１のポリヌクレオチド配列に対して少なくとも９５％の同一性を有するポリヌクレオチド配列を含むポリヌクレオチド、
（ｂ）配列番号１のポリヌクレオチドに対して少なくとも９５％の同一性を有するポリヌクレオチド、
（ｃ）配列番号２のポリペプチド配列に対して少なくとも９５％の同一性を有するポリペプチド配列をコードするポリヌクレオチド配列を含むポリヌクレオチド、
（ｄ）配列番号２のポリペプチド配列に対して少なくとも９５％の同一性を有するポリペプチド配列をコードするポリヌクレオチド配列を有するポリヌクレオチド、
（ｅ）配列番号１の配列または少なくとも１５塩基長を有するそのフラグメントを有する標識されたプローブを用いた厳格なハイブリダイゼーション条件下でライブラリをスクリーニングすることにより得られる少なくとも１００塩基長の塩基配列を有するポリヌクレオチド、
（ｆ）（ａ）～（ｅ）のポリヌクレオチドのＲＮＡ等価体であるポリヌクレオチド、
（ｇ）（ａ）～（ｆ）のいずれかの前記ポリヌクレオチドに対して相補的なポリヌクレオチド配列、および
（ｈ）（ａ）～（ｇ）のいずれかのポリヌクレオチドの変異体またはフラグメントであるか、あるいは前記のポリヌクレオチドに対してその全長にわたって相補的であるポリヌクレオチド
からなる群から選択されるポリヌクレオチド。
【請求項５】
（ａ）配列番号１のポリヌクレオチドを含むポリヌクレオチド、
（ｂ）配列番号１のポリヌクレオチド、
（ｃ）配列番号２のポリペプチドをコードするポリヌクレオチド配列を含むポリヌクレオチド、および
（ｄ）配列番号２のポリペプチドをコードするポリヌクレオチド
からなる群から選択される、請求項４に記載のポリヌクレオチド。
【請求項６】
該発現ベクターが適合性宿主細胞に存在する際、請求項１～３のいずれかに記載のポリペプチドを産生し得るポリヌクレオチドを含む発現システム。
【請求項７】
請求項１～３のいずれかに記載のポリペプチドを発現する、請求項６に記載される発現ベクターを含む組換え宿主細胞またはその膜。
【請求項８】
請求項１～３のいずれかに記載のポリペプチドを製造する方法であって、請求項７に記載の宿主細胞を前記ポリペプチドの産生に十分な条件下で培養して、該ポリペプチドを該培

養培地から回収する工程を含む方法。
【請求項９】
免疫グロブリンのＦｃ領域と請求項１～３のいずれかに記載のポリペプチドとからなる融合タンパク質。
【請求項１０】
請求項１～３のいずれかに記載のポリペプチドに対して免疫特異的な抗体。
【請求項１１】
請求項１～３のいずれかに記載のポリペプチドの機能または濃度を刺激または阻害する化合物を同定するためのスクリーニング方法であって、
（ａ）該ポリペプチド（または該ポリペプチドを発現する細胞もしくは膜）またはその融合タンパク質に対する候補化合物の結合を、該候補化合物に直接的または間接的に結合している標識によって定量的または定性的に測定または検出すること、
（ｂ）該ポリペプチド（または該ポリペプチドを発現する細胞もしくは膜）あるいはその融合タンパク質に対する候補化合物の結合の競合を、標識された競合剤の存在下で測定すること、
（ｃ）該ポリペプチドの活性化または阻害により発生するシグナルを該候補化合物がもたらすか否かを、該ポリペプチドを発現する細胞または細胞膜に適切な検出システムを使用して試験すること、
（ｄ）候補化合物を、請求項１～３のいずれかに記載のポリペプチドを含有する溶液と混合して混合物を作製し、該混合物中の該ポリペプチドの活性を測定し、そして、何らの候補化合物を含有しない対照混合物に対して、該混合物の活性を比較すること、または
（ｅ）細胞中における前記ポリペプチドをコードするｍＲＮＡまたは前記ポリペプチドの産生に対する候補化合物の作用を、例えばＥＬＩＳＡアッセイを使用して検出すること、からなる群から選択される方法と、
（ｆ）生物工学的または化学的な標準的な技術に従って前記化合物を製造すること、
を含む方法。
【発明の詳細な説明】
【０００１】
（発明の分野）
本発明は、以降、時にサバイビン相互作用タンパク質１（Ｓｕｒｖｉｖｉｎ　Ｉｎｔｅｒａｃｔｉｎｇ　Ｐｒｏｔｅｉｎ　１）または「ＳＩＰ－１」と称する、新たに同定されたポリペプチド、およびかかるポリペプチドをコードするポリヌクレオチド、診断ならび治療上潜在的に有用となるアゴニスト、アンタゴニストとなり得る化合物を同定する際のそれらの使用、ならびにかかるポリペプチドおよびポリヌクレオチドの製造とに関する。
【０００２】
（発明の背景）
創薬プロセスは現在、「機能ゲノム科学」、すなわちハイ・スループットのゲノムまたは遺伝子に基づく生物学を取り入れ、根本的改革が進められている。この方法は、治療の標的として遺伝子および遺伝子産物を同定する手段として、迅速に、「ポジショナルクローニング」に基づいたより初期の方法の代わりになりつつある。生物機能または遺伝子の疾患である表現型を同定し、次いでこれに対する原因遺伝子がその遺伝子地図の位置に基づいて突き止められる。
【０００３】
機能ゲノム科学は、ハイ・スループットＤＮＡ配列決定技術、および現在入手可能な多くの分子生物学データベースから、潜在的に注目される遺伝子配列を同定するためにバイオインフォマティクスの様々なツールに大きく依存している。創薬の標的として、さらなる遺伝子およびその関連するポリペプチド／タンパク質を同定し、その特定を行うことが引き続き求められている。この領域における最近の進歩は、新しく同定されたタンパク質の機能に注目して、進歩したスクリーニングシステムを適用することによって主に進められている。このタイプのスクリーニングに関して興味深い適用は腫瘍のアポトーシスの分野

においてである。

【０００４】

腫瘍細胞のアポトーシスは、例えば、よく知られているｐ５３腫瘍抑制因子タンパク質において頻繁に見出される変異のような、重要な遺伝子における変異によって阻止され得る。アポトーシスを阻止する別の機構が、ＤｅｖｅｒａｕｘおよびＲｅｅｄ（Ｇｅｎｅｓ＆　Ｄｅｖｅｌｏｐｍｅｎｔ、１９９９）により総説される「アポトーシス阻害剤」（ＩＡＰ）のタンパク質ファミリーによって引き起こされている。このファミリーの１つのメンバーがサバイビン（Ｓｕｒｖｉｖｉｎ）遺伝子によってコードされる。

【０００５】

ガンにおけるＩＡＰ関与に対する最も有力な証拠がサバイビンについて見出されている（ＡｌｔｉｅｒｉおよびＭａｒｃｈｉｓｉｏ（Ｌａｂｏｒａｔｏｒｙ　Ｉｎｖｅｓｔｉｇａｔｉｏｎ、１９９９）の総説を参照）。成体の分化した組織では認められていないが、サバイビンは、形質転換された細胞株において、そして肺、結腸、膵臓、前立腺および乳房の最も一般的なヒトのガンのすべてにおいてインビボで顕著に発現している。サバイビンはまた、進行度が大きい非ホジキンリンパ腫（胚中心細胞性、免疫芽細胞性）の約５０％において見出されているが、進行度が低いリンパ腫（リンパ球性）では見出されていない。サバイビンは、Ｆａｓ（ＣＤ９５）、Ｂａｘ、様々なカスパーゼおよび抗ガン薬によって誘導されるカスパーゼ活性およびアポトーシスを阻害する。

【０００６】

さらに、サバイビンは、細胞周期のＧ２／Ｍ期において４０倍アップレギュレーションされ、有糸分裂の紡錘体（微小管）に結合するが、紡錘体におけるその役割は依然として不明である。アポトーシスおよび有糸分裂紡錘体チェックポイントのサバイビンによる連携した制御が存在し得る。サバイビン－微小管の相互作用の破壊は、サバイビンの抗アポトーシス機能を失わせ、カスパーゼ－３の活性を増大させる。サバイビンは、Ｇ２／Ｍ期におけるアポトーシスの不十分な（ｄｅｆａｕｌｔ）誘導に対抗して作用し得る。ガンにおけるサバイビンの過剰発現は、このアポトーシスチェックポイントを乗り越えることができ、そして形質転換された細胞の有糸分裂の異常な進行に加担し得る（Ｆｅｎｇｚｈｉ他、Ｎａｔｕｒｅ、１９９８）。これに関連して、Ｆａｓにより媒介される細胞死を阻止するためにＣｄｋ４と相互作用する結果として、サバイビンにより、プロカスパーゼ３／ｐ２１複合体の形成が開始されることが最近示されたことに注目することは重要である（Ｓｕｚｕｋｉ他、Ｏｎｃｏｇｅｎｅ、２０００）。従って、ＳＩＰ－１は免疫系からの知られている腫瘍細胞逃避に関与し得る。サバイビンがその抗アポトーシス活性を媒介し得るその他の生化学的機能は、現在、様々なカスパーゼに対するその阻害的な結合を除き、明らかにされていない。

【０００７】

したがって本発明において、本発明者らはサバイビンと相互作用する潜在的なリガンドについてスクリーニングした。

【０００８】

（発明の概要）

本発明は、サバイビン相互作用タンパク質１（＊遺伝子名）に関し、特には＊遺伝子名ポリペプチドおよび＊遺伝子名ポリヌクレオチド、組換え体、ならびにそれらの製造方法に関する。かかるポリペプチドおよびポリヌクレオチドは、サバイビンに対する天然のリガンドであるので、ある種の疾患を処置する方法に関連して注目されている。指摘されているように、この分子は、抗アポトーシス活性における顕著な役割を有する。

【０００９】

サバイビンは、肺、肝臓、結腸、胃、皮膚、膵臓、前立腺、卵巣および乳房の非常に一般的なヒトのガンにおいてインビボで顕著に発現している。サバイビンはまた、進行度が高い非ホジキンリンパ腫の約５０％において見出されており、したがって、本発明の新しく記載される分子ＳＩＰ－１との相互作用は、これらの疾患を妨げる注目される選択肢を提供する。従って、それらは以降「本発明の疾患」と呼ばれる。さらなる態様において、本

発明は、本発明により提供される材料を用いてアゴニストおよびアンタゴニスト（例えば阻害剤）を同定する方法、ならびに同定された化合物を用いて、＊遺伝子名の活性に関連した状態を処置する方法に関する。さらにさらなる態様において、本発明は、＊遺伝子名の不適切な活性および濃度に付随する疾患を検出する診断アッセイに関する。

【００１０】

（発明の説明）

第一の態様において、本発明は＊遺伝子名ポリペプチドに関する。かかるポリペプチドには下記が含まれる：

（ａ）配列番号１の配列を含むポリヌクレオチドによってコードされるポリペプチド、

（ｂ）配列番号２のポリペプチド配列に対して少なくとも９５％、９６％、９７％、９８％または９９％の同一性を有するポリペプチド配列を含むポリペプチド、

（ｃ）配列番号２のポリペプチド配列を含むポリペプチド、

（ｄ）配列番号２のポリペプチド配列に対して少なくとも９５％、９６％、９７％、９８％または９９％の同一性を有するポリペプチド、

（ｅ）配列番号２のポリペプチド配列、および

（ｆ）配列番号２のポリペプチド配列と比較して、０．９５、０．９６、０．９７、０．９８または０．９９の同一性指標を有するポリペプチド配列を有するか、またはそのようなポリペプチド配列を含むポリペプチド、

（ｇ）（ａ）～（ｆ）に記載のかかるポリペプチドのフラグメントおよび変異体。

【００１１】

本発明のポリペプチドは、サバイビンに対するリガンドであると考えられ、従って、アポトーシスの不十分な誘導に対抗して作用することができる。さらに、サバイビンおよびそのリガンドＳＩＰ－１はアポトーシスチェックポイントを制御することができ、そしてＳＩＰ－１はサバイビンの抗アポトーシス活性の媒介因子であり得る。

【００１２】

ＳＩＰ－１ポリペプチドは、免疫細胞のシグナルのような様々なアポトーシスシグナル、例えば、化学療法剤または生理学的な死シグナルに対する腫瘍細胞の抵抗性をもたらし得る生存シグナルを腫瘍細胞が確立することに関与すると考えられる。

【００１３】

＊遺伝子名の生物学的性質は、以降、「＊遺伝子名の生物学的活性」または「＊遺伝子名活性」と称する。好ましくは、本発明のポリペプチドは少なくとも１つの＊遺伝子名の生物学的活性を示す。

【００１４】

本発明のポリペプチドには、すべての対立遺伝子形態およびスプライス変異体を含む上記ポリペプチドの変異体もまた含まれる。かかるポリペプチドは、挿入、欠失、ならびに保存的または非保存的であり得る置換、あるいはそれらの任意の組合せによって基準ポリペプチドから変異している。特に好ましい変異体は、いくつかの、例えば、５０～３０、３０～２０、２０～１０、１０～５、５～３、３～２、２～１または１個のアミノ酸が任意の組合せで挿入、置換または欠失されている変異体である。

【００１５】

本発明のポリペプチドの好ましいフラグメントには、配列番号２のアミノ酸配列に由来する少なくとも３０、５０または１００個の連続したアミノ酸を有するアミノ酸配列を含むポリペプチド、あるいは配列番号２のアミノ酸配列から、少なくとも３０、５０または１００個の連続したアミノ酸が短縮化または欠失しているアミノ酸配列を含むポリペプチドが含まれる。好ましいフラグメントは、ＡＮＩＣ－ＢＰリガンドの生物学的活性をもたらす生物学的に活性なフラグメントである。これには、類似する活性または改善された活性を有するか、あるいは望ましくない活性が低下しているフラグメントが含まれる。また、動物、特にヒトにおいて抗原性または免疫原性であるそのようなフラグメントも好ましい。

【００１６】

本発明のポリペプチドのフラグメントは、対応する完全長型ポリペプチドをペプチド合成によって製造するために用いることができる。したがって、これらの変異体は本発明の完全長型ポリペプチドを製造するための中間体として用いることができる。本発明のポリペプチドは、「成熟型」タンパク質の形態であってもよく、あるいは前駆体または融合タンパク質などのより大きなタンパク質の一部であってもよい。分泌配列またはリーダー配列、プロ配列、精製を助ける配列、例えば反復するヒスチジン残基、あるいは組換え産生時の安定性に必要なさらなる配列を含有する付加的なアミノ酸配列を含むことは、しばしば、有益である。

【００１７】

本発明のポリペプチドは、任意の好適な方法で、例えば、天然に存在する供給源から、または発現システム（下記参照）を含む遺伝子操作された宿主細胞から単離することによって、あるいは例えば、自動化されたペプチド合成機を使用する化学合成によって、かかるそのような方法の組合せによって調製することができる。かかるポリペプチドを調製するための手段は当該分野では十分に理解されている。

【００１８】

さらなる態様において、本発明は* 遺伝子名ポリヌクレオチドに関する。そのようなポリヌクレオチドには下記が含まれる：

（ａ）配列番号１のポリヌクレオチド配列に対する同一性が少なくとも９５％、９６％、９７％、９８％または９９％であるポリヌクレオチド配列を含むポリヌクレオチド、

（ｂ）配列番号１のポリヌクレオチドを含むポリヌクレオチド、

（ｃ）配列番号１のポリヌクレオチドに対する同一性が少なくとも９５％、９６％、９７％、９８％または９９％であるポリヌクレオチド、

（ｄ）配列番号１のポリヌクレオチド、

（ｅ）配列番号２のポリペプチド配列に対して少なくとも９５％、９６％、９７％、９８％または９９％の同一性を有するポリペプチド配列をコードするポリヌクレオチド配列を含むポリヌクレオチド、

（ｆ）配列番号２のポリペプチドをコードするポリヌクレオチド配列を含むポリヌクレオチド、

（ｇ）配列番号２のポリペプチド配列に対して少なくとも９５％、９６％、９７％、９８％または９９％の同一性を有するポリペプチド配列をコードするポリヌクレオチド配列を有するポリヌクレオチド、

（ｈ）配列番号２のポリペプチドをコードするポリヌクレオチド、

（ｉ）配列番号１のポリヌクレオチド配列と比較して、０．９５、０．９６、０．９７、０．９８または０．９９の同一性指標を有するポリヌクレオチド配列を有するか、またはそのようなポリヌクレオチド配列を含むポリヌクレオチド、

（ｊ）配列番号２のポリペプチド配列と比較して、０．９５、０．９６、０．９７、０．９８または０．９９の同一性指標を有するポリペプチド配列をコードするポリヌクレオチド配列を有するか、またはそのようなポリヌクレオチド配列を含むポリヌクレオチド、ならびに

上記ポリヌクレオチドのフラグメントおよび変異体であるポリヌクレオチド、あるいは上記ポリヌクレオチドに対してその全長にわたって相補的であるポリヌクレオチド。

【００１９】

本発明のポリヌクレオチドの好ましいフラグメントには、配列番号１の配列に由来する少なくとも１５、３０、５０または１００個の連続した塩基を有する塩基配列を含むポリヌクレオチド、あるいは配列番号１の配列から、少なくとも３０、５０または１００個の連続した塩基が短縮化または欠失している配列を含むポリヌクレオチドが含まれる。

【００２０】

本発明のポリヌクレオチドの好ましい変異体には、スプライス変異体、対立遺伝子変異体、および１つまたは複数の単一塩基多型（ＳＮＰ）を有するポリヌクレオチドを含む多型体が含まれる。

【００２１】
本発明のポリヌクレオチドには、配列番号２のアミノ酸配列を含み、かついくつかの、例えば、５０～３０、３０～２０、２０～１０、１０～５、５～３、３～２、２～１または１個のアミノ酸残基が任意の組合せで置換、欠失または付加されているポリペプチド変異体をコードするポリヌクレオチドもまた含まれる。
【００２２】
さらなる態様において、本発明は、本発明のＤＮＡ配列のＲＮＡ転写物であるポリヌクレオチドを提供する。したがって、下記のＲＮＡポリヌクレオチドが提供される：
（ａ）配列番号２のポリペプチドをコードするＤＮＡ配列のＲＮＡ転写物を含むＲＮＡポリヌクレオチド、
（ｂ）配列番号２のポリペプチドをコードするＤＮＡ配列のＲＮＡ転写物であるＲＮＡポリヌクレオチド、
（ｃ）配列番号１のＤＮＡ配列のＲＮＡ転写物を含むＲＮＡポリヌクレオチド、または
（ｄ）配列番号１のＤＮＡ配列のＲＮＡ転写物であるＲＮＡポリヌクレオチド、ならびにそれらに対して相補的であるＲＮＡポリヌクレオチド。
【００２３】
配列番号１のポリヌクレオチド配列は、配列番号２のポリペプチドをコードするｃＤＮＡ配列である。配列番号２のポリペプチドをコードするポリヌクレオチド配列は、配列番号１の配列をコードするポリペプチドと同一であってもよく、あるいは遺伝子暗号の縮退（縮重性）の結果として配列番号２のポリペプチドを同様にコードする、配列番号１以外の配列であってもよい。配列番号２のポリペプチドは、スペクトリン様ドメイン、コイルドコイル、グアノシン交換因子ドメインおよびプレクストリン相同性ドメインをコードするヒトＤｕｏ　ｍＲＮＡとの相同性および／または構造的類似性を有するグアニン交換因子（ＧＥＦ）のファミリーの他のタンパク質との関連性を有し［Ｃｏｌｏｍｅｒ，Ｖ．、Ｅｎｇｅｌｅｎｄｅｒ，Ｓ．、Ｓｈａｒｐ，Ａ．Ｈ．、Ｄｕａｎ，Ｋ．、Ｃｏｏｐｅｒ，Ｊ．Ｋ．、Ｌａｎａｈａｎ，Ａ．、Ｌｙｆｏｒｄ，Ｇ．、Ｗｏｒｌｅｙ，Ｐ．およびＲｏｓｓ，Ｃ．Ａ．、ハンチントン関連タンパク質１はｒａｃ１グアニンヌクレオチド交換因子ドメインによりＴｒｉｏ様ポリペプチドに結合する、Ｈｕｍ．Ｍｏｌ．Ｇｅｎｅｔ．６（９）、１５１９～１５２５（１９９７）］、そしてプロテインキナーゼ、ｒａｃグアニン交換因子、ｒｈｏグアニン交換因子およびスペクトリン様反復に対する類似性を示すヒトＴｒｉｏ　ｍＲＮＡとの関連性を有する［Ｄｅｂａｎｔ，Ａ．、Ｓｅｒｒａ－Ｐａｇｅｓ，Ｃ．、Ｓｅｉｐｅｌ，Ｋ．、Ｏ’Ｂｒｉｅｎ，Ｓ．、Ｔａｎｇ，Ｍ．、Ｐａｒｋ，Ｓ．Ｈ．およびＳｔｒｅｕｌｉ，Ｍ．、マルチドメインタンパク質ＴｒｉｏはＬＡＲ膜貫通チロシンホスファターゼに結合し、プロテインキナーゼドメインを含み、離れたｒａｃ特異的グアニンヌクレオチド交換因子およびｒｈｏ特異的グアニンヌクレオチド交換因子を有する、Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．９３（１１）、５４６６～５４７１（１９９６）］。
【００２４】
本発明の好ましいポリペプチドおよびポリヌクレオチドは、特に、それらの相同的なポリペプチドおよびポリヌクレオチドと類似する生物学的な機能／性質を有することが期待される。さらに、本発明の好ましいポリペプチドおよびポリヌクレオチドは少なくとも１つの＊遺伝子名活性を有する。
【００２５】
本発明のポリヌクレオチドは、ヒトＨｅＬａ細胞（ＡＴＣＣ：ＣＣＬ－２）のｍＲＮＡに由来するｃＤＮＡライブラリから、そしてヒト胎児脳組織に由来するｃＤＮＡライブラリから標準的なクローニング技術およびスクリーニング技術を使用して得ることができる（例えば、Ｓａｍｂｒｏｏｋ他、Ｍｏｌｅｃｕｌａｒ　Ｃｌｏｎｉｎｇ：Ａ　Ｌａｂｏｒａｔｏｒｙ　Ｍａｎｕａｌ、第２版、Ｃｏｌｄ　Ｓｐｒｉｎｇ　Ｈａｒｂｏｒ　Ｌａｂｏｒａｔｏｒｙ　Ｐｒｅｓｓ、Ｃｏｌｄ　Ｓｐｒｉｎｇ　Ｈａｒｂｏｒ、Ｎ．Ｙ．（１９８９）参照）。本発明のポリヌクレオチドはまた、ゲノムＤＮＡライブラリなどの天然の供給

源から得ることができ、あるいは周知の技術および市販の技術を用いて合成することができる。

【００２６】

本発明のポリヌクレオチドを本発明のポリペプチドの組換え製造のために使用する際には、該ポリヌクレオチドは、成熟型ポリペプチドのコード配列、それ自体、あるいはリーダー配列もしくは分泌配列、プレタンパク質配列もしくはプロタンパク質配列もしくはプレプロタンパク質配列、または他の融合ペプチド部分をコードするコード配列などの他のコード配列と読み枠を合わせた成熟型ポリペプチドのコード配列を含むことができる。例えば、融合ポリペプチドの精製を容易にするマーカー配列をコードさせることができる。本発明のこの態様のいくつかの好ましい実施形態において、該マーカー配列は、ｐＱＥベクター（Ｑｉａｇｅｎ，Ｉｎｃ．）において提供され、そしてＧｅｎｔｚ他、Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、（１９８９）８６：８２１～８２４に記載されている、ヘキサ・ヒスチジン・ペプチド、あるいはＨＡタグである。該ポリヌクレオチドはまた、転写される非翻訳配列、スプライシングシグナルおよびポリアデニレーション・シグナル、リボソーム結合部位、ならびにｍＲＮＡを安定化させる配列などの５’および３’非コード配列を含有することができる。

【００２７】

配列番号１のポリヌクレオチド配列に対して同一であるか、または十分な同一性を有するポリヌクレオチドは、ｃＤＮＡおよびゲノムＤＮＡに対するハイブリダイゼーション・プローブとして、あるいは核酸増幅反応（例えばＰＣＲ）用のプライマーとして使用することができる。かかるプローブおよびプライマーは、本発明のポリペプチドをコードする完全長型ｃＤＮＡおよびゲノム・クローンを単離するために、ならびに配列番号１に対する高い配列類似性（概して少なくとも９５％の同一性）を有する他の遺伝子（ヒト供給源に由来するパラログならびにヒト以外の種に由来するオルソログおよびパラログをコードする遺伝子を含む）のｃＤＮＡクローンおよびゲノム・クローンを単離するために使用することができる。好ましいプローブおよびプライマーは、一般には、少なくとも１５塩基、好ましくは少なくとも３０塩基を含み、そして少なくとも１００塩基ではなくとも、少なくとも５０塩基を有し得る。特に好ましいプローブは３０～５０塩基を有する。特に好ましいプライマーは２０～２５塩基を有する。

【００２８】

ヒト以外の種に由来するホモログを含む、本発明のポリペプチドをコードするポリヌクレオチドは、配列番号１の配列またはそのフラグメントを有する、好ましくは少なくとも１５塩基の標識されたプローブを用いて厳格なハイブリダイゼーション条件下でライブラリをスクリーニングする過程；および前記ポリヌクレオチド配列を含有する完全長型ｃＤＮＡクローンおよびゲノム・クローンを単離する過程を含む工程によって得ることができる。かかるハイブリダイゼーション技術は当業者には十分に知られている。好ましい厳格なハイブリダイゼーション条件には、５０％ホルムアミド、５×ＳＳＣ（１５０ｍＭのＮａＣｌ、１５ｍＭのクエン酸三ナトリウム）、５０ｍＭのリン酸ナトリウム（ｐＨ７．６）、５×デンハルト溶液、１０％のデキストラン硫酸および２０マイクログラム／ｍｌの変性させた剪断サケ精子ＤＮＡを含む溶液において４２℃で終夜インキュベーションし、その後、０．１×ＳＳＣにおいて約６５℃にてフィルターを洗浄することが含まれる。したがって、本発明はまた、配列番号１の配列またはそのフラグメントを有する、好ましくは少なくとも１５塩基の標識されたプローブを用いて厳格なハイブリダイゼーション条件下でライブラリをスクリーニングすることによって得られるポリヌクレオチド、好ましくは少なくとも１００塩基の塩基配列を有するポリヌクレオチドを含む。

【００２９】

当業者は、多くの場合において、該ポリペプチドをコードする領域が５’末端に至るまで必ずしも完全に伸長していない点で、単離されたｃＤＮＡ配列が不完全であることを理解している。これは、本質的に「プロセシング能」（ポリメリゼーション反応の間、テンプレートに結合した状態を維持する酵素の能力の指標）が低い酵素である逆転写酵素が、第

１鎖ｃＤＮＡ合成の際にｍＲＮＡテンプレートのＤＮＡコピーを完成させることができない結果である。

【００３０】

全長型ｃＤＮＡを得るために、あるいは短いｃＤＮＡを伸長させるために利用でき、かつ当業者に周知の方法がいくつかある。例えば、ｃＤＮＡ末端の迅速な増幅（ＲＡＣＥ）方法に基づく方法がある（例えば、Ｆｒｏｈｍａｎ他、Ｐｒｏｃ．Ｎａｔ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、８５、８９９８～９００２、１９８８を参照）。例えば、Ｍａｒａｔｈｏｎ（商標）法（Ｃｌｏｎｔｅｃｈ Ｌａｂｏｒａｔｏｒｉｅｓ Ｉｎｃ．）によって例示されるこの技術の近年の改変により、より長いｃＤＮＡに対する探索が著しく単純化されている。Ｍａｒａｔｈｏｎ（商標）法では、選択された組織から抽出されたｍＲＮＡとその両脇に連結された「アダプター」配列とから、ｃＤＮＡが調整される。その後、核酸増幅（ＰＣＲ）を行い、遺伝子特異的オリゴヌクレオチドプライマーおよびアダプター特異的オリゴヌクレオチドプライマーの組合せを使用してｃＤＮＡの「失われている」５’末端を増幅する。その後、「ネスティッド（入れこ型）」プライマー、すなわち、増幅産物の内部にアニーリングするように設計されたプライマー（典型的には、アダプター配列においてさらに３’側にアニーリングするアダプター特異的プライマー、および既知の遺伝子配列においてさらに５’側にアニーリングする遺伝子特異的プライマー）を使用して、ＰＣＲ反応を繰り返す。その後、この反応の生成物はＤＮＡ配列決定によって分析することができる。生成物を既存のｃＤＮＡに直接結合して完全な配列を得るか、または５’プライマーを設計するための新しい配列情報を使用して別の完全長ＰＣＲを行うことによって、完全長型のｃＤＮＡを構築することができる。

【００３１】

本発明の組換えポリペプチドは、発現システムを含む遺伝子操作された宿主細胞からこの分野で周知の方法によって調製することができる。従って、さらなる態様において、本発明は、本発明のポリヌクレオチドの１つまたは複数を含む発現システムと、そのような発現システムで遺伝子操作されている宿主細胞と、および組換え技術による本発明のポリペプチドの製造とに関する。本発明のＤＮＡ構築物に由来するＲＮＡを用いてかかるタンパク質を製造するために、無細胞翻訳システムを用いることもできる。

【００３２】

組換え製造のために、宿主細胞を遺伝子操作し本発明のポリヌクレオチドに対する発現システムまたはその一部を取り込ませることができる。ポリヌクレオチドは、Ｄａｖｉｓ他、Ｂａｓｉｃ Ｍｅｔｈｏｄｓ ｉｎ Ｍｏｌｅｃｕｌａｒ Ｂｉｏｌｏｇｙ（１９８６）およびＳａｍｂｒｏｏｋ他（同上）などの多くの標準的な実験室マニュアルに記載される方法によって宿主細胞に導入することができる。ポリヌクレオチドを宿主細胞に導入する好ましい方法には、例えば、リン酸カルシウム・トランスフェクション、ＤＥＡＥ－デキストラン媒介トランスフェクション、トランスベクション、マイクロインジェクション、カチオン性脂質媒介トランスフェクション、エレクトロポレーション、トランスダクション、スクレイプ負荷、バリスティック導入または感染が含まれる。

【００３３】

適当な宿主の代表的な例には、ストレプトコッカス属、スタフィロコッカス属、大腸菌、ストレプトミセス属ならびに枯草菌細胞などの細菌細胞；酵母細胞やアスペルギルス属細胞などの菌類細胞；ショウジョウバエ（Ｄｒｏｓｏｐｈｉｌａ）Ｓ２細胞やＳｐｏｄｏｐｔｅｒａ Ｓｆ９細胞などの昆虫細胞；ＣＨＯ、ＣＯＳ、ＨｅＬａ、Ｃ１２７、３Ｔ３、ＢＨＫ、ＨＥＫ２９３およびＢｏｗｅｓメラノーマ細胞などの動物細胞；ならびに植物細胞が含まれる。

【００３４】

多様な発現システムを使用することができる。例えば、染色体、エピソームおよびウイルスに由来するシステム、例えば、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、バキュロウイルス、パポバウイルス（ＳＶ４０など）、ワクシニアウイルス、アデノウイルス、鶏痘ウイルス、仮性

狂犬病ウイルスやレトロウイルスなどのウイルスに由来するベクター、ならびに、プラスミドおよびバクテリオファージの遺伝子エレメントから誘導されたものなどの、それらの組合せに由来するベクター（コスミドおよびファージミドなど）を使用することができる。該発現システムは、発現を生じさせるとともに、発現を調節するための制御領域を含有することができる。一般に、宿主内においてポリペプチドを産生させるためにポリヌクレオチドを維持し、または伝搬させ、または発現させることができるシステムまたはベクターはどれも使用することができる。適切なポリヌクレオチド配列を、例えば、Ｓａｍｂｒｏｏｋ他（同上）に示されている技術などの様々なよく知られている日常的な技術のいずれかによって発現システムに挿入することができる。適切な分泌シグナルを所望するポリペプチドに組み込んで、翻訳されたタンパク質を小胞体の内腔、ペリプラズム腔または細胞外環境に分泌させることができる。これらのシグナルは、ポリペプチドに対して内因性であってもよく、あるいは異種のシグナルであってもよい。

【００３５】

スクリーニングアッセイで用いるために本発明のポリペプチドを発現させる場合、該ポリペプチドを細胞の表面に産生させることが一般には好ましい。この場合細胞を集菌し、その後スクリーニングアッセイにおいて使用することができる。該ポリペプチドが培地中に分泌される場合には、培地を回収してポリペプチドの回収および精製をすることができる。細胞内に産生された場合には、細胞を最初に溶解して、その後ポリペプチドを回収しなければならない。

【００３６】

本発明のポリペプチドは、硫酸アンモニウムまたはエタノール沈殿、酸抽出、アニオンまたはカチオン交換クロマトグラフィ、ホスホセルロース・クロマトグラフィ、疎水性相互作用クロマトグラフィ、アフィニティ・クロマトグラフィ、ヒドロキシルアパタイト・クロマトグラフィおよびレクチン・クロマトグラフィを含む周知の方法によって組換え細胞培養物から回収および精製することができる。最も好ましくは、ハイ・パフォーマンス液体クロマトグラフィが精製に用いられる。該ポリペプチドが細胞内合成、単離および／または精製のときに変性する際には、タンパク質をリフォールディングさせるために周知の技術を用いて、活性な立体配座を再生させることができる。

【００３７】

本発明のポリヌクレオチドは、関連遺伝子における変異を検出することによって診断試薬として使用することができる。ｃＤＮＡ配列またはゲノム配列における配列番号１のポリヌクレオチドにより特定される遺伝子で、機能不全に関連している遺伝子の突然変異型を検出することによって、その遺伝子の過少発現、過剰発現あるいは変化した空間的または時間的な発現から生じる疾患の診断またはそのような疾患に対する感受性の診断の一助になり得るか、またはそのような診断を規定し得る診断ツールが提供される。遺伝子に変異を有する個体を、この分野で周知な様々な技術によってＤＮＡレベルで検出することができる。

【００３８】

診断に必要な核酸は、血液、尿、唾液、組織生検体または解剖体などの被験体の細胞から得ることできる。ゲノムＤＮＡは、検出のために直接使用することができ、あるいは分析に先立ち、ＰＣＲ、好ましくはＲＴ－ＰＣＲ、または他の増幅法を使用して酵素的に増幅してもよい。ＲＮＡまたはｃＤＮＡもまた同様の方法で使用することができる。欠失および挿入は、正常な遺伝子型と比較して増幅産物のサイズにおける変化によって検出することができる。点変異は、増幅されたＤＮＡを＊遺伝子名の標識されたヌクレオチド配列にハイブリダイゼーションさせることによって同定することができる。完全に一致する配列は、リボヌクレアーゼ（ＲＮａｓｅ）消化、あるいは融解温度の差によってミスマッチした二重鎖から区別することができる。ＤＮＡ配列の違いはまた、変性剤の存在下または非存在下でのゲルにおけるＤＮＡフラグメントの電気泳動移動度の変化によって、あるいは直接的なＤＮＡ配列決定によって検出することができる（例えば、Ｍｙｅｒｓ他、Ｓｃｉｅｎｃｅ、（１９８５）２３０：１２４２）。特定の位置における配列の変化もまた、リ

ボヌクレアーゼ保護またはＳ１保護などのヌクレアーゼ保護アッセイあるいは化学的な切断方法によって明らかにすることができる（Ｃｏｔｔｏｎ他、Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、（１９８５）８５：４３９７～４４０１）。

【００３９】

＊遺伝子名ポリヌクレオチド配列またはそのフラグメントを含むオリゴヌクレオチド・プローブのアレイを構築して、例えば、遺伝子変異の効率的なスクリーニングを行うことができる。かかるアレイは、高密度のアレイまたは格子状であることが好ましい。アレイ技術法は周知であり、一般的な適用性を有し、遺伝子発現、遺伝子連鎖および遺伝子変動性を含む分子遺伝学における様々な問題を検討するために使用することができ、例えば、Ｍ．Ｃｈｅｅ他、Ｓｃｉｅｎｃｅ、２７４、６１０～６１３（１９９６）およびそれに引用されている他の参考文献を参照する。

【００４０】

異常に、低下または増大しているポリペプチドまたはｍＲＮＡの発現レベルの検出もまた、本発明の疾患に対する被験体の感受性を診断または決定するために使用することができる。発現の減少または増加は、ポリヌクレオチドを定量することに関してこの分野で周知の方法、例えば、核酸増幅、例えばＰＣＲ、ＲＴ－ＰＣＲ、リボヌクレアーゼ保護、ノーザン・ブロッティングおよび他のハイブリダイゼーション法のいずれかを使用してＲＮＡレベルで測定することができる。宿主に由来するサンプルにおける本発明のポリペプチドなどのタンパク質のレベルを決定するために使用することができるアッセイ手法は当業者には周知である。かかるアッセイ方法には、放射免疫アッセイ、結合タンパク質競合アッセイ、ウエスタン・ブロット分析およびＥＬＩＳＡアッセイが含まれる。

【００４１】

従って、別の態様において本発明は下記を含む診断キットに関する：

（ａ）本発明のポリヌクレオチド、好ましくは配列番号１のヌクレオチド配列またはそのフラグメントもしくはＲＮＡ転写物；

（ｂ）（ａ）のヌクレオチド配列に対して相補的なヌクレオチド配列、

（ｃ）本発明のポリペプチド、好ましくは配列番号２のポリペプチドまたはそのフラグメント、あるいは

（ｄ）本発明のポリペプチドに対する抗体、好ましくは配列番号２のポリペプチドに対する抗体。

【００４２】

任意のかかるキットにおいて、（ａ）、（ｂ）、（ｃ）または（ｄ）は実質的な成分を構成することができることは理解される。かかるキットは、疾患または疾患に対する感受性、中でも特に本発明の疾患を診断する際に有用である。

【００４３】

本発明のポリヌクレオチド配列は染色体局在化研究に有益である。該配列は、個々のヒト染色体における特定位置に対して特異的に標的化され、かつその特定の位置とハイブリダイゼーションし得る。本発明に従って関連する配列を染色体にマッピングすることは、そのような配列を遺伝子関連疾患と相関させる際の重要な最初の過程である。配列が正確な染色体位置にマッピングされると、染色体上における配列の物理的な位置を遺伝地図データと相関させることができる。かかるデータは、例えば、Ｖ．ＭｃＫｕｓｉｃｋの「ヒトにおけるメンデル遺伝」において見出される（これはＪｏｈｎｓ　Ｈｏｐｋｉｎｓ大学Ｗｅｌｃｈ　Ｍｅｄｉｃａｌ　Ｌｉｂｒａｒｙからオンラインで入手可能）。遺伝子と、同じ染色体領域にマッピングされている疾患との関係が、その後、連鎖解析（物理的に隣接する遺伝子の同時遺伝）によって同定される。ゲノム配列（遺伝子フラグメントなど）に関する正確なヒト染色体局在化を放射ハイブリッド（ＲＨ）マッピングを使用して決定することができる（Ｗａｌｔｅｒ，Ｍ．、Ｓｐｉｌｌｅｔｔ，Ｄ．、Ｔｈｏｍａｓ，Ｐ．、Ｗｅｉｓｓｅｎｂａｃｈ，Ｊ．およびＧｏｏｄｆｅｌｌｏｗ，Ｐ．（１９９４）、ゲノム全体の放射ハイブリッドマップを構築する方法、Ｎａｔｕｒｅ　Ｇｅｎｅｔｉｃｓ、７、２２～２８）。多数のＲＨパネルをＲｅｓｅａｒｃｈ　Ｇｅｎｅｔｉｃｓ（Ｈｕｎｔｓｖ

ｉｌｌｅ、ＡＬ、米国）から得ることができる。例えば、ＧｅｎｅＢｒｉｄｇｅ４ ＲＨパネル（Ｈｕｍ Ｍｏｌ Ｇｅｎｅｔ、１９９６、Ｍａｒ；５（３）：３３９～４６、ヒトゲノムの放射ハイブリッドマップ。Ｇｙａｐａｙ Ｇ.、Ｓｃｈｍｉｔｔ Ｋ.、Ｆｉｚａｍｅｓ Ｃ.、Ｊｏｎｅｓ Ｈ.、Ｖｅｇａ－Ｃｚａｒｎｙ Ｎ.、Ｓｐｉｌｌｅｔｔ Ｄ.、Ｍｕｓｅｌｅｔ Ｄ.、Ｐｒｕｄ’Ｈｏｍｍｅ ＪＦ、Ｄｉｂ Ｃ.、Ａｕｆｆｒａｙ Ｃ.、Ｍｏｒｉｓｓｅｔｔｅ Ｊ.、Ｗｅｉｓｓｅｎｂａｃｈ，Ｊ.、Ｇｏｏｄｆｅｌｌｏｗ，ＰＮ）。このパネルを使用して遺伝子の染色体上の位置を決定するために、９３回のＰＣＲが、ＲＨ ＤＮＡについて目的とする遺伝子から設計されたプライマーを用いて行われる。これらのＤＮＡはそれぞれが、ハムスターのバックグラウンド（ヒト／ハムスターのハイブリッド細胞株）中に維持されたランダムなヒト・ゲノム・フラグメントを含有する。これらのＰＣＲにより、目的とする遺伝子のＰＣＲ産物の存在の有無を示す９３のスコアが得られる。これらのスコアを既知の位置のゲノム配列に由来するＰＣＲ産物を使用して作製されたスコアと比較される。この比較は、ｈｔｔｐ：／／ｗｗｗ．ｇｅｎｏｍｅ．ｗｉ．ｍｉｔ．ｅｄｕ／において行われる。本発明の遺伝子はヒト染色体の第２染色体（Ｄ２Ｓ３３５－Ｄ２Ｓ２２５７）にマッピングされる。

【００４４】

本発明のポリヌクレオチド配列はまた組織発現研究に対する有益なツールである。かかる研究は、それらをコードするｍＲＮＡを検出することによって、コードされたポリペプチドの組織内の発現パターンに関する指標を与える、本発明のポリヌクレオチドの発現パターンの決定を可能とする。使用される技術は、当該分野では周知であり、ｃＤＮＡマイクロアレイハイブリダイゼーション（Ｓｃｈｅｎａ他、Ｓｃｉｅｎｃｅ、２７０、４６７～４７０、１９９５およびＳｈａｌｏｎ他、Ｇｅｎｏｍｅ Ｒｅｓ、６、６３９～６４５、１９９６）などの、格子上に配置された様々なクローンに対するインサイチュー（ｉｎ ｓｉｔｕ）・ハイブリダイゼーション技術、およびＰＣＲなどのヌクレオチド増幅技術を含む。好ましい方法では、Ｐｅｒｋｉｎ Ｅｌｍｅｒから入手可能なＴＡＱＭＡＮ（商標）法が使用される。これらの研究結果により、生物におけるポリペプチドの正常な機能の指標を得ることができる。加えて、ｍＲＮＡの正常な発現パターンと、同じ遺伝子の別の形態（例えば、潜在的または調節的な変異をコードするポリペプチドにおける変化を有するもの）によってコードされるｍＲＮＡの発現パターンとの比較研究により、本発明のポリペプチドの役割に対する有益な洞察がもたらされ得るか、または疾患におけるその不適切な発現の役割に対する有益な洞察がもたらされ得る。そのような不適切な発現は、時間的、空間的または単に量的な性質であり得る。

【００４５】

本発明のさらなる態様は抗体に関する。本発明のポリペプチドまたはそのフラグメントあるいはそれらを発現する細胞は、本発明のポリペプチドに対して免疫特異的である抗体を産生させるための免疫原として使用することができる。用語「免疫特異的」は、抗体が、先行技術における他の関連するポリペプチドに対するその親和性よりも実質的に大きい親和性を本発明のポリペプチドに対して有することを意味する。

【００４６】

本発明のポリペプチドに対して生じる抗体は、慣用的なプロトコルを使用して、該ポリペプチドまたはエピトープ保持したフラグメントまたは細胞を動物、好ましくは非ヒト動物に投与することによって得ることができる。モノクローナル抗体を調製する場合、継代的な細胞株培養物によって産生される抗体を提供するいかなる技術を使用することができる。例としては、ハイブリドーマ技術（Ｋｏｈｌｅｒ，Ｇ.およびＭｉｌｓｔｅｉｎ，Ｃ.、Ｎａｔｕｒｅ（１９７５）、２５６：４９５～４９７）、トリオーマ技術、ヒトＢ細胞ハイブリドーマ技術（Ｋｏｚｂｏｒ他、Ｉｍｍｕｎｏｌｏｇｙ Ｔｏｄａｙ（１９８３）、４：７２）、およびＥＢＶハイブリドーマ技術（Ｃｏｌｅ他、Ｍｏｎｏｃｌｏｎａｌ Ａｎｔｉｂｏｄｉｅｓ ａｎｄ Ｃａｎｃｅｒ Ｔｈｅｒａｐｙ、７７～９６、Ａｌａｎ Ｒ.Ｌｉｓｓ，Ｉｎｃ.、１９８５）が含まれる。

【００４７】

米国特許第４，９４６，７７８号に記載される技術などの一本鎖抗体の製造技術もまた、本発明のポリペプチドに対する一本鎖抗体を製造するために適応させることができる。また、トランスジェニックマウスまたは他の哺乳動物を含む他の生物を使用して、ヒト化抗体を発現させることができる。

【００４８】

上記の抗体を用いて、該ポリペプチドを発現するクローンを単離または同定する、もしくはアフィニティ・クロマトグラフィによって該ポリペプチドを精製することができる。本発明のポリペプチドに対する抗体は、また、特に本発明の疾患を治療するために用いることができる。

【００４９】

本発明のポリペプチドおよびポリヌクレオチドをワクチンとして使用することもできる。したがって、さらなる態様において、本発明は、哺乳動物における免疫学的応答を誘導する方法に関する。この方法は、前記動物を疾患から、その疾患が個体において既に慢性化しているか否かに関わらず保護するために抗体応答および／またはＴ細胞免疫応答（例えば、サイトカイン産生Ｔ細胞または細胞傷害性Ｔ細胞を含む）を生じさせるのに適する、本発明のポリペプチドを哺乳動物に接種することを含む。哺乳動物における免疫学的応答はまた、本発明にかかる疾患から前記動物を保護する抗体を産生させるようにかかる免疫学的応答を誘導するために、該ポリヌクレオチドの発現を支配し、かつ該ポリペプチドをコードするベクターによって本発明のポリペプチドをインビボで送達することを含む方法によって誘導され得る。そのようなベクターを投与する１つの方法は、粒子またはそれ以外のものの上へのコーティング物として所望する細胞中へのそれを促進することによる。かかる核酸ベクターは、ＤＮＡ、ＲＮＡ、修飾型核酸またはＤＮＡ／ＲＮＡハイブリッドを含むことができる。用途によって、ワクチン、ポリペプチドまたは核酸ベクターは、通常、ワクチン配合物（組成物）として提供される。該配合物はさらに好適なキャリアを含むことができる。ポリペプチドは胃において分解されることもあるため、それは、好ましくは非経口で投与される（例えば皮下、筋肉内、静脈内または皮内注射）。非経口投与に好適な配合物には、抗酸化剤、緩衝剤、静菌剤、および配合物を接種者の血液と等張性にする溶質を含んでもよい、水性および非水性の無菌注射液；ならびに懸濁剤または増粘剤を含んでもよい、水性および非水性の無菌懸濁物が含まれる。該配合物は、単位用量容器または多回用量容器で、例えば密封アンプルおよびバイアルに入れて提供することができ、そして使用直前に無菌の液体キャリアを添加することだけでよい、凍結乾燥状態で保存することができる。該ワクチン配合物はまた、水中油型システムおよび当該分野で既知のその他のシステムなどの、配合物の免疫原性を増強するアジュバント・システムを含むことができる。投与量は該ワクチンの比活性に依存するが、型どおりの実験によって容易に決定することができる。

【００５０】

本発明のポリペプチドは、１つまたは複数の疾患状態、特に、前述の本発明の疾患に関連する１つまたは複数の生物学的機能を有する。従って、該ポリペプチドの機能または濃度を刺激または阻害する化合物を同定することは有用である。従って、さらなる態様において、本発明は、該ポリペプチドの機能または濃度を刺激または阻害する化合物を同定するために化合物をスクリーニングする方法を提供する。かかる方法により、本明細書前記のような本発明のそのような疾患に対する治療目的および予防目的のために用いることができるアゴニストまたはアンタゴニストが同定される。化合物は、様々な供給源、例えば、細胞、無細胞調製物、化学ライブラリ、化学化合物のコレクション、および天然産物の混合物から同定することができる。このように同定される、かかるアゴニストまたはアンタゴニストは、場合によっては、ポリペプチド自体の、天然または修飾された基質、リガンド、受容体、酵素など；その構造的または機能的な模倣体（Ｃｏｌｉｇａｎ ｅｔ ａｌ．，Ｃｕｒｒｅｎｔ Ｐｒｏｔｏｃｏｌｓ ｉｎ Ｉｍｍｕｎｏｌｏｇｙ，１（２）：５章（１９９１）を参照のこと）あるいは小分子であってもよい。そのような小分子は、分子量が２，０００ダルトン未満であることが好ましく、より好ましくは３００～１，

０００ダルトンであり、最も好ましくは４００～７００ダルトンである。これらの小分子は有機分子であることが好ましい。

【００５１】

該スクリーニング方法は、候補化合物に直接的または間接的に連結されている標識を利用して、該ポリペプチド、あるいは該ポリペプチドまたはその融合タンパク質を表出している細胞または膜に対する候補化合物の結合を単に測定することでもよい。あるいは、スクリーニング方法は、標識された競合剤（例えばアゴニストまたはアンタゴニスト）に対抗して候補化合物がポリペプチドに競合的に結合することを（定性的または定量的に）測定または検出することを含み得る。さらにこれらのスクリーニング方法では、該ポリペプチドを表出している細胞に適切な検出システムを使用して候補化合物がポリペプチドの活性化または阻害によって誘起されるシグナルをもたらすか否かを試験することもできる。一般には、活性化阻害剤が既知のアゴニストの存在下でアッセイされ、そして候補化合物の存在下でのアゴニストによる活性化に対する作用が観測される。さらに、該スクリーニング方法は、候補化合物を、本発明のポリペプチドを含有する溶液と混合して、混合物を形成させる工程と、混合物における遺伝子名活性を測定する工程と、および混合物の＊遺伝子名活性を、候補化合物を含有しないコントロール混合物と比較する工程とを単に含み得る。

【００５２】

本発明のポリペプチドは、従来の低い容量のスクリーニング方法において、そしてまた高処理能（ハイ・スループット）スクリーニング（ＨＴＳ）形式において用いることができる。かかるＨＴＳ形式には、十分に確立された９６穴マイクロタイター（ｍｉｃｏｔｉｔｅｒ）プレートおよびより最近には３８４穴マイクロタイター（ｍｉｃｏｔｉｔｅｒ）プレートの使用だけでなく、Ｓｃｈｕｌｌｅｋ他、Ａｎａｌ　Ｂｉｏｃｈｅｍ．、２４６、２０～２９（１９９７）に記載のナノウエル法などの最近現れた方法もまた含まれる。

【００５３】

Ｆｃ部分および＊遺伝子名ポリペプチドから生成される融合タンパク質などの融合タンパク質もまた、本明細書中前記ように、本発明のポリペプチドに対するアンタゴニストを同定するための高処理能スクリーニングアッセイに使用することができる（Ｄ．Ｂｅｎｎｅｔｔ他、Ｊ　Ｍｏｌ　Ｒｅｃｏｇｎｉｔｉｏｎ、８：５２～５８（１９９５）；Ｋ．Ｊｏｈａｎｓｏｎ他、Ｊ　Ｂｉｏｌ　Ｃｈｅｍ、２７０（１６）：９４５９～９４７１（１９９５））。

【００５４】

スクリーニング技術

本発明のポリヌクレオチド、ポリペプチドおよび該ポリペプチドに対する抗体はまた、細胞内のｍＲＮＡおよびポリペプチドの産生に対する添加された化合物の作用を検出するためのスクリーニング方法を形成するために使用することができる。例えば、当該分野で公知の標準的な方法によってモノクローナル抗体およびポリクローナル抗体を使用してポリペプチドの分泌または細胞結合濃度を測定するために、ＥＬＩＳＡアッセイを構築することができる。これは、好適に操作された細胞または組織からのポリペプチドの産生を阻害または増強し得る薬剤（また、それぞれアンタゴニストまたはアゴニストとも呼ばれる）を発見するために使用することができる。

【００５５】

本発明のポリペプチドを用いて、当該分野で公知の標準的な受容体結合技術によって、存在する場合には膜結合型または可溶性の受容体を同定することができる。これらには、これらに限定されないものの、ポリペプチドを放射性同位体（例えば^{125}Ｉ）で標識、化学修飾（例えばビオチン化され）、あるいは検出または精製に好適なペプチド配列に融合して、そして推定される受容体の供給源（細胞、細胞膜、細胞上清、組織抽出物、体液）とインキュベーションするリガンド結合およびクロスリンク・アッセイが含まれるが。他の方法には、表面プラズモン共鳴および分光測定法などの生物物理学的技術が含まれる。これらのスクリーニング方法はまた、それらが存在する場合にはポリペプチドのその受容

体に対する結合と競合する該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使用することができる。かかるアッセイを行うための標準的な方法は当該分野では十分に理解されている。

【００５６】

本発明のポリペプチドのアンタゴニストの例には、抗体、あるいは、ある場合には、該ポリペプチド自体のリガンド、基質、受容体、酵素などに密接に関連するオリゴヌクレオチドまたはタンパク質、場合により、例えば、該リガンド、基質、受容体、酵素などのフラグメント；あるいは本発明のポリペプチドに結合するものの、応答を誘発せず、その結果、ポリペプチドの活性が妨げられる、小分子が含まれる。

【００５７】

スクリーニング方法はまた、トランスジェニック技術および＊遺伝子名遺伝子の使用を含んでもよい。トランスジェニック動物を構築する技術は十分に確立されている。例えば、受精した卵母細胞の雄性前核へのマイクロインジェクション、移植前または移植後の胚へのレトロウイルス移入、エレクトロポレーションなどにより遺伝子操作された胚性幹細胞の宿主胚盤胞への注入によって＊遺伝子名遺伝子を導入することができる。特に有用なトランスジェニック動物は、いわゆる「ノック・イン」動物であり、動物の遺伝子がその動物のゲノム内においてヒトの等価体によって置き換えられている。ノック・イン・トランスジェニック動物は、医薬探索のプロセスにおいて、化合物はヒトの標的に対して特異的であるという、標的の妥当性検証用に有用である。他の有用なトランスジェニック動物は、いわゆる「ノック・アウト」動物であり、内因性ＤＮＡ配列によってコードされている、本発明のポリペプチドに対する動物オルソログの発現が細胞内で部分的または完全に無効とされている。遺伝子のノック・アウトは、技術の限界の結果として、特異的な細胞または組織を対象とする、特定の細胞または組織においてのみ生じてもよく、あるいは動物内のすべての細胞または実質的にすべての細胞において生じ得る。トランスジェニック動物の手法はまた、導入遺伝子が発現されて大量の本発明のポリペプチドが得られる、動物全体の発現－クローニング・システムを提供する。

【００５８】

上記の方法において使用されるスクリーニングキットは、本発明のさらなる態様をなす。かかるスクリーニングキットは下記のものを含む：

（ａ）本発明のポリペプチド、

（ｂ）本発明のポリペプチドを発現する組換え細胞、

（ｃ）本発明のポリペプチドを発現する細胞膜、または

（ｄ）本発明のポリペプチドに対する抗体、

そのようなポリペプチドは、好ましくは配列番号２のポリペプチドである。

【００５９】

かかるキット何れにおいても、（ａ）、（ｂ）、（ｃ）または（ｄ）は実質的な構成要素を構成することは理解される。

【００６０】

（用語集）

下記の定義は、本明細書中で既に頻繁に使用されているいくつかの用語の理解を容易にするために提供される。

【００６１】

本明細書中に用いられる「抗体」は、ポリクローナルおよびモノクローナル抗体、キメラ、一本鎖、ならびにヒト化抗体、同様にＦａｂフラグメントをも包含し、Ｆａｂ発現ライブラリまたは他の免疫グロブリン発現ライブラリの生成物を包含する。

【００６２】

「単離（された）」は、その自然の状態から「ヒトの手によって」変化していること、すなわち、自然界に存在する場合、その本来の環境から変化、または取り出されている、もしくはその両方であることを意味する。例えば、生きた生物中に天然に存在するポリヌクレオチドまたはポリペプチドは「単離」されていないが、その天然状態の共存物質から分

離された同じポリヌクレオチドまたはポリペプチドはこの用語の本明細書中での用法に従うと、「単離」されている。さらに、形質転換、遺伝子操作、または何らかの他の組換え方法によって生物に導入されているポリヌクレオチドまたはポリペプチドは、その生物が生存または非生存かのいずれでも、前記生物内に依然として存在している場合でさえ、「単離」されている。

【００６３】

「ポリヌクレオチド」は、一般には、任意のポリリボヌクレオチド（ＲＮＡ）またはポリデオキシリボヌクレオチド（ＤＮＡ）をいうが、非改変型または改変型のＲＮＡまたはＤＮＡであってもよい。「ポリヌクレオチド」には、一本鎖ＤＮＡおよび二本鎖ＤＮＡ、一本鎖領域および二本鎖領域の混合であるＤＮＡ、一本鎖ＲＮＡおよび二本鎖ＲＮＡ、ならびに一本鎖領域および二本鎖領域の混合であるＲＮＡ、一本鎖またはより典型的には二本鎖であり得るか、または一本鎖領域および二本鎖領域の混合であり得るＤＮＡおよびＲＮＡを含むハイブリッド分子が含まれるが、これらに限定されることはない。さらに、「ポリヌクレオチド」は、ＲＮＡまたはＤＮＡあるいはＲＮＡとＤＮＡとの両方を含む三重鎖領域をもいう。用語「ポリヌクレオチド」はまた、１つまたは複数の修飾された塩基を含有するＤＮＡまたはＲＮＡ、および安定性またはその他の理由のために修飾された骨格を有するＤＮＡまたはＲＮＡを含む。「修飾（された）」塩基には、例えば、トリチル化された塩基、およびイノシンなどの非通常型の塩基が含まれる。様々な修飾をＤＮＡおよびＲＮＡに対して行うことができる。従って、「ポリヌクレオチド」は、自然界に典型的に見出されるようなポリヌクレオチドの化学的、酵素的または代謝的に修飾された形態、ならびにウイルスおよび細胞に特徴的なＤＮＡおよびＲＮＡの化学的形態を包含する。「ポリヌクレオチド」はまた、しばしばオリゴヌクレオチドと称される比較的短いポリヌクレオチドを包含する。

【００６４】

「ポリペプチド」は、ペプチド結合または修飾されたペプチド結合（すなわちペプチド等配電子体）によって互いに連結された２つ以上のアミノ酸を含むポリペプチドの何れをもいう。「ポリペプチド」は、ペプチド、オリゴペプチドまたはオリゴマーと広く呼ばれる短い鎖、ならびに一般にはタンパク質と呼ばれるより長い鎖の両方をいう。ポリペプチドは遺伝子によってコードされる２０のアミノ酸とは異なるアミノ酸を含有することができる。「ポリペプチド」は、翻訳後プロセシングなどの天然のプロセス、または当該分野で周知の化学的な修飾方法のいずれかによって、修飾されたアミノ酸配列を含む。かかる修飾は、基本的な教本、より詳細な専門書ならびに数多くの研究文献に広く記載されている。修飾は、ペプチド骨格、アミノ酸側鎖およびアミノ末端またはカルボキシル末端を含むポリペプチド内の任意のところに存在させることができる。同じ型の修飾が所与ポリペプチド内のいくつかの部位に同じ程度または異なる程度で存在してもよいことが理解される。また、所与ポリペプチドは多くのタイプの修飾を含有することができる。ポリペプチドは、ユビキチン化の結果として分枝が成されてもよく、そして分枝型または非分枝型の環状であり得る。環状、分枝状および分枝した環状ポリペプチドは、翻訳に続く天然のプロセスに起因しても、あるいは合成的方法によって作製されてもよい。修飾には、アセチル化、アシル化、ＡＤＰ－リボシル化、アミド化、ビオチン化、フラビンの共有結合、ヘム成分の共有結合、ヌクレオチドまたはヌクレオチド誘導体の共有結合、脂質または脂質誘導体の共有結合、ホスホチジルイノシトールの共有結合、クロス・リンク形成、環化、ジスルフィド結合の形成、脱メチル化、共有結合架橋の形成、システインの形成、ピログルタミン酸の形成、ホルミル化、γ－カルボキシル化、グリコシル化、ＧＰＩ・アンカーの形成、ヒドロキシル化、ヨウ素化、メチル化、ミリストイル化、酸化、タンパク質分解プロセシング、リン酸化、プレニル化、ラセミ化、セレノイル化、硫酸化、アルギニル化などのタンパク質へのアミノ酸のトランスファー・ＲＮＡ媒介付加、ならびユビキチン化が含まれる（例えば、Ｐｒｏｔｅｉｎｓ－Ｓｔｒｕｃｔｕｒｅ ａｎｄ Ｍｏｌｅｃｕｌａｒ Ｐｒｏｐｅｒｔｉｅｓ、第２版、Ｔ．Ｅ．Ｃｒｅｉｇｈｔｏｎ、Ｗ．Ｈ．Ｆｒｅｅｍａｎ ａｎｄ Ｃｏｍｐａｎｙ、Ｎｅｗ Ｙｏｒｋ、１９９３；Ｗｏｌｄ，Ｆ．、翻訳後

のタンパク質修飾：全体像および展望、１～１２、Ｐｏｓｔ－ｔｒａｎｓｌａｔｉｏｎａｌ Ｃｏｖａｌｅｎｔ Ｍｏｄｉｆｉｃａｔｉｏｎ ｏｆ Ｐｒｏｔｅｉｎｓ、Ｂ．Ｃ．Ｊｏｈｎｓｏｎ編、Ａｃａｄｅｍｉｃ Ｐｒｅｓｓ、Ｎｅｗ Ｙｏｒｋ、１９８３；Ｓｅｉｆｔｅｒ他、「タンパク質修飾および非タンパク質補助因子の分析」、Ｍｅｔｈ Ｅｎｚｙｍｏｌ、１８２、６２６～６４６、１９９０；Ｒａｔｔａｎ他、「タンパク質合成：翻訳後修飾およびエージング」、Ａｎｎ ＮＹ Ａｃａｄ Ｓｃｉ、６６３、４８～６２、１９９２）。

【００６５】

ポリペプチド配列の「フラグメント」は、基準配列よりも短いが基準ポリペプチドとなるポリペプチドと同じ生物学的な機能または活性を本質的に保持しているポリペプチド配列をいう。ポリヌクレオチド配列の「フラグメント」は配列番号１の基準配列よりも短いポリヌクレオチド配列をいう。

【００６６】

「変異体」は、基準のポリヌクレオチドまたはポリペプチドとは異なるが、その本質的な性質を保持しているポリヌクレオチドまたはポリペプチドをいう。ポリヌクレオチドの典型的な変異体は、基準ポリヌクレオチドに対して塩基配列に相違がある。変異体の塩基配列における変化により、基準ポリヌクレオチドによってコードされるポリペプチドのアミノ酸配列が変化してもよく、あるいは変化しなくてもよい。ヌクレオチドの変化は、後述するとおり、基準配列によってコードされるポリペプチドにおけるアミノ酸の置換、付加、欠失、融合および短縮化をもたらし得る。ポリペプチドの典型的な変異体は、アミノ酸配列が基準ポリペプチドとは異なる。一般に、改変は基準ポリペプチドおよび変異体の配列が全体的に非常に類似し、そして多くの領域において同一であるように制限される。変異体ポリペプチドおよび基準ポリペプチドは、アミノ酸配列が、１つまたは複数の置換、挿入、欠失の任意の組合せによって異なりうる。置換または挿入されるアミノ酸残基は、遺伝子コードによってコードされるアミノ酸残基であっても、あるいはなくてもよい。典型的な保存的置換には、Ｇｌｙ、Ａｌａ；Ｖａｌ、Ｉｌｅ、Ｌｅｕ；Ａｓｐ、Ｇｌｕ；Ａｓｎ、Ｇｌｎ；Ｓｅｒ、Ｔｈｒ；Ｌｙｓ、Ａｒｇ；ならびにＰｈｅおよびＴｙｒが含まれる。ポリヌクレオチドまたはポリペプチドの変異体は、対立遺伝子などの天然に存在するものであってもよく、あるいは天然に存在することが知られていない変異体であってもよい。ポリヌクレオチドおよびポリペプチドの天然に存在しない変異体は、変異誘発方法によって、あるいは直接的な合成によって作製することができる。また、１つ以上の翻訳後修飾、例えば、グリコシル化、リン酸化、メチル化、ＡＤＰリボシル化などを有するポリペプチドもまた変異体として含まれる。実施形態には、Ｎ末端アミノ酸のメチル化、セリンおよびトレオニンのリン酸化、ならびにＣ末端グリシンの修飾が含まれる。

【００６７】

「対立遺伝子」は、ゲノム内の所与遺伝子座に存在する遺伝子の２つ以上の選択的な形態の１つをいう。

【００６８】

「多型」は、集団内のゲノムにおける所与の位置での塩基配列（関連する場合にはコードされるポリペプチド配列）の変動をいう。

【００６９】

「単一塩基多型」（ＳＮＰ）は、集団内においてゲノム内の１つの塩基位置における塩基変動の発生をいう。ＳＮＰは遺伝子内にまたはゲノムの遺伝子間領域内に存在し得る。ＳＮＰは対立遺伝子特異的増幅（ＡＳＡ）を使用してアッセイすることができる。該プロセスには少なくとも３つのプライマーが必要である。共通プライマーがアッセイされる多型に対して逆方向に相補となるように使用される。この共通プライマーは多型の塩基から５０ｂｐから１５００ｂｐの間で隔たったものとできる。それ以外の２つ（またはそれ以上）のプライマーは、最後の３’塩基が、多型を構成する２つ（またはそれ以上）の対立遺伝子の１つと一致するように変化している点を除いて互いに同一である。そして、それぞれ、共通プライマーおよび１つの対立遺伝子特異的プライマーを使用して、２つ（または

それ以上）のＰＣＲ反応を、サンプルＤＮＡについて行う。

【００７０】

本発明において用いられる「スプライス変異体」は、同じゲノムＤＮＡ配列から一旦転写されたが、但し、択一的なＲＮＡスプライシングを受けたＲＮＡ分子から生成したｃＤＮＡを意味する。択一的なＲＮＡスプライシングは、一般にはイントロンを除くために一次ＲＮＡ転写物がスプライシングを受けているときに生じる。その結果、それぞれが異なるアミノ酸配列をコードし得る１つ以上のｍＲＮＡ分子が生じる。スプライス変異体の用語はまた、上記のｃＤＮＡ分子によってコードされるタンパク質をも示す。

【００７１】

「同一性」は、その配列を比較することによって決定される、２つ以上のポリペプチド配列または２つ以上のポリヌクレオチド配列の間における相互関係を反映する。一般に、同一性とは、比較されている配列の長さにわたって２つのポリヌクレオチド配列または２つのポリペプチド配列のそれぞれの塩基毎またはアミノ酸毎の厳密な一致をいう。

【００７２】

「％同一性」－正確な一致が存在しない配列については、「％同一性」を決定することがある。一般に、対比すべき２つの配列を、最大の相関が配列間に得られるようにアラインメントされる。これには、アラインメントの程度を高めるために「ギャップ」をいずれか一方または両方の配列に挿入することが含まれ得る。％同一性は、比較されている配列のそれぞれの長さ全体にわたって決定することができる（いわゆる全体的アラインメント）：これは同じ長さまたは非常に類似する長さの配列の場合には特に好適である；あるいはより短い限定された長さにわたって決定することができる（いわゆる局所的アラインメント）：これは長さが等しくない配列の場合にはより好適である。

【００７３】

「類似性」は、２つのポリペプチド配列の間における関係に対する、より精巧な、さらなる尺度である。一般に「類似性」は、残基毎に基づき、（同一性に関してと、同様に）比較されている配列それぞれからの、相互に対をなす残基間における正確な一致だけでなく、正確な一致が存在しない場合にも、進化的な基準に基づいて、１つの残基は、他方に対する適当な置換であるかどうかをも考慮する、２つのポリペプチド鎖のアミノ酸間での比較を意味する。この蓋然性は付随した「スコア」を有し、２つの配列の「％類似性」はそれに基づき決定することができる。

【００７４】

２つ以上の配列の同一性および類似性を比較する方法は当該分野では周知となっている。従って、例えば、ウイスコンシン配列分析パッケージ（バージョン９．１；Ｄｅｖｅｒｅｕｘ Ｊ．他、Ｎｕｃｌｅｉｃ Ａｃｉｄｓ Ｒｅｓ．１２、３８７～３９５、１９８４；Ｇｅｎｅｔｉｃｓ Ｃｏｍｐｕｔｅｒ Ｇｒｏｕｐ（Ｍａｄｉｓｏｎ、Ｗｉｓｃｏｎｓｉｎ、米国）から入手可能）において利用可能なプログラム、例えば、ＢＥＳＴＦＩＴおよびＧＡＰ プログラムを使用して、２つのポリヌクレオチド間の％同一性ならびに２つのポリペプチド配列間の％同一性および％類似性を決定することができる。ＢＥＳＴＦＩＴでは、ＳｍｉｔｈおよびＷａｔｅｒｍａｎの「局所的相同性」アルゴリズム（Ｊ Ｍｏｌ Ｂｉｏｌ、１４７、１９５～１９７、１９８１、Ａｄｖａｎｃｅｓ ｉｎ Ａｐｐｌｉｅｄ Ｍａｔｈｅｍａｔｉｃｓ、２、４８２～４８９、１９８１）を用いて、２つの配列間における類似性の最も良い１つの領域が見出される。ＢＥＳＴＦＩＴは、長さが類似していない２つのポリヌクレオチド配列または２つのポリペプチド配列を比較することに対してより適している。該プログラムでは、より短い配列がより長い配列の一部を表すことが仮定されている。一方、ＧＡＰは、ＮｅｄｄｌｅｍａｎおよびＷｕｎｓｃｈのアルゴリズム（Ｊ Ｍｏｌ Ｂｉｏｌ、４８、４４３～４５３、１９７０）に従って２つの配列をアラインメントし、これにより「最大の類似性」を見出す。ＧＡＰは、ほぼ同じ長さであり、また、アラインメントが長さ全体にわたって期待される配列を比較することに対してより適している。好ましくは、各プログラムにおいて使用される「ギャップ加重（Ｇａｐ Ｗｅｉｇｈｔ）」および「長さ加重（Ｌｅｎｇｔｈ Ｗｅｉｇｈｔ）」のパラメー

ターは、それぞれ、ポリヌクレオチド配列の場合には５０および３であり、ポリペプチド配列の場合には１２および４である。好ましくは、比較されている２つの配列が最適にアラインメントした上で％同一性および類似性を決定される。

【００７５】

配列間の同一性および／または類似性を決定するための他のプログラムもまた当該分野では知られている：例えば、ＢＬＡＳＴファミリーのプログラム（Ａｌｔｓｃｈｕｌ　ＳＦ他、Ｊ　Ｍｏｌ　Ｂｉｏｌ、２１５、４０３～４１０、１９９０；Ａｌｔｓｃｈｕｌ　ＳＦ他、Ｎｕｃｌｅｉｃ　Ａｃｉｄｓ　Ｒｅｓ．、２５：３８９～３４０２、１９９７；これはＮａｔｉｏｎａｌ　Ｃｅｎｔｅｒ　ｆｏｒ　Ｂｉｏｔｅｃｈｎｏｌｏｇｙ　Ｉｎｆｏｒｍａｔｉｏｎ（ＮＣＢＩ）（Ｂｅｔｈｅｓｄａ、Ｍａｒｙｌａｎｄ、米国）から入手可能であり、ｗｗｗ．ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖにおけるＮＣＢＩのホームページからアクセス可能である）、およびＦＡＳＴＡ（Ｐｅａｒｓｏｎ　ＷＲ、Ｍｅｔｈｏｄｓ　ｉｎ　Ｅｎｚｙｍｏｌｏｇｙ、１８３、６３～９９、１９９０；Ｐｅａｒｓｏｎ　ＷＲおよびＬｉｐｍａｎ　ＤＪ、Ｐｒｏｃ．Ｎａｔ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、８５、２４４４～２４４８、１９８８；これはウイスコンシン配列分析パッケージの一部として入手可能である）。

【００７６】

好ましくは、ＢＬＯＳＵＭ６２アミノ酸置換行列（Ｈｅｎｉｋｏｆｆ　ＳおよびＨｅｎｉｋｏｆｆ　ＪＧ、Ｐｒｏｃ．Ｎａｔ．Ａｃａｄ．Ｓｃｉ．ＵＳＡ、８９、１０９１５～１０９１９、１９９２）が、比較前にヌクレオチド配列がアミノ酸配列に最初に翻訳される場合をも含みポリペプチド配列比較において使用される。

【００７７】

好ましくは、プログラムＢＥＳＴＦＩＴが、基準ポリヌクレオチド配列またはポリペプチド配列に関して、検討するポリヌクレオチドまたはポリペプチド配列の％同一性を決定するために使用される。この場合、前記したように、検討と基準とする配列とは、最適にアラインメントされ、そしてプログラムのパラメーターは暫定の値に設定されている。

【００７８】

「同一性指標」は、候補配列（ポリヌクレオチドまたはポリペプチド）と基準配列とを比較するために使用することができる、配列関連性の尺度である。したがって、例えば基準ポリヌクレオチド配列と比較したときに、例えば０．９５の同一性指標を有する候補ポリヌクレオチド配列は、該候補ポリヌクレオチド配列は基準配列の各１００塩基あたり平均して５個までの相違を含んでもよい点を除いて基準配列と同一である。かかる相違は、少なくとも１つの塩基の欠失、トランジションおよびトランスバージョンを含む置換、または挿入からなる群から選択される。これらの相違は、基準ポリヌクレオチド配列の５’または３’末端部位に、あるいはこれらの末端部位間における任意の位置で、基準配列内の塩基間に個々に、あるいは基準配列内に１つまたは複数の連続した群で点在して存在してもよい。すなわち、基準ポリヌクレオチド配列と比較したときに０．９５の同一性指標を有するポリヌクレオチド配列を得るためには、既に記載のように、基準配列において塩基１００個毎に平均して５個までが任意の組合せで欠失、置換または挿入されていてもよい。同じことが同一性指標の他の値、例えば０．９６、０．９７、０．９８および０．９９についても必要に応じて変更して適用される。

【００７９】

同様に、ポリペプチドの場合、基準ポリペプチド配列と比較したときに、例えば０．９５の同一性指標を有する候補ポリペプチド配列は、基準配列の各１００アミノ酸あたり平均して５個までの相違を該ポリペプチド配列が含んでよいことを除いて基準配列と同一である。かかる相違は、少なくとも１つのアミノ酸の欠失、保存的および非保存的な置換を含む置換、または挿入からなる群から選択される。これらの違いは、基準ポリペプチド配列のアミノ末端またはカルボキシ末端部位に、あるいはこれらの末端部位間の任意の位置に、基準配列内のアミノ酸間に個々に、あるいは基準配列内に１つまたは複数の連続した群中で点在して起きてもよい。すなわち、基準ポリペプチド配列と比較した際に０．９５の

同一性指標を有するポリペプチド配列を得るためには、既に記載のように、基準配列内のアミノ酸の１００個毎に平均して５個まで任意の組み合せで欠失、置換または挿入がなされていてもよい。同じことが同一性指標の他の値、例えば、０．９６、０．９７、０．９８および０．９９についても必要に応じて変更して適用される。

【００８０】

塩基またはアミノ酸の相違数と同一性指標との関係は下記の式で表すことができる：

$n_a \leqq x_a - (x_a \cdot I)$

式中、

n_aは塩基またはアミノ酸の相違数であり、

x_aは配列番号１または配列番号２における塩基またはアミノ酸のそれぞれの総数であり、

Ｉは同一性指標であり、

・は乗算演算子に対する記号であり、

この場合、x_aとＩとの非整数の積は最も近い整数に切り捨てられ、その後x_aからその値を引く。

【００８１】

「ホモログ」は、基準配列に対する高度の配列関連性を有するポリヌクレオチド配列またはポリペプチド配列を示すために、当該分野で使用されている総称である。かかる関連性は、既に定義されているように２つの配列間の同一性および／または類似性の程度を決定することによって定量化することができる。この総称には、「オルソログ」および「パラログ」の用語が含まれる。「オルソログ」は、別の種中における該ポリヌクレオチドまたはポリペプチドの機能的等価体であるポリヌクレオチドまたはポリペプチドをいう。「パラログ」は、機能的に類似している同じ種内におけるポリヌクレオチドまたはポリペプチドをいう。

【００８２】

「融合タンパク質」は、２つの関連しない融合された遺伝子またはそのフラグメントによってコードされるタンパク質をいう。その例が米国特許第５５４１０８７号、米国特許第５７２６０４４号に開示されている。Ｆｃ－ＳＩＰ－１リガンドの場合、融合タンパク質の一部として免疫グロブリンのＦｃ領域を用いることは、Ｆｃ－ＳＩＰ－１リガンドまたはリガンドのフラグメントを機能的に発現させて、治療に使用する際のかかる融合タンパク質の薬物動態学的性質を改善するために、ならびに二量体のＳＩＰ－１リガンドを生成させるためには好都合である。Ｆｃ－ＳＩＰ－１リガンドのＤＮＡ構築物は、５’から３’の方向で、分泌カセット、すなわち、哺乳動物細胞からの細胞外への輸送を引き起こすシグナル配列、融合パートナーとして免疫グロブリンのＦｃ領域フラグメントをコードするＤＮＡ、およびＳＩＰ－１リガンドをコードするＤＮＡまたはその断片を含む。ある用途では、融合タンパク質の残部には手を触れることなく、機能的なＦｃ側を変異させ、その固有的な機能的性質（補体結合、Ｆｃ受容体結合）を変える、あるいは発現後にＦｃ部分を完全に除くことを可能とすることが望ましい。

【００８３】

特許および特許出願に限らず、これらを含む、本明細書中で引用されている刊行物および参考文献の全ては、十分に述べているように、個々の刊行物または参考文献を、参照して組み込むために、明示的かつ個別的に示唆されているように、その全部を、参照することで、本明細書中に組み込まれる。本出願が優先権を主張する特許出願はいずれも、先に刊行物および参考文献に関して記載したと同様に、その全部を、参照することで、本明細書中に組み込まれる。

【００８４】

（図面の説明）

（図１）

Ｓ．ｃｅｒｅｖｉｓｉａｅ　ＥＧＹ４８／ｐＳＨ１８－３２におけるサバイビンとＳＩＰ－１との相互作用を示す図である。

【００８５】
Ｓ．ｃｅｒｅｖｉｓｉａｅ　ＥＧＹ４８／１８－３２（Ｃｌｏｎｔｅｃｈ）を、ＬｅｘＡ－およびＧａｌ４－活性化ドメイン融合タンパク質をコードするプラスミドを使用して形質転換した。単一コロニーを単離して、ヒスチジン、トリプトファンおよびウラシルを含まない培地（－ＷＨＵ）に、そしてヒスチジン、トリプトファン、ロイシンおよびウラシルを含まず、４０ｍｇ／ｌのＸ－Ｇａｌを含む培地（－ＷＨＬＵ－ＸＧａｌ）に画線培養した。すべての菌株が－ＷＨＵ培地で増殖したことにより、プラスミドの存在が示される。レポーター遺伝子の発現はツーハイブリッドセレクションスキームに依存しているので、ロイシンを含まない培地における酵母の青色コロニーおよび増殖はおとりとえものとの相互作用を示している。レーン１および２が不特定のコントロールとして使用される。レーン３は、この酵母ツーハイブリッドシステムにおけるサバイビンとＳＩＰ－１との特異的相互作用を示している。

【００８６】
（図２）
ＳＩＰ－１ノーザンブロットを示す図である：ランダムプライム処理した^{32}Ｐ－放射能標識ＳＩＰ－１プローブをハイブリダイゼーションのために使用した。このプローブは、ＳＩＰ－１ｃｄｓの５５３～１１４３位に対応する５９０塩基対からなる。ブロットには、レーンあたり２μｇのポリ（Ａ）ＲＮＡが負荷された（Ｉｎｖｉｔｒｏｇｅｎから得られたブロット：「リアルヒト腫瘍パネルブロット３＃Ｄ３８０３－５０）。

【００８７】
２．８ｋｂのＳＩＰ－１　ｍＲＮＡが、腎臓の正常組織、腎臓の腫瘍組織、肺の正常組織および肺の腫瘍組織のｍＲＮＡサンプルにおいて検出される。

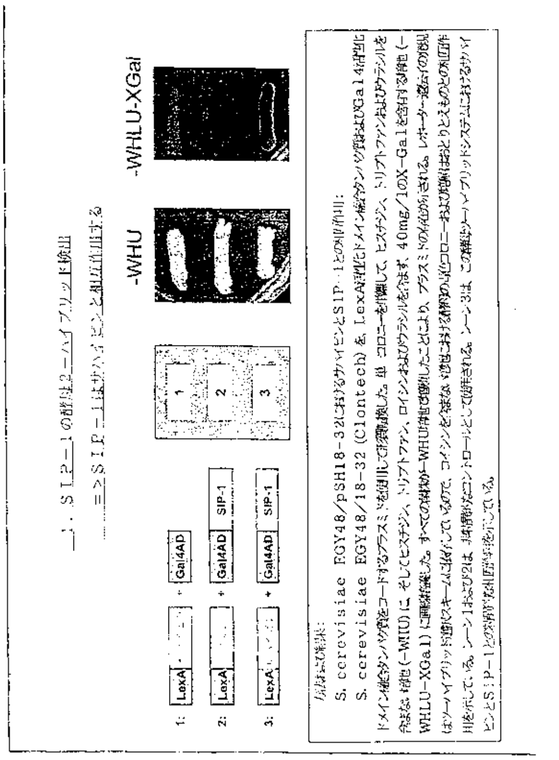
【００８８】
（図３）
「遺伝子特異的相対的ＲＴ－ＰＣＲキット」を使用する、ヒトＨｅＬａ細胞に由来するｍＲＮＡの多重ＰＣＲ（プロトコルはこのキットシステムの製造者（Ａｍｂｉｏｎ）に従う）を示す図である：
ヒトＨｅＬａ細胞の細胞期を、下記の処理を使用して停止させた（Ｇ１期：Ｌ－ミモシン［４００μＭ］；Ｓ期：チミジン［２ｍＭ］；Ｇ２／Ｍ期：ノコダゾール［４０ｎｇ／ｍｌ］）。細胞周期状態はＦＡＣＳ分析によって制御された。コントロール細胞は１ｍｌの培地あたり１μｌのＤＭＳＯで処理された。ＳＩＰ－１のｍＲＮＡ発現は、細胞周期のＳ期およびＧ２／Ｍ期のときにアップレギュレーションされているようである。

【００８９】
（図４）
ＳＩＰ－１／サバイビンの同時免疫沈降を示す図である：ＳＩＰ－１およびサバイビンは、インビトロ翻訳されたタンパク質として発現させた後、ヒトＨｅＬａ細胞において相互作用する。Ｖ５抗体により媒介されるＳＩＰ－１／サバイビン相互作用の分析は免疫沈降（ＩＰ）アッセイにより行われた。

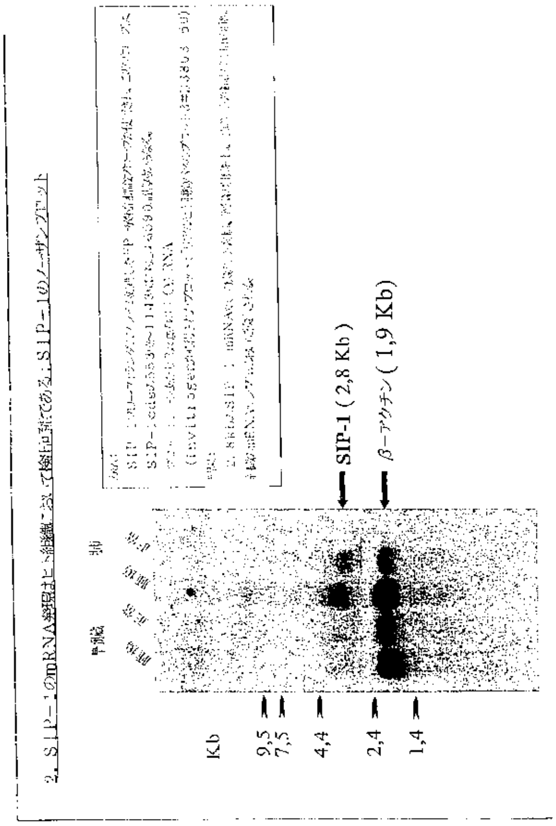
【００９０】
（実施例）
プラスミド構築
ｐＤＢＬｅｕ－サバイビンのクローニング：サバイビン遺伝子を、ＮｈｅＩおよびＳｔｕＩの制限部位を組み込んだ遺伝子特異的ＰＣＲプライマー対を用いてヒトＰＣ－３細胞のｃＤＮＡからクローン化し、サバイビンのコード配列を増幅した。ＰＣＲ産物をＴＡクローニングによってｐＣＲＩＩベクター（Ｉｎｖｉｔｒｏｇｅｎ）にクローン化した。インサートをＮｈｅＩ／ＳｔｕＩ制限によって単離し、Ｙ２ＨベクターのｐＤＢＬｅｕ（ＰＲＯＱＵＥＳＴ（登録商標）ツーハイブリッドシステム、ＧｉｂｃｏＢＲＬ　ＬＩＦＥ　ＴＥＣＨＮＯＬＯＧＩＥＳ）にクローン化して、ｐＤＢＬｅｕ－サバイビンを作製した。これにより、全長のサバイビンペプチドに融合したＧａｌ４ＤＢドメインを発現させるための融合遺伝子構築物が作製された。すべてのベクターは配列決定により確認された。

【００９１】
サバイビン遺伝子をクローニングするためのＰＣＲプライマーは下記のプライマーであった：
ＮｈｅＩ－５’サバイビンプライマー：プライマー１
ＳｔｕＩ－３’サバイビンプライマー：プライマー２
サバイビンと相互作用するタンパク質の選択
ｐＤＢＬｅｕ－サバイビンをおとり構築物として使用する酵母ツーハイブリッドスクリーニング（Ｐｒｏｑｕｅｓｔ、Ｌｉｆｅ　Ｔｅｃｈｎｏｌｏｇｉｅｓ）を、ヒトＨｅＬａ細胞のｃＤＮＡ－Ｙ２Ｈライブラリとヒト胎児脳組織に由来するｃＤＮＡライブラリとを使用して行った（セレクションは、２５ｍＭの３－アミノトリアゾールを含有する－Ｔｒｐ、－Ｌｅｕ、－Ｈｉｓ最少培地において行われた）。ＳＩＰ－１クローンが両方のライブラリから単離された。相互作用は、β－ガラクトシダーゼフィルターアッセイとウラシル、製造者のプロトコルに記載されるように、トリプトファンおよびロイシンを含まない培地によって確認された。
【００９２】
１つの陽性クローンが単離され、配列決定された。対応する相互作用タンパク質リガンドをＳＩＰ－１（サバイビン相互作用タンパク質１）リガンドと名付けた。配列は配列番号１および配列番号２に開示されている。
【図面の簡単な説明】
【図１】
Ｓ．ｃｅｒｅｖｉｓｉａｅ　ＥＧＹ４８／ｐＳＨ１８－３２におけるサバイビンとＳＩＰ－１との相互作用を示す図である。
【図２】
ＳＩＰ－１ノーザンブロットを示す図である。
【図３】
「遺伝子特異的相対的ＲＴ－ＰＣＲキット」を使用する、ヒトＨｅＬａ細胞に由来するｍＲＮＡの多重ＰＣＲ（プロトコルはこのキットシステムの製造者（Ａｍｂｉｏｎ）に従う）を示す図である。
【図４】
ＳＩＰ－１／サバイビンの同時免疫沈降を示す図である。

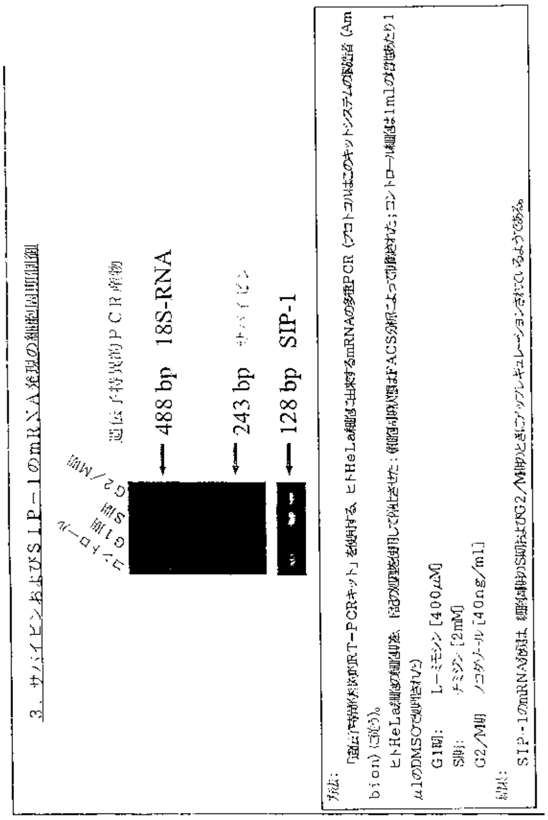
【図１】


-WHU
-WHLU-XGal
LexA
Gal4AD
SIP-1

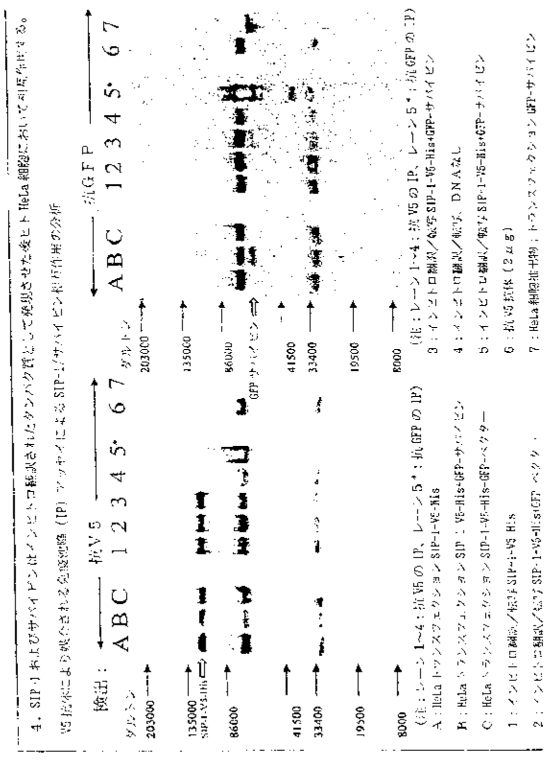
【図２】


SIP-1 (2,8 Kb)
β-アクチン (1,9 Kb)
Kb
9,5
7,5
4,4
2,4
1,4

【図３】


488 bp 18S-RNA
243 bp
128 bp SIP-1

【図４】


抗V5
抗GFP

【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
4 October 2001 (04.10.2001)

PCT

(10) International Publication Number
WO 01/73014 A1

(51) International Patent Classification⁷: C12N 15/12, C07K 14/47, C12N 15/62, C07K 16/18

(21) International Application Number: PCT/EP01/03204

(22) International Filing Date: 21 March 2001 (21.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00106108.8 24 March 2000 (24.03.2000) EP

(71) Applicant (*for all designated States except US*): **MERCK PATENT GMBH** [DE/DE]; Frankfurter Strasse 250, 64293 Darmstadt (DE).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **HENTSCH, Bernd** [DE/DE]; Kampstrasse 48, 64289 Darmstadt (DE). **DÜCKER, Klaus** [DE/DE]; Rheinstrasse 5, 64293 Darmstadt (DE). **HOCK, Björn** [DE/DE]; Nordstrasse 2a, 63477 Maintal (DE).

(74) Common Representative: **MERCK PATENT GMBH**, 64271 Darmstadt (DE).

(81) Designated States (*national*): CA, JP, US.

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

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 01/73014 A1

(54) Title: HUMAN SURVIVIN INTERACTING PROTEIN 1 (SIP-1)

(57) Abstract: Survivin Interacting Protein 1 and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing SIP 1 polypeptides and polynucleotides in diagnostic assays.

HUMAN SURVIVIN INTERACTING PROTEIN 1 (SIP-1)

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides sometimes hereinafter referred to as Survivin Interacting Protein 1 or „SIP-1", , to their use in diagnosis and in identifying compounds that may be agonists, antagonists that are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces "functional genomics", that is, high throughput genome- or gene-based biology. This approach as a means to identify genes and gene products as therapeutic targets is rapidly superseding earlier approaches based on "positional cloning". A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery. Recent advances in this area are mainly driven by applying advanced screening systems with focus to the function of newly identified proteins. Interesting application for this type of screens are whithin the field of tumor apoptosis.

Apoptosis of tumor cells can be blocked e.g. by mutations in crucial genes like those frequently found in the well known p53 tumor suppressor protein. Another mechanism that blocks apoptosis is caused by the protein family of „inhibitors of apoptosis" (IAPs), reviewed by Deveraux

and Reed (Genes & Development 1999). One member of this family is encoded by the *Survivin* gene.

The strongest evidence for an IAP involvement in cancer is seen for *Survivin* (see review of Altieri and Marchisio, Laboratory Investigation 1999). Although not observed in adult differentiated tissue, *Survivin* becomes prominently expressed in transformed cell lines and in all of the most common human cancers of lung, colon, pancreas, prostate and breast, in vivo. *Survivin* is also found in approximately 50% of high-grade non-Hodgkin's lymphomas (centroblastic, immunoblastic), but not in low-grade lymphomas (lymphocytic). *Survivin* inhibits caspase activity and apoptosis induced by Fas (CD95), Bax, Caspases, and anti-cancer drugs.

In addition, *Survivin* is upregulated 40-fold at G2/M phase of the cell cycle and binds to mitotic spindles (micro-tubules), although its role at the spindle is still unclear. There might be a connected control of apoptosis and mitotic spindle checkpoint by *Survivin*. Disruption of *Survivin*-microtubule interactions results in loss of *Survivin*'s anti-apoptosis function and increased caspase-3 activity. *Survivin* may counteract a default induction of apoptosis in G2/M phase. The overexpression of *Survivin* in cancer may overcome this apoptotic checkpoint and favour aberrant progression of transformed cells through mitosis (Fengzhi et al., Nature 1998). In this repect it is important to note that it was recently shown that *Survivin* initiates procaspase 3/p21 complex formation as a result of interaction with Cdk4 to resist Fas-mediated cell death (Suzuki et al., Oncogene 2000). SIP-1 may therefore be implicated in the known tumor-cell-escape from the immune system. The other biochemical mechanisms, besides its inhibitory binding to Caspases, by which *Survivin* might mediate its anti-apoptotic activity are currently unclear.

Therefore in this invention we have screened for potential ligands that interact with *Survivin*.

Summary of the Invention

The present invention relates to Survivin Interacting Protein 1 (*GENE NAME), in particular *GENE NAME polypeptides and *GENE NAME

polynucleotides, recombinant materials and methods for their production. Such polypeptides and polynucleotides are of interest in relation to methods of treatment of certain diseases, because they are natural ligands for Survivin. As pointed out this molecule has a prominent role in anti-apoptotic activity.

Survivin becomes prominently expressed in most common human cancers of lung, liver, colon, stomach, skin, pancreas, prostate, ovary and breast, in vivo. *Survivin* is also found in approximately 50% of high-grade non-Hodgkin's lymphomas thus the interaction with the newly described molecule of this invention SIP-1 offers an attractive option to interfere with this diseases therefore they are hereinafter referred to as "diseases of the invention". In a further aspect, the invention relates to methods for identifying agonists and antagonists (*e.g.*, inhibitors) using the materials provided by the invention, and treating conditions associated with *GENE NAME activity with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate *GENE NAME activity or levels.

Description of the Invention

In a first aspect, the present invention relates to *GENE NAME polypeptides. Such polypeptides include:

(a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1;

(b) an polypeptide comprising a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(c) a polypeptide comprising the polypeptide sequence of SEQ ID NO:2;

(d) a polypeptide having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(e) the polypeptide sequence of SEQ ID NO:2; and

(f) a polypeptide having or comprising a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2;

(g) fragments and variants of such polypeptides in (a) to (f).

Polypeptides of the present invention are believed to represent ligands for *Survivin* and thus may counteract a default induction of apoptosis. Furthermore Survivin and its ligand SIP-1 may control apoptotic checkpoint and SIP-1 may be a mediator of Survivin's anti-apoptotic activity.

SIP-1 polypeptides may be involved in the establishment of a survival signal of tumor cells which may result in resistance of these cells against apoptotic signals, e.g. chemotherapeutic agents or physiological death signals as such of immune cells.

The biological properties of the *GENE NAME are hereinafter referred to as "biological activity of *GENE NAME" or "*GENE NAME activity". Preferably, a polypeptide of the present invention exhibits at least one biological activity of *GENE NAME.

Polypeptides of the present invention also includes variants of the aforementioned polypeptides, including all allelic forms and splice variants. Such polypeptides vary from the reference polypeptide by insertions, deletions, and substitutions that may be conservative or non-conservative, or any combination thereof. Particularly preferred variants are those in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acids are inserted, substituted, or deleted, in any combination.

Preferred fragments of polypeptides of the present invention include a polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO: 2, or a polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids truncated or deleted from the amino acid sequence of SEQ ID NO: 2. Preferred fragments are biologically active fragments that mediate the biological activity of *GENE NAME, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also preferred are those fragments that are antigenic or immunogenic in an animal, especially in a human.

Fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis;

therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention. The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence that contains secretory or leader sequences, pro-sequences, sequences that aid in purification, for instance multiple histidine residues, or an additional sequence for stability during recombinant production.

Polypeptides of the present invention can be prepared in any suitable manner, for instance by isolation form naturally occurring sources, from genetically engineered host cells comprising expression systems (*vide infra*) or by chemical synthesis, using for instance automated peptide synthesizers, or a combination of such methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to *GENE NAME polynucleotides. Such polynucleotides include:

(a) a polynucleotide comprising a polynucleotide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide comprising the polynucleotide of SEQ ID NO:1;

(c) a polynucleotide having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide of SEQ ID NO:1;

(d) the polynucleotide of SEQ ID NO:1;

(e) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(f) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2;

(g) a polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(h) a polynucleotide encoding the polypeptide of SEQ ID NO:2;

(i) a polynucleotide having or comprising a polynucleotide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polynucleotide sequence of SEQ ID NO:1;

(j) a polynucleotide having or comprising a polynucleotide sequence encoding a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2; and

polynucleotides that are fragments and variants of the above mentioned polynucleotides or that are complementary to above mentioned polynucleotides, over the entire length thereof.

Preferred fragments of polynucleotides of the present invention include a polynucleotide comprising an nucleotide sequence having at least 15, 30, 50 or 100 contiguous nucleotides from the sequence of SEQ ID NO: 1, or a polynucleotide comprising an sequence having at least 30, 50 or 100 contiguous nucleotides truncated or deleted from the sequence of SEQ ID NO: 1.

Preferred variants of polynucleotides of the present invention include splice variants, allelic variants, and polymorphisms, including polynucleotides having one or more single nucleotide polymorphisms (SNPs).

Polynucleotides of the present invention also include polynucleotides encoding polypeptide variants that comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acid residues are substituted, deleted or added, in any combination.

In a further aspect, the present invention provides polynucleotides that are RNA transcripts of the DNA sequences of the present invention. Accordingly, there is provided an RNA polynucleotide that:

(a) comprises an RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2;

(b) is the RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2;

(c) comprises an RNA transcript of the DNA sequence of SEQ ID NO:1; or

(d) is the RNA transcript of the DNA sequence of SEQ ID NO:1;

and RNA polynucleotides that are complementary thereto.

The polynucleotide sequence of SEQ ID NO:1 is a cDNA sequence that encodes the polypeptide of SEQ ID NO:2. The polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence of SEQ ID NO:1 or it may be a sequence other than SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is related to other proteins of the family of guanine exchange factors (GEFs), having homology and/or structural similarity with Homo sapiens Duo mRNA, which encodes a spectrin-like domain, coiled coil, guanosine exchange factor domain and pleckstrin homology domain [Colomer, V., Engelender, S., Sharp, A.H., Duan, K., Cooper, J.K., Lanahan, A., Lyford, G., Worley, P. and Ross, C.A.: Huntingtin-associated protein 1 binds to a Trio-like polypeptide, with a rac1 guanine nucleotide exchange factor domain. Hum. Mol. Genet. 6 (9), 1519-1525 (1997)] and to Homo sapiens Trio mRNA which displays similarities to protein kinase, *rac* guanine exchange factor, *rho* guanine exchange factor and spectrin-like repeats [Debant, A., Serra-Pages, C., Seipel, K., O'Brien, S., Tang, M., Park, S.H. and Streuli, M.: The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. Proc. Natl. Acad. Sci. U.S.A. 93 (11), 5466-5471, (1996)].

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one *GENE NAME activity.

Polynucleotides of the present invention may be obtained using standard cloning and screening techniques from a cDNA library derived from mRNA of human HeLa cells (ATCC: CCL-2) and in a cDNA library derived from human fetal brain tissue (see for instance, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself, or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Polynucleotides that are identical, or have sufficient identity to a polynucleotide sequence of SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification reaction (for instance, PCR). Such probes and primers may be used to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1, typically at least 95% identity. Preferred probes and primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50, if not at least 100 nucleotides. Particularly preferred probes will have

between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process comprising the steps of screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof, preferably of at least 15 nucleotides; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides, preferably with a nucleotide sequence of at least 100, obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof, preferably of at least 15 nucleotides.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide does not extend all the way through to the 5' terminus. This is a consequence of reverse transcriptase, an enzyme with inherently low "processivity" (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during first strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., Proc Nat Acad Sci USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon (trade mark) technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon (trade mark) technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to

amplify the "missing" 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adapter specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analyzed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems comprising a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Polynucleotides may be introduced into host cells by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook *et al.*(*ibid*). Preferred methods of introducing polynucleotides into host cells include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, micro-injection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci*, *Staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as

CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector that is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate polynucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, (*ibid*). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate

active conformation when the polypeptide is denatured during intracellular synthesis, isolation and/or purification.

Polynucleotides of the present invention may be used as diagnostic reagents, through detecting mutations in the associated gene. Detection of a mutated form of the gene characterized by the polynucleotide of SEQ ID NO:1 in the cDNA or genomic sequence and 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 well known in the art.

Nucleic acids 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 it may be amplified enzymatically by using PCR, preferably RT-PCR, or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled *GENE NAME nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence difference 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 (see, for instance, 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).

An array of oligonucleotides probes comprising *GENE NAME polynucleotide sequence or fragments thereof can be constructed to conduct efficient screening of *e.g.*, genetic mutations. Such arrays are preferably high density arrays or grids. 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, 274, 610-613 (1996) and other references cited therein.

Detection of abnormally decreased or increased levels of polypeptide or mRNA expression may also be used for diagnosing or determining susceptibility of a subject to a disease of the invention. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the 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 protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radio-immunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit comprising:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment or an RNA transcript thereof;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a disease, particularly diseases of the invention, amongst others.

The polynucleotide sequences of the present invention are valuable for chromosome localisation studies. The sequence 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 first step in correlating those sequences with 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 relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (co-inheritance of physically adjacent genes). Precise human chromosomal localisations for a genomic sequence (gene fragment etc.) can be determined using Radiation Hybrid (RH) Mapping (Walter, M. Spillett, D., Thomas, P., Weissenbach, J., and Goodfellow, P., (1994) A method for constructing radiation hybrid maps of whole genomes, Nature Genetics 7, 22-28). A number of RH panels are available from Research Genetics (Huntsville, AL, USA) e.g. the GeneBridge4 RH panel (Hum Mol Genet 1996 Mar;5(3):339-46 A radiation hybrid map of the human genome. Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN). To determine the chromosomal location of a gene using this panel, 93 PCRs are performed using primers designed from the gene of interest on RH DNAs. Each of these DNAs contains random human genomic fragments maintained in a hamster background (human / hamster hybrid cell lines). These PCRs result in 93 scores indicating the presence or absence of the PCR product of the gene of interest. These scores are compared with scores created using PCR products from genomic sequences of known location. This comparison is conducted at http://www.genome.wi.mit.edu/. The gene of the present invention maps to human chromosome Chr. 2 (D2S335-D2S2257).

The polynucleotide sequences of the present invention are also valuable tools for tissue expression studies. Such studies allow the determination of expression patterns of polynucleotides of the present invention which may give an indication as to the expression patterns of the encoded polypeptides in tissues, by detecting the mRNAs that encode them. The techniques used are well known in the art and include in situ hydridization techniques to clones arrayed on a grid, such as cDNA microarray hybridization (Schena *et al*, Science, 270, 467-470, 1995 and Shalon *et al*, Genome Res, 6, 639-645, 1996) and nucleotide amplification techniques such as PCR. A preferred method uses the TAQMAN (Trade mark) technology available from Perkin Elmer. Results from these studies can provide an indication of the normal function of the polypeptide in the

organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by an alternative form of the same gene (for example, one having an alteration in polypeptide coding potential or a regulatory mutation) can provide valuable insights into the role of the polypeptides of the present invention, or that of inappropriate expression thereof in disease. Such inappropriate expression may be of a temporal, spatial or simply quantitative nature.

A further aspect of the present invention relates to antibodies. The polypeptides of the invention or their fragments, or cells expressing them, can be used as immunogens to produce antibodies that are immunospecific for polypeptides of the present invention. 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.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology Today (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography. Antibodies against polypeptides of the present invention may also be employed to treat diseases of the invention, amongst others.

Polypeptides and polynucleotides of the present invention may also be used as vaccines. Accordingly, in a further aspect, the present invention

relates to a method for inducing an immunological response in a mammal that comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said animal from disease, whether that disease is already established within the individual or not. An immunological response in a mammal may also be induced by a method comprises delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases of the invention. One way of administering the vector is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid. For use a vaccine, a polypeptide or a nucleic acid vector will be normally provided as a vaccine formulation (composition). The formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intra-muscular, intravenous, or intra-dermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions that may contain anti-oxidants, buffers, bacteriostats and solutes that render the formulation instonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions that may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose 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 vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention have one or more biological functions that are of relevance in one or more disease states, in particular the diseases of the invention hereinbefore mentioned. It is therefore useful to identify compounds that stimulate or inhibit the function or level of the

polypeptide. Accordingly, in a further aspect, the present inven provides for a method of screening compounds to identify those stimulate or inhibit the function or level of the polypeptide. Such meth identify agonists or antagonists that may be employed for therapeutic prophylactic purposes for such diseases of the invention as hereinbe mentioned. Compounds may be identified from a variety of sources, example, cells, cell-free preparations, chemical libraries, collections chemical compounds, and natural product mixtures. Such agonists antagonists so-identified may be natural or modified substrates, ligar receptors, enzymes, etc., as the case may be, of the polypeptide structural or functional mimetic thereof (see Coligan *et al.*, Cur Protocols in Immunology 1(2):Chapter 5 (1991)) or a small molecule. S small molecules preferably have a molecular weight below 2,000 dalt more preferably between 300 and 1,000 daltons, and most prefer between 400 and 700 daltons. It is preferred that these small molec are organic molecules.

The screening method may simply measure the binding of a candi compound to the polypeptide, or to cells or membranes bearing polypeptide, or a fusion protein thereof, by means of a label direct indirectly associated with the candidate compound. Alternatively, screening method may involve measuring or detecting (qualitativel quantitatively) the competitive binding of a candidate compound to polypeptide against a labeled competitor (*e.g.* agonist or antagor Further, these screening methods may test whether the candi compound results in a signal generated by activation or inhibition of polypeptide, using detection systems appropriate to the cells bearing polypeptide. Inhibitors of activation are generally assayed in presence of a known agonist and the effect on activation by the ag by the presence of the candidate compound is observed. Further, screening methods may simply comprise the steps of mixing a candi compound with a solution containing a polypeptide of the pre invention, to form a mixture, measuring a *GENE NAME* activity in mixture, and comparing the *GENE NAME* activity of the mixture control mixture which contains no candidate compound.

Polypeptides of the present invention may be employed in conventi low capacity screening methods and also in high-throughput scree (HTS) formats. Such HTS formats include not only the well-establi

use of 96- and, more recently, 384-well micotiter plates but also emerg methods such as the nanowell method described by Schullek et al, A Biochem., 246, 20-29, (1997).

Fusion proteins, such as those made from Fc portion and *GENE NU polypeptide, as hereinbefore described, can also be used high-throughput screening assays to identify antagonists for polypeptide of the present invention (see D. Bennett *et al.*, J Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Ch 270(16):9459-9471 (1995)).

Screening techniques

The polynucleotides, polypeptides and antibodies to the polypeptide of present invention may also be used to configure screening methods detecting the effect of added compounds on the production of mRNA a polypeptide in cells. For example, an ELISA assay may be construc for measuring secreted or cell associated levels of polypeptide us monoclonal and polyclonal antibodies by standard methods known in art. This can be used to discover agents that may inhibit or enhance production of polypeptide (also called antagonist or agonist, respectiv from suitably manipulated cells or tissues.

A polypeptide of the present invention may be used to identify membra bound or soluble receptors, if any, through standard receptor bind techniques known in the art. These include, but are not limited to, liga binding and crosslinking assays in which the polypeptide is labeled wit radioactive isotope (for instance, ^{125}I), chemically modified (for instan biotinylated), or fused to a peptide sequence suitable for detection purification, and incubated with a source of the putative receptor (ce cell membranes, cell supernatants, tissue extracts, bodily fluids). Ot methods include biophysical techniques such as surface plasm resonance and spectroscopy. These screening methods may also used to identify agonists and antagonists of the polypeptide that comp with the binding of the polypeptide to its receptors, if any. Standa methods for conducting such assays are well understood in the art.

Examples of antagonists of polypeptides of the present invention incl antibodies or, in some cases, oligonucleotides or proteins that are clos related to the ligands, substrates, receptors, enzymes, etc., as the ca

may be, of the polypeptide, e.g., a fragment of the ligands, substra receptors, enzymes, etc.; or a small molecule that bind to the polypeptid the present invention but do not elicit a response, so that the activity of polypeptide is prevented.

Screening methods may also involve the use of transgenic technol and *GENE NAME gene. The art of constructing transgenic animal well established. For example, the *GENE NAME gene may be introdu through microinjection into the male pronucleus of fertilized oocy retroviral transfer into pre- or post-implantation embryos, or injectior genetically modified, such as by electroporation, embryonic stem c into host blastocysts. Particularly useful transgenic animals are so-ca "knock-in" animals in which an animal gene is replaced by the hur equivalent within the genome of that animal. Knock-in transgenic anin are useful in the drug discovery process, for target validation, where compound is specific for the human target. Other useful transge animals are so-called "knock-out" animals in which the expression of animal ortholog of a polypeptide of the present invention and encodec an endogenous DNA sequence in a cell is partially or comple annulled. The gene knock-out may be targeted to specific cells tissues, may occur only in certain cells or tissues as a consequence the limitations of the technology, or may occur in all, or substantially cells in the animal. Transgenic animal technology also offers a wh animal expression-cloning system in which introduced genes expressed to give large amounts of polypeptides of the present inventi

Screening kits for use in the above described methods form a furt aspect of the present invention. Such screening kits comprise:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present inventic

(c) a cell membrane expressing a polypeptide of the present invention

(d) an antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may compr a substantial component.

Glossary

The following definitions are provided to facilitate understanding of certa terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclor antibodies, chimeric, single chain, and humanized antibodies, as well Fab fragments, including the products of an

Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from its natural state, *i.* if it occurs in nature, it has been changed or removed from its origir environment, or both. For example, a polynucleotide or a polypepti naturally present in a living organism is not "isolated," but the sar polynucleotide or polypeptide separated from the coexisting materials its natural state is "isolated", as the term is employed herein. Moreov a polynucleotide or polypeptide that is introduced into an organism transformation, genetic manipulation or by any other recombinant meth is "isolated" even if it is still present in said organism, which organi may be living or non-living.

"Polynucleotide" generally refers to any polyribonucleotide (RNA) polydeoxribonucleotide (DNA), which may be unmodified or modifi RNA or DNA. "Polynucleotides" include, without limitation, single- a double-stranded DNA, DNA that is a mixture of single- and doub stranded regions, single- and double-stranded RNA, and RNA that mixture of single- and double-stranded regions, hybrid molecu comprising DNA and RNA that may be single-stranded or, more typica double-stranded or a mixture of single- and double-stranded regions. addition, "polynucleotide" refers to triple-stranded regions compris RNA or DNA or both RNA and DNA. The term "polynucleotide" a includes DNAs or RNAs containing one or more modified bases a DNAs or RNAs with backbones modified for stability or for other reaso "Modified" bases include, for example, tritylated bases and unusual bas such as inosine. A variety of modifications may be made to DNA a RNA; thus, "polynucleotide" embraces chemically, enzymatically metabolically modified forms of polynucleotides as typically found

nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any polypeptide comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques that are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, biotinylation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, Proteins - Structure and Molecular Properties, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, 1-12, in

Post-translational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", Meth Enzymol, 182, 626-646, 1990, and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci, 663, 48-62, 1992).

"Fragment" of a polypeptide sequence refers to a polypeptide sequence that is shorter than the reference sequence but that retains essentially the same biological function or activity as the reference polypeptide. "Fragment" of a polynucleotide sequence refers to a polynucleotide sequence that is shorter than the reference sequence of SEQ ID NO:1.

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains the essential properties thereof. A typical variant of a polynucleotide differs in nucleotide sequence from the reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from the reference polypeptide. Generally, alterations are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, insertions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. Typical conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe and Tyr. A variant of a polynucleotide or polypeptide may be naturally occurring such as an allele, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis. Also included as variants are polypeptides having one or more post-translational modifications, for instance glycosylation, phosphorylation, methylation, ADP ribosylation and the like. Embodiments include methylation of the N-terminal amino acid, phosphorylations of serines and threonines and modification of C-terminal glycines.

"Allele" refers to one of two or more alternative forms of a gene occurring at a given locus in the genome.

"Polymorphism" refers to a variation in nucleotide sequence (and encoded polypeptide sequence, if relevant) at a given position in the genome within a population.

"Single Nucleotide Polymorphism" (SNP) refers to the occurrence of nucleotide variability at a single nucleotide position in the genome, within a population. An SNP may occur within a gene or within intergenic regions of the genome. SNPs can be assayed using Allele Specific Amplification (ASA). For the process at least 3 primers are required. A common primer is used in reverse complement to the polymorphism being assayed. This common primer can be between 50 and 1500 bps from the polymorphic base. The other two (or more) primers are identical to each other except that the final 3' base wobbles to match one of the two (or more) alleles that make up the polymorphism. Two (or more) PCR reactions are then conducted on sample DNA, each using the common primer and one of the Allele Specific Primers.

"Splice Variant" as used herein refers to cDNA molecules produced from RNA molecules initially transcribed from the same genomic DNA sequence but which have undergone alternative RNA splicing. Alternative RNA splicing occurs when a primary RNA transcript undergoes splicing, generally for the removal of introns, which results in the production of more than one mRNA molecule each of that may encode different amino acid sequences. The term splice variant also refers to the proteins encoded by the above cDNA molecules.

"Identity" reflects a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, determined by comparing the sequences. In general, identity refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of the two polynucleotide or two polypeptide sequences, respectively, over the length of the sequences being compared.

"% Identity" - For sequences where there is not an exact correspondence, a "% identity" may be determined. In general, the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting "gaps" in either one

or both sequences, to enhance the degree of alignment. A % identity may be determined over the whole length of each of the sequences being compared (so-called global alignment), that is particularly suitable for sequences of the same or very similar length, or over shorter, defined lengths (so-called local alignment), that is more suitable for sequences of unequal length.

"Similarity" is a further, more sophisticated measure of the relationship between two polypeptide sequences. In general, "similarity" means a comparison between the amino acids of two polypeptide chains, on a residue by residue basis, taking into account not only exact correspondences between a between pairs of residues, one from each of the sequences being compared (as for identity) but also, where there is not an exact correspondence, whether, on an evolutionary basis, one residue is a likely substitute for the other. This likelihood has an associated "score" from which the "% similarity" of the two sequences can then be determined.

Methods for comparing the identity and similarity of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J et al, Nucleic Acids Res, 12, 387-395, 1984, available from Genetics Computer Group, Madison, Wisconsin, USA), for example the programs BESTFIT and GAP, may be used to determine the % identity between two polynucleotides and the % identity and the % similarity between two polypeptide sequences. BESTFIT uses the "local homology" algorithm of Smith and Waterman (J Mol Biol, 147,195-197, 1981, Advances in Applied Mathematics, 2, 482-489, 1981) and finds the best single region of similarity between two sequences. BESTFIT is more suited to comparing two polynucleotide or two polypeptide sequences that are dissimilar in length, the program assuming that the shorter sequence represents a portion of the longer. In comparison, GAP aligns two sequences, finding a "maximum similarity", according to the algorithm of Neddleman and Wunsch (J Mol Biol, 48, 443-453, 1970). GAP is more suited to comparing sequences that are approximately the same length and an alignment is expected over the entire length. Preferably, the parameters "Gap Weight" and "Length Weight" used in each program are 50 and 3, for polynucleotide sequences and 12 and 4 for polypeptide sequences, respectively. Preferably, % identities and

similarities are determined when the two sequences being compared are optimally aligned.

Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Altschul S F et al, J Mol Biol, 215, 403-410, 1990, Altschul S F et al, Nucleic Acids Res., 25:389-3402, 1997, available from the National Center for Biotechnology Information (NCBI), Bethesda, Maryland, USA and accessible through the home page of the NCBI at www.ncbi.nlm.nih.gov) and FASTA (Pearson W R, Methods in Enzymology, 183, 63-99, 1990; Pearson W R and Lipman D J, Proc Nat Acad Sci USA, 85, 2444-2448,1988, available as part of the Wisconsin Sequence Analysis Package).

Preferably, the BLOSUM62 amino acid substitution matrix (Henikoff S and Henikoff J G, Proc. Nat. Acad Sci. USA, 89, 10915-10919, 1992) is used in polypeptide sequence comparisons including where nucleotide sequences are first translated into amino acid sequences before comparison.

Preferably, the program BESTFIT is used to determine the % identity of a query polynucleotide or a polypeptide sequence with respect to a reference polynucleotide or a polypeptide sequence, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value, as hereinbefore described.

"Identity Index" is a measure of sequence relatedness which may be used to compare a candidate sequence (polynucleotide or polypeptide) and a reference sequence. Thus, for instance, a candidate polynucleotide sequence having, for example, an Identity Index of 0.95 compared to a reference polynucleotide sequence is identical to the reference sequence except that the candidate polynucleotide sequence may include on average up to five differences per each 100 nucleotides of the reference sequence. Such differences are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion. These differences may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between these terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or

more contiguous groups within the reference sequence. In other words, to obtain a polynucleotide sequence having an Identity Index of 0.95 compared to a reference polynucleotide sequence, an average of up to 5 in every 100 of the nucleotides of the in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

Similarly, for a polypeptide, a candidate polypeptide sequence having, for example, an Identity Index of 0.95 compared to a reference polypeptide sequence is identical to the reference sequence except that the polypeptide sequence may include an average of up to five differences per each 100 amino acids of the reference sequence. Such differences are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion. These differences may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between these terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. In other words, to obtain a polypeptide sequence having an Identity Index of 0.95 compared to a reference polypeptide sequence, an average of up to 5 in every 100 of the amino acids in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

The relationship between the number of nucleotide or amino acid differences and the Identity Index may be expressed in the following equation:

$$n_a \leq x_a - (x_a \bullet I),$$

in which:

n_a is the number of nucleotide or amino acid differences,

x_a is the total number of nucleotides or amino acids in SEQ ID NO:1 or SEQ ID NO:2, respectively,

I is the Identity Index,

• is the symbol for the multiplication operator, and

in which any non-integer product of x_a and I is rounded down to the nearest integer prior to subtracting it from x_a.

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a reference sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the two sequences as hereinbefore defined. Falling within this generic term are the terms "ortholog", and "paralog". "Ortholog" refers to a polynucleotide or polypeptide that is the functional equivalent of the polynucleotide or polypeptide in another species. "Paralog" refers to a polynucleotideor polypeptide that within the same species which is functionally similar.

"Fusion protein" refers to a protein encoded by two, unrelated, fused genes or fragments thereof. Examples have been disclosed in US 5541087, 5726044. In the case of Fc-SIP-1 ligand, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for performing the functional expression of Fc-SIP-1 ligand or fragments of the ligand, to improve pharmacokinetic properties of such a fusion protein when used for therapy and to generate a dimeric SIP-1 ligand. The Fc-SIP-1 ligand DNA construct comprises in 5' to 3' direction, a secretion cassette, i.e. a signal sequence that triggers export from a mammalian cell, DNA encoding an immunoglobulin Fc region fragment, as a fusion partner, and a DNA encoding SIP-1 ligand or fragments thereof. In some uses it would be desirable to be able to alter the intrinsic functional properties (complement binding, Fc-Receptor binding) by mutating the functional Fc sides while leaving the rest of the fusion protein untouched or delete the Fc part completely after expression.

All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set forth. Any patent application to which this application claims priority is also incorporated by reference herein in

its entirety in the manner described above for publications and references.

Discription of figures

Figure 1

Interaction of Survivin with SIP-1 in *S. cerevisiae* EGY48/pSH18-32:
S. cerevisiae EGY48/18-32 (Clontech) was transformed using plasmids encoding LexA- and Gal4-activation domain fusion proteins. Single colonies were isolated and streaked onto Media lacking Histidine, Tryptophane and Uracil (-WHU) and onto Medium lacking Histidine, Tryptophane, Leucine and Uracil containing 40 mg/l X-Gal (-WHLU-XGal). Growth of all strains on -WHU-Medium indicates presence of plasmids. Blue colour of yeast and growth on Medium lacking Leucine indicates interacion of baits and preys since expression of reportergenes depends on a Two-Hybrid selection scheme. Lanes 1 and 2 are used as unspecific controls. Lane 3 indicates the specific interaction of Survivin and SIP-1 in this Yeast-Two-Hybrid System.

Figure 2

SIP-1 Northern blot: Random primed ^{32}P-radiolabelled SIP-1 probe was used for hybridisation consisting of 590 pase pairs corresponding to position 553-1143 of the SIP-1 cds. The blot was loaded with 2µg Poly (A) RNA per lane (Blot from Invitrogen: „real human tumor panel blot 3# D3803-50).
A 2,8 kb SIP-1 mRNA is detected in mRNA-samples of kidney normal tissue, kidney tumor tissue, lung normal tissue and lung tumor tissue

Figure 3

Multiplex PCR of mRNA derived from human HeLa cells using the „Gene Specific Relative RT-PCR Kit„ (Protocol according to producer of this kit system: Ambion):
Human HeLa cells were cell phase arrested using the following treatment (G1-Phase: L-Mimosine [400µM]; S-Phase: Thymidine [2mM]; G2/M Phase: Nocodazole [40 ng/ml]). The cell cycle status was controlled by FACS analysis; control cells were treated with 1µl DMSO/ ml medium.
SIP-1 mRNA expression appears to be upregulated during S-phase and G2/M phase of the cell cycle

Figure 4

SIP-1/Survivin coimmunoprecipitation: SIP-1 and Survivin interact in human HeLa cells and after expression as in vitro translated proteins. Analysis of the SIP-1/Survivin interaction via V5-antibody mediated immunoprecipitation (IP) Assays

Further examples

Plasmid constructs

Cloning of pDBLeu – survivin: The survivin gene was cloned from cDNA of human PC-3 cells by using gene specific PCR primer pairs with in-built restriction sites for NheI and StuI, amplifying the coding sequence of survivin. The PCR product was cloned via TA-cloning into the pCRII vector (Invitrogen). The insert was isolated via NheI/StuI restriction and cloned into the Y2H vector pDBLeu (PROQUEST™ Two-Hybrid System, GibcoBRL LIFE TECHNOLOGIES) creating pDBLeu-Survivin. This results in generation of a fusion gene construct for the expression of the Gal4 DB domain fused to the full length survivin peptid. All vectors were confirmed by sequencing.

PCR primers for the cloning of the survivin gene were the following:

NheI-5´survivin primer: Primer 1

StuI-3´survivin primer: Primer 2

Selection of Survivin interacting proteins

A yeast Two Hybrid screen (Proquest, Life Technologies) using pDBLeu-Survivin as the bait construct was performed using a human HeLa cell cDNA-Y2H-library and a cDNA library derived from human fetal brain tissue (selection was performed on –Trp, -Leu, -His Minimalmedium containing 25 mM 3-Aminotriazole). SIP-1 clones were isolated from both libraries. Interaction was confirmed by ß-Galactosidase-filter assay

and Uracile, Tryptophane and Leucine lacking medium as described in the manufacturers protocoll.

One positive clone was isolated and sequenced. The corresponding interacting protein ligand was termed SIP-1 (Survivin-Interacting-Protein-1) ligand and the sequences have been disclosed in SEQ ID NO: 1 and NO: 2

Claims

1. A polypeptide selected from the group consisting of:

(a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1;

(b) a polypeptide comprising a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

c) a polypeptide having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

d) the polypeptide sequence of SEQ ID NO:2 and

(e) fragments and variants of such polypeptides in (a) to (d).

2. The polypeptide of claim 1 comprising the polypeptide sequence of SEQ ID NO:2.

3. The polypeptide of claim 1 which is the polypeptide sequence of SEQ ID NO:2.

4. A polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising a polynucleotide sequence having at least 95% identity to the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide having at least 95% identity to the polynucleotide of SEQ ID NO:1;

(c) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

(d) a polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

(e) a polynucleotide with a nucleotide sequence of at least 100 nucleotides obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof having at least 15 nucleotides;

(f) a polynucleotide which is the RNA equivalent of a polynucleotide of (a) to (e);

(g) a polynucleotide sequence complementary to said polynucleotide of any one of (a) to (f), and

(h) polynucleotides that are variants or fragments of the polynucleotides of any one of (a) to (g) or that are complementary to above mentioned polynucleotides, over the entire length thereof.

5. A polynucleotide of claim 4 selected from the group consisting of:

(a) a polynucleotide comprising the polynucleotide of SEQ ID NO:1;

(b) the polynucleotide of SEQ ID NO:1;

(c) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2; and

(d) a polynucleotide encoding the polypeptide of SEQ ID NO:2.

6. An expression system comprising a polynucleotide capable of producing a polypeptide of any one of claim 1-3 when said expression vector is present in a compatible host cell.

7. A recombinant host cell comprising the expression vector of claim 6 or a membrane thereof expressing the polypeptide of any one of claim 1-3.

8. A process for producing a polypeptide of any one of claim 1-3 comprising the step of culturing a host cell as defined in claim 7 under conditions sufficient for

the production of said polypeptide and recovering the polypeptide from the culture medium.

9. A fusion protein consisting of the Immunoglobulin Fc-region and a polypeptide any one one of claims 1-3.

10. An antibody immunospecific for the polypeptide of any one of claims 1 to 3.

11. A method for screening to identify compounds that stimulate or inhibit the function or level of the polypeptide of any one of claim 1-3 comprising a method selected from the group consisting of:

(a) measuring or, detecting, quantitatively or qualitatively, the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

(b) measuring the competition of binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof in the presence of a labeled competitior;

(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes expressing the polypeptide;

(d) mixing a candidate compound with a solution containing a polypeptide of any one of claims 1-3, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a control mixture which contains no candidate compound; or

(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide or said polypeptide in cells, using for instance, an ELISA assay, and

(f) producing said compound according to biotechnological or chemical standard techniques.

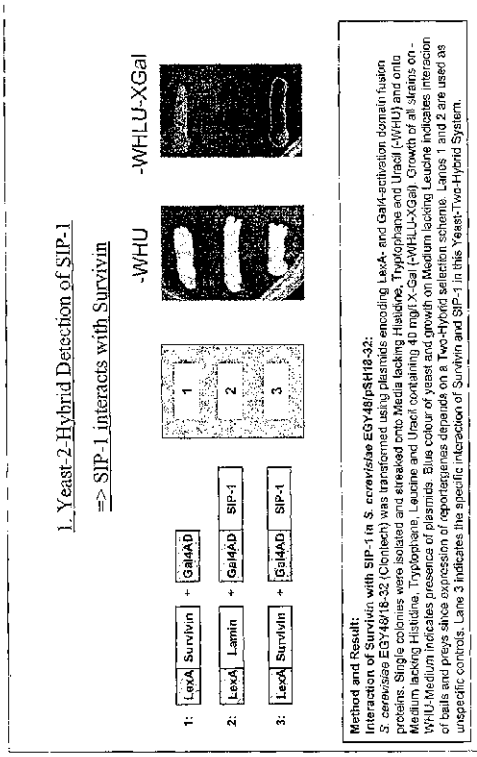


Figure1 from 4

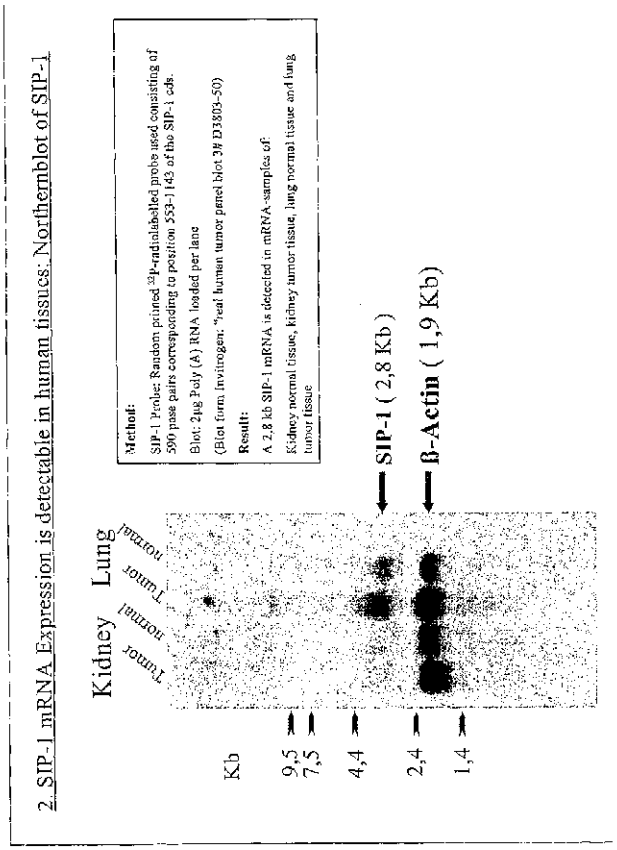


Figure 2 from 4

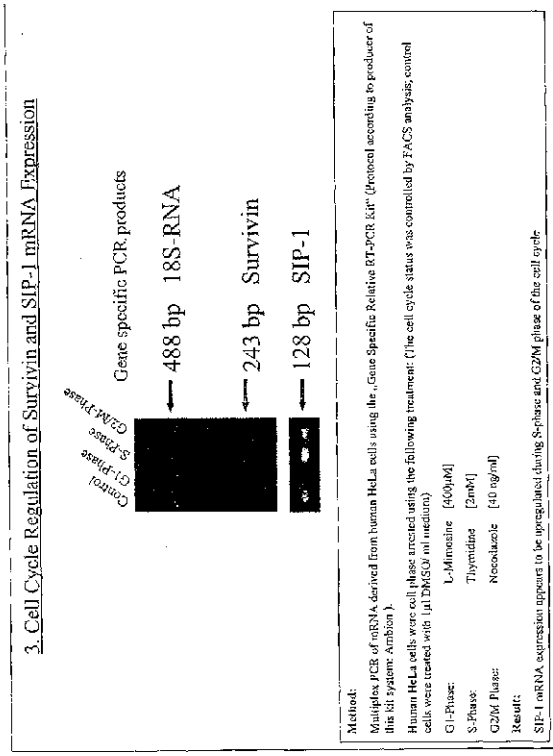


Figure 3 from 4

4: SIP-1 and Survivin interact in human HeLa cells and after expression as in vitro translated proteins.

Analysis of the SIP-1/Survivin Interaction via V5-antibody mediated Immunoprecipitation (IP) Assays

(Note: lanes 1-4: anti V5 IPs but lane 5*: anti-GFP IP)

(Note: lanes 1-4: anti V5 IPs but lane 5*: anti-GFP IP)

A: Hela Transfection SIP-1-V5-His

B: Hela Transfection SIP-1-V5-His + GFP-Survivin

C: Hela Transfection SIP-1-V5-His + GFP-vector

1: In vitro translation/transcription SIP-1-V5-His

2: In vitro translation/transcription SIP-1-V5-His + GFP-vector

3: In vitro translation/transcription SIP-1-V5-His + GFP-Survivin

4: In vitro translation/transcription without DNA

5: In vitro translation/transcription SIP-1-V5-His + GFP-Survivin

6: Anti-V5 Antibody (2µg)

7: Hela cell extract, transfection GFP-Survivin

Figure 4 from 4

SEQUENCE LISTING

<110> Merck Patent GmbH
Merck Patent GmbH

<120> Survivin Interacting Protein I

<130> SIPDHWS

<140>
<141>

<160> 4

<170> PatentIn Ver. 2.1

<210> 1
<211> 3263
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (233)..(2320)

<400> 1
tgacgcacgg aggggccggg acgcgatttg ccgcgagtga ctggagccgc aaccgtcgcc 60

gctgctgccg ccgccgcccg ggcccgctcc gcccccagcc cgncttnctc ccgccgcgcn 120

ctccgcctcc gcccgcaatt cggctacagt tcccctgtct tcccccaacgc cccggagccg 180

ccggccgcta gcgtcagcgc cagccagaat taaggaagtt cactggagta aa atg gag 238
Met Glu
1

gcc tca gta ata tta ccc att ctg aag aaa aaa cta gcc ttc ctt tca 286
Ala Ser Val Ile Leu Pro Ile Leu Lys Lys Lys Leu Ala Phe Leu Ser
5 10 15

gga gga aag gac aga cgg agt ggc ctc att ttg aca att cca tta tgc 334
Gly Gly Lys Asp Arg Arg Ser Gly Leu Ile Leu Thr Ile Pro Leu Cys
20 25 30

ctc gaa cag aca aat atg gat gag ctg agt gtc acc tta gac tac cta 382
Leu Glu Gln Thr Asn Met Asp Glu Leu Ser Val Thr Leu Asp Tyr Leu
35 40 45 50

ctt agc att cca agt gag aag tgt aag gct aga gga ttt acc gtg att 430
Leu Ser Ile Pro Ser Glu Lys Cys Lys Ala Arg Gly Phe Thr Val Ile
55 60 65

gtg gat ggc aga aaa tca cag tgg aat gtg gtg aaa aca gta gtc gta 478
Val Asp Gly Arg Lys Ser Gln Trp Asn Val Val Lys Thr Val Val Val
70 75 80

atg cta cag aat gtt gtt cca gct gag gtg tcc ctt gtt tgt gtg gta 526
Met Leu Gln Asn Val Val Pro Ala Glu Val Ser Leu Val Cys Val Val
85 90 95

```
aag cca gat gaa ttc tgg gat aag aaa gta acg cat ttt tgt ttt tgg   574
Lys Pro Asp Glu Phe Trp Asp Lys Lys Val Thr His Phe Cys Phe Trp
    100                 105                 110

aag gag aag gat aga ctt ggc ttt gag gtt att tta gtg tcc gcc aac   622
Lys Glu Lys Asp Arg Leu Gly Phe Glu Val Ile Leu Val Ser Ala Asn
115                 120                 125                 130

aaa ttg act cgt tat ata gaa cca tgc caa tta aca gaa gat ttt ggt   670
Lys Leu Thr Arg Tyr Ile Glu Pro Cys Gln Leu Thr Glu Asp Phe Gly
                135                 140                 145

ggg agt ctc acc tat gat cac atg gac tgg tta aat aag agg ctg gtt   718
Gly Ser Leu Thr Tyr Asp His Met Asp Trp Leu Asn Lys Arg Leu Val
            150                 155                 160

ttt gag aag ttt aca aag gaa tct aca tca tta tta gat gaa ctt gct   766
Phe Glu Lys Phe Thr Lys Glu Ser Thr Ser Leu Leu Asp Glu Leu Ala
        165                 170                 175

ttg att aac aat gga agt gat aaa gga aat cag caa gag aaa gaa agg   814
Leu Ile Asn Asn Gly Ser Asp Lys Gly Asn Gln Gln Glu Lys Glu Arg
    180                 185                 190

tct gtg gat tta aac ttt ctt cca tcg gtt gat cct gaa aca gtt ctt   862
Ser Val Asp Leu Asn Phe Leu Pro Ser Val Asp Pro Glu Thr Val Leu
195                 200                 205                 210

cag aca ggg cat gaa ttg ttg tcc gaa tta cag cag cgt cga ttt aat   910
Gln Thr Gly His Glu Leu Leu Ser Glu Leu Gln Gln Arg Arg Phe Asn
                215                 220                 225

ggc tca gac gga ggg gtt tca tgg tct cct atg gat gat gaa ctt ctt   958
Gly Ser Asp Gly Gly Val Ser Trp Ser Pro Met Asp Asp Glu Leu Leu
            230                 235                 240

gca cag cca cag gtt atg aaa tta tta gat tca ctc cga gag caa tat  1006
Ala Gln Pro Gln Val Met Lys Leu Leu Asp Ser Leu Arg Glu Gln Tyr
        245                 250                 255

acc cgc tac cag gaa gtt tgt agg caa cgt agc aag cgc aca cag tta  1054
Thr Arg Tyr Gln Glu Val Cys Arg Gln Arg Ser Lys Arg Thr Gln Leu
    260                 265                 270

gaa gag att caa cag aag gta atg cag gtg gtg aac tgg cta gaa ggg  1102
Glu Glu Ile Gln Gln Lys Val Met Gln Val Val Asn Trp Leu Glu Gly
275                 280                 285                 290

cct gga tca gaa caa cta aga gcc cag tgg ggc att gga gac tcc att  1150
Pro Gly Ser Glu Gln Leu Arg Ala Gln Trp Gly Ile Gly Asp Ser Ile
                295                 300                 305

agg gcc tcc cag gcc cta cag cag aaa cac gaa gag att gag agc cag  1198
Arg Ala Ser Gln Ala Leu Gln Gln Lys His Glu Glu Ile Glu Ser Gln
            310                 315                 320

cac agt gaa tgg ttt gca gtg tat gtg gaa ctt aat cag caa att gca  1246
His Ser Glu Trp Phe Ala Val Tyr Val Glu Leu Asn Gln Gln Ile Ala
        325                 330                 335
```

```
gca ctc ttg aat gct ggc gat gag gaa gat ctt gtg gaa cta aag tca  1294
Ala Leu Leu Asn Ala Gly Asp Glu Glu Asp Leu Val Glu Leu Lys Ser
    340                 345                 350

ctg cag caa caa ctt agt gat gtt tgt tat cga cag gcc agt cag ctg  1342
Leu Gln Gln Gln Leu Ser Asp Val Cys Tyr Arg Gln Ala Ser Gln Leu
355                 360                 365                 370

gaa ttt agg caa aat ctc tta caa gca gct ctt gaa ttt cat ggt gtt  1390
Glu Phe Arg Gln Asn Leu Leu Gln Ala Ala Leu Glu Phe His Gly Val
                375                 380                 385

gcc caa gat ttg tct cag cag ttg gat ggc tta tta ggg atg ttg tgc  1438
Ala Gln Asp Leu Ser Gln Gln Leu Asp Gly Leu Leu Gly Met Leu Cys
            390                 395                 400

gta gat gta gca cca gct gat gga gca tcg att cag caa act tta aaa  1486
Val Asp Val Ala Pro Ala Asp Gly Ala Ser Ile Gln Gln Thr Leu Lys
        405                 410                 415

ctg ctt gaa gag aag ctg aaa agt gtt gat gtg gga ttg caa ggt ttg  1534
Leu Leu Glu Glu Lys Leu Lys Ser Val Asp Val Gly Leu Gln Gly Leu
    420                 425                 430

cgt gaa aaa ggt caa ggt ctc ctg gat cag atc tcc aat cag gca tcc  1582
Arg Glu Lys Gly Gln Gly Leu Leu Asp Gln Ile Ser Asn Gln Ala Ser
435                 440                 445                 450

tgg gcc tat gga aag gat gta acc att gaa aat aaa gaa aat gtg gac  1630
Trp Ala Tyr Gly Lys Asp Val Thr Ile Glu Asn Lys Glu Asn Val Asp
                455                 460                 465

cac ata caa gga gtg atg gaa gat atg cag ctt aga aaa caa aga tgt  1678
His Ile Gln Gly Val Met Glu Asp Met Gln Leu Arg Lys Gln Arg Cys
            470                 475                 480

gaa gac atg gta gat gtg cga agg tta aag atg ctt cag atg gtg cag  1726
Glu Asp Met Val Asp Val Arg Arg Leu Lys Met Leu Gln Met Val Gln
        485                 490                 495

ttg ttt aaa tgt gaa gaa gat gct gcc cag gca gta gaa tgg cta agt  1774
Leu Phe Lys Cys Glu Glu Asp Ala Ala Gln Ala Val Glu Trp Leu Ser
    500                 505                 510

gaa ctt ctg gat gct ctg ctt aag act cac atc aga ttg ggc gat gat  1822
Glu Leu Leu Asp Ala Leu Leu Lys Thr His Ile Arg Leu Gly Asp Asp
515                 520                 525                 530

gct caa gaa acg aaa gtt ttg ctg gaa aag cat aga aaa ttt gtt gat  1870
Ala Gln Glu Thr Lys Val Leu Leu Glu Lys His Arg Lys Phe Val Asp
                535                 540                 545

gtt gca cag agc act tat gac tat ggc agg cag ttg cta cag gcc aca  1918
Val Ala Gln Ser Thr Tyr Asp Tyr Gly Arg Gln Leu Leu Gln Ala Thr
            550                 555                 560

gtt gtg tta tgc caa tct ttg cgc tgc act tct cgg tca tct ggg gat  1966
Val Val Leu Cys Gln Ser Leu Arg Cys Thr Ser Arg Ser Ser Gly Asp
        565                 570                 575
```

```
aca ctt cct cga ctg aac aga gta tgg aaa caa ttt aca ata gca tct  2014
Thr Leu Pro Arg Leu Asn Arg Val Trp Lys Gln Phe Thr Ile Ala Ser
    580                 585                 590

gaa gag aga gta cat aga ttg gaa atg gct att gca ttt cac tca aat  2062
Glu Glu Arg Val His Arg Leu Glu Met Ala Ile Ala Phe His Ser Asn
595                 600                 605                 610

gct gaa aag att ttg cag gac tgt cca gaa gag cct gaa gct att aat  2110
Ala Glu Lys Ile Leu Gln Asp Cys Pro Glu Glu Pro Glu Ala Ile Asn
                615                 620                 625

gat gag gag caa ttt gat gaa att gaa gca gtt ggg aaa tca ctt ttg  2158
Asp Glu Glu Gln Phe Asp Glu Ile Glu Ala Val Gly Lys Ser Leu Leu
            630                 635                 640

gat aga tta act gtt cca gta gtt tat cct gat gga acc gaa caa tat  2206
Asp Arg Leu Thr Val Pro Val Val Tyr Pro Asp Gly Thr Glu Gln Tyr
        645                 650                 655

ttt ggg agt cca agt gac atg gct tct act gca gaa aac atc aga gac  2254
Phe Gly Ser Pro Ser Asp Met Ala Ser Thr Ala Glu Asn Ile Arg Asp
    660                 665                 670

agg atg aaa cta gtt aat ctc aaa agg cag cag ctg aga cat cct gaa  2302
Arg Met Lys Leu Val Asn Leu Lys Arg Gln Gln Leu Arg His Pro Glu
675                 680                 685                 690

atg gtg acc aca gag agc taatagctac cagctaccta cagatttgca         2350
Met Val Thr Thr Glu Ser
                695

gttcataatc ccgcatgttg tcaacatact acagcattag ccaccacacc ttaagatgca 2410

tttcacagcc aaaataagtc tcatttcttt tcatgacaca tttctcttta catgttaaca 2470

cottgctact accaaggcat aattacttaa catgcttcga ggctgtagat tccaagtatc 2530

ttaaaagaag gaactataaa cattgcactg aaaacttgct ttaaagcttt acctgacctg 2590

tcagtttgta gacaaacaac tgataatsag ctttgaatgg tgctaataag agtaggaatt 2650

ctctctatta aaaaaaaaaa aaaaagttgc cctttcttcca caggtgattt agtaaattta 2710

gacagtagtt aaactcttgt tagtagacag tggtgtcctc aaaatttcac tttgtaattc 2770

ttcagaattg attatttttа ttgtgtcaat acagagaaag cctttcagat ctttgatata 2830

tcatagtcat taaaagacct tttcctattt gtattgatca tgtattaaaa gttgtttgtg 2890

cttaataaaa gacttcttta aacatcttat ttaatttagt agttacatcc tatttccaaa 2950

catgagtgcc ttatttaaaa gggcatttctt aggactgtga ggatggttta atatttgttt 3010

ttcatggtgg ttgcatgtat tttagacago aaatacatat gtaagcatgt gtntataata 3070

aataancatg ttttatcatg aaaaaattat tgtgaacaat ttaaatcttt aagaacttat 3130

taataatgga atactattct aatttttcctc tttttcacтt gaaaatttcc caaattatta 3190

actttcccaa aanatttgtn tttaggggaa gaaggtnaag aaaagggttc ttattcctcn 3250
```

tgttancctt aat 3263

<210> 2
<211> 696
<212> PRT
<213> Homo sapiens

<400> 2
Met Glu Ala Ser Val Ile Leu Pro Ile Leu Lys Lys Lys Leu Ala Phe
1 5 10 15

Leu Ser Gly Gly Lys Asp Arg Arg Ser Gly Leu Ile Leu Thr Ile Pro
20 25 30

Leu Cys Leu Glu Gln Thr Asn Met Asp Glu Leu Ser Val Thr Leu Asp
35 40 45

Tyr Leu Leu Ser Ile Pro Ser Glu Lys Cys Lys Ala Arg Gly Phe Thr
50 55 60

Val Ile Val Asp Gly Arg Lys Ser Gln Trp Asn Val Val Lys Thr Val
65 70 75 80

Val Val Met Leu Gln Asn Val Val Pro Ala Glu Val Ser Leu Val Cys
85 90 95

Val Val Lys Pro Asp Glu Phe Trp Asp Lys Lys Val Thr His Phe Cys
100 105 110

Phe Trp Lys Glu Lys Asp Arg Leu Gly Phe Glu Val Ile Leu Val Ser
115 120 125

Ala Asn Lys Leu Thr Arg Tyr Ile Glu Pro Cys Gln Leu Thr Glu Asp
130 135 140

Phe Gly Gly Ser Leu Thr Tyr Asp His Met Asp Trp Leu Asn Lys Arg
145 150 155 160

Leu Val Phe Glu Lys Phe Thr Lys Glu Ser Thr Ser Leu Leu Asp Glu
165 170 175

Leu Ala Leu Ile Asn Asn Gly Ser Asp Lys Gly Asn Gln Gln Glu Lys
180 185 190

Glu Arg Ser Val Asp Leu Asn Phe Leu Pro Ser Val Asp Pro Glu Thr
195 200 205

Val Leu Gln Thr Gly His Glu Leu Leu Ser Glu Leu Gln Gln Arg Arg
210 215 220

Phe Asn Gly Ser Asp Gly Gly Val Ser Trp Ser Pro Met Asp Asp Glu
225 230 235 240

Leu Leu Ala Gln Pro Gln Val Met Lys Leu Leu Asp Ser Leu Arg Glu
245 250 255

Gln Tyr Thr Arg Tyr Gln Glu Val Cys Arg Gln Arg Ser Lys Arg Thr
260 265 270

Gln Leu Glu Glu Ile Gln Gln Lys Val Met Gln Val Val Asn Trp Leu
275 280 285

Gln Gly Pro Gly Ser Glu Gln Leu Arg Ala Gln Trp Gly Ile Gly Asp
290 295 300

Ser Ile Arg Ala Ser Gln Ala Leu Gln Gln Lys His Glu Glu Ile Glu
305 310 315 320

Ser Gln His Ser Glu Trp Phe Ala Val Tyr Val Glu Leu Asn Gln Gln
325 330 335

Ile Ala Ala Leu Leu Asn Ala Gly Asp Glu Glu Asp Leu Val Glu Leu
340 345 350

Lys Ser Leu Gln Gln Gln Leu Ser Asp Val Cys Tyr Arg Gln Ala Ser
355 360 365

Gln Leu Glu Phe Arg Gln Asn Leu Leu Gln Ala Ala Leu Glu Phe His
370 375 380

Gly Val Ala Gln Asp Leu Ser Gln Gln Leu Asp Gly Leu Leu Gly Met
385 390 395 400

Leu Cys Val Asp Val Ala Pro Ala Asp Gly Ala Ser Ile Gln Gln Thr
405 410 415

Leu Lys Leu Leu Glu Glu Lys Leu Lys Ser Val Asp Val Gly Leu Gln
420 425 430

Gly Leu Arg Glu Lys Gly Gln Gly Leu Leu Asp Gln Ile Ser Asn Gln
435 440 445

Ala Ser Trp Ala Tyr Gly Lys Asp Val Thr Ile Glu Asn Lys Glu Asn
450 455 460

Val Asp His Ile Gln Gly Val Met Glu Asp Met Gln Leu Arg Lys Gln
465 470 475 480

Arg Cys Glu Asp Met Val Asp Val Arg Arg Leu Lys Met Leu Gln Met
485 490 495

Val Gln Leu Phe Lys Cys Glu Glu Asp Ala Ala Gln Ala Val Glu Trp
500 505 510

Leu Ser Glu Leu Leu Asp Ala Leu Leu Lys Thr His Ile Arg Leu Gly
515 520 525

Asp Asp Ala Gln Glu Thr Lys Val Leu Leu Glu Lys His Arg Lys Phe
530 535 540

Val Asp Val Ala Gln Ser Thr Tyr Asp Tyr Gly Arg Gln Leu Leu Gln
545 550 555 560

Ala Thr Val Val Leu Cys Gln Ser Leu Arg Cys Thr Ser Arg Ser Ser
565 570 575

Gly Asp Thr Leu Pro Arg Leu Asn Arg Val Trp Lys Gln Phe Thr Ile
580 585 590

Ala Ser Glu Glu Arg Val His Arg Leu Glu Met Ala Ile Ala Phe His
595 600 605

Ser Asn Ala Glu Lys Ile Leu Gln Asp Cys Pro Glu Glu Pro Glu Ala
610 615 620

Ile Asn Asp Glu Glu Gln Phe Asp Glu Ile Glu Ala Val Gly Lys Ser
625 630 635 640

Leu Leu Asp Arg Leu Thr Val Pro Val Val Tyr Pro Asp Gly Thr Glu
645 650 655

Gln Tyr Phe Gly Ser Pro Ser Asp Met Ala Ser Thr Ala Glu Asn Ile
660 665 670

Arg Asp Arg Met Lys Leu Val Asn Leu Lys Arg Gln Gln Leu Arg His
675 680 685

Pro Glu Met Val Thr Thr Glu Ser
690 695

<210> 3
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer 1
(NheI-5' survivin primer)

<400> 3
ctagctagca tgggtgcccc gacgttg 27

<210> 4
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer 2
(StuI-3' survivin primer)

<400> 4
taggccttca atccatggca gccag 25

【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)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 October 2001 (04.10.2001)

PCT

(10) International Publication Number
WO 01/073014 A1

(51) International Patent Classification⁷: C12N 15/12, C07K 14/47, C12N 15/62, C07K 16/18

(21) International Application Number: PCT/EP01/03203

(22) International Filing Date: 21 March 2001 (21.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00106408.8 24 March 2000 (24.03.2000) EP

(71) Applicant (*for all designated States except US*): MERCK PATENT GMBH [DE/DE]; Frankfurter Strasse 250, 64293 Darmstadt (DE).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): HENTSCH, Bernd [DE/DE]; Kampstrasse 48, 64289 Darmstadt (DE); DÜCKER, Klaus [DE/DE]; Dieselstr. 8, 64293 Darmstadt (DE); HOCK, Björn [DE/DE]; Nordstrasse 2a, 63477 Maintal (DE).

(74) Common Representative: MERCK PATENT GMBH; 64271 Darmstadt (DE).

(81) Designated States (*national*): CA, JP, US.

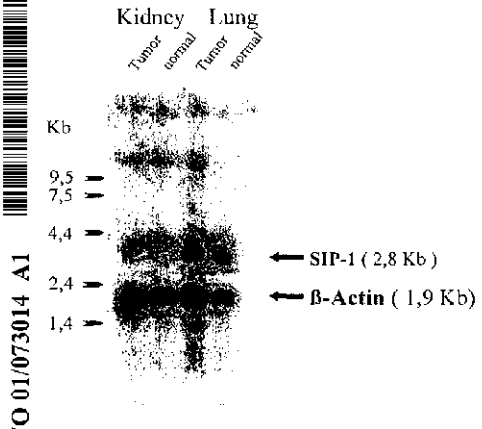
(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*

(48) Date of publication of this corrected version:
10 April 2003

(Continued on next page)

(54) Title: HUMAN SURVIVIN INTERACTING PROTEIN 1 (SIP-1)



(57) Abstract: Survivin Interacting Protein 1 and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing SIP-1 polypeptides and polynucleotides in diagnostic assays.

WO 01/073014 A1

WO 01/073014 A1

(15) Information about Correction:
see PCT Gazette No. 15/2003 of 10 April 2003, Section II

HUMAN SURVIVIN INTERACTING PROTEIN 1 (SIP-1)

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides sometimes hereinafter referred to as Survivin Interacting Protein 1 or „SIP-1", , to their use in diagnosis and in identifying compounds that may be agonists, antagonists that are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces "functional genomics", that is, high throughput genome- or gene-based biology. This approach as a means to identify genes and gene products as therapeutic targets is rapidly superseding earlier approaches based on "positional cloning". A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery. Recent advances in this area are mainly driven by applying advanced screening systems with focus to the function of newly identified proteins. Interesting application for this type of screens are whithin the field of tumor apoptosis.

Apoptosis of tumor cells can be blocked e.g. by mutations in crucial genes like those frequently found in the well known p53 tumor suppressor protein. Another mechanism that blocks apoptosis is caused by the protein family of „inhibitors of apoptosis" (IAPs), reviewed by Deveraux

and Reed (Genes & Development 1999). One member of this family is encoded by the *Survivin* gene.

The strongest evidence for an IAP involvement in cancer is seen for *Survivin* (see review of Altieri and Marchisio, Laboratory Investigation 1999). Although not observed in adult differentiated tissue, *Survivin* becomes prominently expressed in transformed cell lines and in all of the most common human cancers of lung, colon, pancreas, prostate and breast, in vivo. *Survivin* is also found in approximately 50% of high-grade non-Hodgkin's lymphomas (centroblastic, immunoblastic), but not in low-grade lymphomas (lymphocytic). *Survivin* inhibits caspase activity and apoptosis induced by Fas (CD95), Bax, Caspases, and anti-cancer drugs.

In addition, *Survivin* is upregulated 40-fold at G2/M phase of the cell cycle and binds to mitotic spindles (micro-tubules), although its role at the spindle is still unclear. There might be a connected control of apoptosis and mitotic spindle checkpoint by *Survivin*. Disruption of *Survivin*-microtubule interactions results in loss of *Survivin*'s anti-apoptosis function and increased caspase-3 activity. *Survivin* may counteract a default induction of apoptosis in G2/M phase. The overexpression of *Survivin* in cancer may overcome this apoptotic checkpoint and favour aberrant progression of transformed cells through mitosis (Fengzhi et al., Nature 1998). In this repect it is important to note that it was recently shown that *Survivin* initiates procaspase 3/p21 complex formation as a result of interaction with Cdk4 to resist Fas-mediated cell death (Suzuki et al., Oncogene 2000). SIP-1 may therefore be implicated in the known tumor-cell-escape from the immune system. The other biochemical mechanisms, besides its inhibitory binding to Caspases, by which *Survivin* might mediate its anti-apoptotic activity are currently unclear.

Therefore in this invention we have screened for potential ligands that interact with *Survivin*.

Summary of the invention

The present invention relates to Survivin Interacting Protein 1 (*GENE NAME), in particular *GENE NAME polypeptides and *GENE NAME

polynucleotides, recombinant materials and methods for their production. Such polypeptides and polynucleotides are of interest in relation to methods of treatment of certain diseases, because they are natural ligands for Survivin. As pointed out this molecule has a prominent role in anti-apoptotic activity.

Survivin becomes prominently expressed in most common human cancers of lung, liver, colon, stomach, skin, pancreas, prostate, ovary and breast, in vivo. *Survivin* is also found in approximately 50% of high-grade non-Hodgkin's lymphomas thus the interaction with the newly described molecule of this invention SIP-1 offers an attractive option to interfere with this diseases therefore they are hereinafter referred to as "diseases of the invention". In a further aspect, the invention relates to methods for identifying agonists and antagonists (*e.g.*, inhibitors) using the materials provided by the invention, and treating conditions associated with *GENE NAME activity with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate *GENE NAME activity or levels.

Description of the Invention

In a first aspect, the present invention relates to *GENE NAME polypeptides. Such polypeptides include:

(a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1;

(b) an polypeptide comprising a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(c) a polypeptide comprising the polypeptide sequence of SEQ ID NO:2;

(d) a polypeptide having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(e) the polypeptide sequence of SEQ ID NO:2; and

(f) a polypeptide having or comprising a polypeptide sequence that has an identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2;

(g) fragments and variants of such polypeptides in (a) to (f).

Polypeptides of the present invention are believed to represent ligands for *Survivin* and thus may counteract a default induction of apoptosis. Furthermore Survivin and its ligand SIP-1 may control apoptotic checkpoint and SIP-1 may be a mediator of Survivin's anti-apoptotic activity.

SIP-1 polypeptides may be involved in the establishment of a survival signal of tumor cells which may result in resistance of these cells against apoptotic signals, e.g. chemotherapeutic agents or physiological death signals as such of immune cells.

The biological properties of the *GENE NAME are hereinafter referred to as "biological activity of *GENE NAME" or "*GENE NAME activity". Preferably, a polypeptide of the present invention exhibits at least one biological activity of *GENE NAME.

Polypeptides of the present invention also includes variants of the aforementioned polypeptides, including all allelic forms and splice variants. Such polypeptides vary from the reference polypeptide by insertions, deletions, and substitutions that may be conservative or non-conservative, or any combination thereof. Particularly preferred variants are those in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acids are inserted, substituted, or deleted, in any combination.

Preferred fragments of polypeptides of the present invention include a polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO: 2, or a polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids truncated or deleted from the amino acid sequence of SEQ ID NO: 2. Preferred fragments are biologically active fragments that mediate the biological activity of *GENE NAME, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also preferred are those fragments that are antigenic or immunogenic in an animal, especially in a human.

Fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis;

therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention. The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence that contains secretory or leader sequences, pro-sequences, sequences that aid in purification, for instance multiple histidine residues, or an additional sequence for stability during recombinant production.

Polypeptides of the present invention can be prepared in any suitable manner, for instance by isolation form naturally occurring sources, from genetically engineered host cells comprising expression systems (*vide infra*) or by chemical synthesis, using for instance automated peptide synthesizers, or a combination of such methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to *GENE NAME polynucleotides. Such polynucleotides include:

(a) a polynucleotide comprising a polynucleotide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide comprising the polynucleotide of SEQ ID NO:1;

(c) a polynucleotide having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide of SEQ ID NO:1;

(d) the polynucleotide of SEQ ID NO:1;

(e) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(f) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2;

(g) a polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;

(h) a polynucleotide encoding the polypeptide of SEQ ID NO:2;

(i) a polynucleotide having or comprising a polynucleotide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polynucleotide sequence of SEQ ID NO:1;

(j) a polynucleotide having or comprising a polynucleotide sequence encoding a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2; and

polynucleotides that are fragments and variants of the above mentioned polynucleotides or that are complementary to above mentioned polynucleotides, over the entire length thereof.

Preferred fragments of polynucleotides of the present invention include a polynucleotide comprising an nucleotide sequence having at least 15, 30, 50 or 100 contiguous nucleotides from the sequence of SEQ ID NO: 1, or a polynucleotide comprising an sequence having at least 30, 50 or 100 contiguous nucleotides truncated or deleted from the sequence of SEQ ID NO: 1.

Preferred variants of polynucleotides of the present invention include splice variants, allelic variants, and polymorphisms, including polynucleotides having one or more single nucleotide polymorphisms (SNPs).

Polynucleotides of the present invention also include polynucleotides encoding polypeptide variants that comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acid residues are substituted, deleted or added, in any combination.

In a further aspect, the present invention provides polynucleotides that are RNA transcripts of the DNA sequences of the present invention. Accordingly, there is provided an RNA polynucleotide that:

(a) comprises an RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2;

(b) is the RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2;

(c) comprises an RNA transcript of the DNA sequence of SEQ ID NO:1; or

(d) is the RNA transcript of the DNA sequence of SEQ ID NO:1;

and RNA polynucleotides that are complementary thereto.

The polynucleotide sequence of SEQ ID NO:1 is a cDNA sequence that encodes the polypeptide of SEQ ID NO:2. The polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence of SEQ ID NO:1 or it may be a sequence other than SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is related to other proteins of the family of guanine exchange factors (GEFs), having homology and/or structural similarity with Homo sapiens Duo mRNA, which encodes a spectrin-like domain, coiled coil, guanosine exchange factor domain and pleckstrin homology domain [Colomer, V., Engelender, S., Sharp, A.H., Duan, K., Cooper, J.K., Lanahan, A., Lyford, G., Worley, P. and Ross, C.A.: Huntingtin-associated protein 1 binds to a Trio-like polypeptide, with a rac1 guanine nucleotide exchange factor domain. Hum. Mol. Genet. 6 (9), 1519-1525 (1997)] and to Homo sapiens Trio mRNA which displays similarities to protein kinase, rac guanine exchange factor, rho guanine exchange factor and spectrin-like repeats [Debant, A., Serra-Pages, C., Seipel, K., O'Brien, S., Tang, M., Park, S.H. and Streuli, M.: The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. Proc. Natl. Acad. Sci. U.S.A. 93 (11), 5466-5471, (1996)].

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one *GENE NAME activity.

Polynucleotides of the present invention may be obtained using standard cloning and screening techniques from a cDNA library derived from mRNA of human HeLa cells (ATCC: CCL-2) and in a cDNA library derived from human fetal brain tissue (see for instance, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself, or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Polynucleotides that are identical, or have sufficient identity to a polynucleotide sequence of SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification reaction (for instance, PCR). Such probes and primers may be used to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1, typically at least 95% identity. Preferred probes and primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50, if not at least 100 nucleotides. Particularly preferred probes will have

between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process comprising the steps of screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof, preferably of at least 15 nucleotides; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides, preferably with a nucleotide sequence of at least 100, obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof, preferably of at least 15 nucleotides.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide does not extend all the way through to the 5' terminus. This is a consequence of reverse transcriptase, an enzyme with inherently low "processivity" (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during first strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., Proc Nat Acad Sci USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon (trade mark) technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon (trade mark) technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to

amplify the "missing" 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adapter specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analyzed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems comprising a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Polynucleotides may be introduced into host cells by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook *et al.*(*ibid*). Preferred methods of introducing polynucleotides into host cells include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, micro-injection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci*, *Staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as

CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector that is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate polynucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, (*ibid*). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate

active conformation when the polypeptide is denatured during intracellular synthesis, isolation and/or purification.

Polynucleotides of the present invention may be used as diagnostic reagents, through detecting mutations in the associated gene. Detection of a mutated form of the gene characterized by the polynucleotide of SEQ ID NO:1 in the cDNA or genomic sequence and 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 well known in the art.

Nucleic acids 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 it may be amplified enzymatically by using PCR, preferably RT-PCR, or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled *GENE NAME nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence difference 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 (see, for instance, 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).

An array of oligonucleotides probes comprising *GENE NAME polynucleotide sequence or fragments thereof can be constructed to conduct efficient screening of *e.g.*, genetic mutations. Such arrays are preferably high density arrays or grids. 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, 274, 610-613 (1996) and other references cited therein.

Detection of abnormally decreased or increased levels of polypeptide or mRNA expression may also be used for diagnosing or determining susceptibility of a subject to a disease of the invention. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the 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 protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radio-immunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagnostic kit comprising:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment or an RNA transcript thereof;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a disease, particularly diseases of the invention, amongst others.

The polynuclectide sequences of the present invention are valuable for chromosome localisation studies. The sequence 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 first step in correlating those sequences with 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 relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (co-inheritance of physically adjacent genes). Precise human chromosomal localisations for a genomic sequence (gene fragment etc.) can be determined using Radiation Hybrid (RH) Mapping (Walter, M. Spillett, D., Thomas, P., Weissenbach, J., and Goodfellow, P., (1994) A method for constructing radiation hybrid maps of whole genomes, Nature Genetics 7, 22-28). A number of RH panels are available from Research Genetics (Huntsville, AL, USA) e.g. the GeneBridge4 RH panel (Hum Mol Genet 1996 Mar;5(3):339-46 A radiation hybrid map of the human genome. Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN). To determine the chromosomal location of a gene using this panel, 93 PCRs are performed using primers designed from the gene of interest on RH DNAs. Each of these DNAs contains random human genomic fragments maintained in a hamster background (human / hamster hybrid cell lines). These PCRs result in 93 scores indicating the presence or absence of the PCR product of the gene of interest. These scores are compared with scores created using PCR products from genomic sequences of known location. This comparison is conducted at http://www.genome.wi.mit.edu/. The gene of the present invention maps to human chromosome Chr. 2 (D2S336-D2S2257).

The polynucleotide sequences of the present invention are also valuable tools for tissue expression studies. Such studies allow the determination of expression patterns of polynucleotides of the present invention which may give an indication as to the expression patterns of the encoded polypeptides in tissues, by detecting the mRNAs that encode them. The techniques used are well known in the art and include in situ hydridization techniques to clones arrayed on a grid, such as cDNA microarray hybridization (Schena *et al*, Science, 270, 467-470, 1995 and Shalon *et al*, Genome Res, 6, 639-645, 1996) and nucleotide amplification techniques such as PCR. A preferred method uses the TAQMAN (Trade mark) technology available from Perkin Elmer. Results from these studies can provide an indication of the normal function of the polypeptide in the

organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by an alternative form of the same gene (for example, one having an alteration in polypeptide coding potential or a regulatory mutation) can provide valuable insights into the role of the polypeptides of the present invention, or that of inappropriate expression thereof in disease. Such inappropriate expression may be of a temporal, spatial or simply quantitative nature.

A further aspect of the present invention relates to antibodies. The polypeptides of the invention or their fragments, or cells expressing them, can be used as immunogens to produce antibodies that are immunospecific for polypeptides of the present invention. 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.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, Immunology Today (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography. Antibodies against polypeptides of the present invention may also be employed to treat diseases of the invention, amongst others.

Polypeptides and polynucleotides of the present invention may also be used as vaccines. Accordingly, in a further aspect, the present invention

relates to a method for inducing an immunological response in a mammal that comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said animal from disease, whether that disease is already established within the individual or not. An immunological response in a mammal may also be induced by a method comprises delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases of the invention. One way of administering the vector is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid. For use a vaccine, a polypeptide or a nucleic acid vector will be normally provided as a vaccine formulation (composition). The formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance. subcutaneous, intra-muscular, intravenous, or intra-dermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions that may contain anti-oxidants, buffers, bacteriostats and solutes that render the formulation instonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions that may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose 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 vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention have one or more biological functions that are of relevance in one or more disease states, in particular the diseases of the invention hereinbefore mentioned. It is therefore useful to identify compounds that stimulate or inhibit the function or level of the

polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those that stimulate or inhibit the function or level of the polypeptide. Such methods identify agonists or antagonists that may be employed for therapeutic and prophylactic purposes for such diseases of the invention as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, collections of chemical compounds, and natural product mixtures. Such agonists or antagonists so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; a structural or functional mimetic thereof (see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)) or a small molecule. Such small molecules preferably have a molecular weight below 2,000 daltons, more preferably between 300 and 1,000 daltons, and most preferably between 400 and 700 daltons. It is preferred that these small molecules are organic molecules.

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof, by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve measuring or detecting (qualitatively or quantitatively) the competitive binding of a candidate compound to the polypeptide against a labeled competitor (*e.g.* agonist or antagonist). Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring a *GENE NAME activity in the mixture, and comparing the *GENE NAME activity of the mixture to a control mixture which contains no candidate compound.

Polypeptides of the present invention may be employed in conventional low capacity screening methods and also in high-throughput screening (HTS) formats. Such HTS formats include not only the well-established

use of 96- and, more recently, 384-well micotiter plates but also emerging methods such as the nanowell method described by Schullek et al, Anal Biochem., 246, 20-29, (1997).

Fusion proteins, such as those made from Fc portion and *GENE NAME polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

Screening techniques

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents that may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

A polypeptide of the present invention may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide that compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of antagonists of polypeptides of the present invention include antibodies or, in some cases, oligonucleotides or proteins that are closely related to the ligands, substrates, receptors, enzymes, etc., as the case

may be, of the polypeptide, *e.g.*, a fragment of the ligands, substrates, receptors, enzymes, etc.; or a small molecule that bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Screening methods may also involve the use of transgenic technology and *GENE NAME gene. The art of constructing transgenic animals is well established. For example, the *GENE NAME gene may be introduced through microinjection into the male pronucleus of fertilized oocytes, retroviral transfer into pre- or post-implantation embryos, or injection of genetically modified, such as by electroporation, embryonic stem cells into host blastocysts. Particularly useful transgenic animals are so-called "knock-in" animals in which an animal gene is replaced by the human equivalent within the genome of that animal. Knock-in transgenic animals are useful in the drug discovery process, for target validation, where the compound is specific for the human target. Other useful transgenic animals are so-called "knock-out" animals in which the expression of the animal ortholog of a polypeptide of the present invention and encoded by an endogenous DNA sequence in a cell is partially or completely annulled. The gene knock-out may be targeted to specific cells or tissues, may occur only in certain cells or tissues as a consequence of the limitations of the technology, or may occur in all, or substantially all, cells in the animal. Transgenic animal technology also offers a whole animal expression-cloning system in which introduced genes are expressed to give large amounts of polypeptides of the present invention

Screening kits for use in the above described methods form a further aspect of the present invention. Such screening kits comprise:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present invention;

(c) a cell membrane expressing a polypeptide of the present invention; or

(d) an antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

Glossary

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an

Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Moreover, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living.

"Polynucleotide" generally refers to any polyribonucleotide (RNA) or polydeoxribonucleotide (DNA), which: may be unmodified or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in

nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any polypeptide comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques that are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, biotinylation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, Proteins - Structure and Molecular Properties, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, 1-12, in

Post-translational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", Meth Enzymol, 182, 626-646, 1990, and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci, 663, 48-62, 1992).

"Fragment" of a polypeptide sequence refers to a polypeptide sequence that is shorter than the reference sequence but that retains essentially the same biological function or activity as the reference polypeptide. "Fragment" of a polynucleotide sequence refers to a polynucleotide sequence that is shorter than the reference sequence of SEQ ID NO:1.

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains the essential properties thereof. A typical variant of a polynucleotide differs in nucleotide sequence from the reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from the reference polypeptide. Generally, alterations are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, insertions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. Typical conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe and Tyr. A variant of a polynucleotide or polypeptide may be naturally occurring such as an allele, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis. Also included as variants are polypeptides having one or more post-translational modifications, for instance glycosylation, phosphorylation, methylation, ADP ribosylation and the like. Embodiments include methylation of the N-terminal amino acid, phosphorylations of serines and threonines and modification of C-terminal glycines.

"Allele" refers to one of two or more alternative forms of a gene occurring at a given locus in the genome.

"Polymorphism" refers to a variation in nucleotide sequence (and encoded polypeptide sequence, if relevant) at a given position in the genome within a population.

"Single Nucleotide Polymorphism" (SNP) refers to the occurrence of nucleotide variability at a single nucleotide position in the genome, within a population. An SNP may occur within a gene or within intergenic regions of the genome. SNPs can be assayed using Allele Specific Amplification (ASA). For the process at least 3 primers are required. A common primer is used in reverse complement to the polymorphism being assayed. This common primer can be between 50 and 1500 bps from the polymorphic base. The other two (or more) primers are identical to each other except that the final 3' base wobbles to match one of the two (or more) alleles that make up the polymorphism. Two (or more) PCR reactions are then conducted on sample DNA, each using the common primer and one of the Allele Specific Primers.

"Splice Variant" as used herein refers to cDNA molecules produced from RNA molecules initially transcribed from the same genomic DNA sequence but which have undergone alternative RNA splicing. Alternative RNA splicing occurs when a primary RNA transcript undergoes splicing, generally for the removal of introns, which results in the production of more than one mRNA molecule each of that may encode different amino acid sequences. The term splice variant also refers to the proteins encoded by the above cDNA molecules.

"Identity" reflects a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, determined by comparing the sequences. In general, identity refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of the two polynucleotide or two polypeptide sequences, respectively, over the length of the sequences being compared.

"% Identity" - For sequences where there is not an exact correspondence, a "% identity" may be determined. In general, the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting "gaps" in either one

or both sequences, to enhance the degree of alignment. A % identity may be determined over the whole length of each of the sequences being compared (so-called global alignment), that is particularly suitable for sequences of the same or very similar length, or over shorter, defined lengths (so-called local alignment), that is more suitable for sequences of unequal length.

"Similarity" is a further, more sophisticated measure of the relationship between two polypeptide sequences. In general, "similarity" means a comparison between the amino acids of two polypeptide chains, on a residue by residue basis, taking into account not only exact correspondences between a between pairs of residues, one from each of the sequences being compared (as for identity) but also, where there is not an exact correspondence, whether, on an evolutionary basis, one residue is a likely substitute for the other. This likelihood has an associated "score" from which the "% similarity" of the two sequences can then be determined.

Methods for comparing the identity and similarity of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J et al, Nucleic Acids Res, 12, 387-395, 1984, available from Genetics Computer Group, Madison, Wisconsin, USA), for example the programs BESTFIT and GAP, may be used to determine the % identity between two polynucleotides and the % identity and the % similarity between two polypeptide sequences. BESTFIT uses the "local homology" algorithm of Smith and Waterman (J Mol Biol, 147,195-197, 1981, Advances in Applied Mathematics, 2, 482-489, 1981) and finds the best single region of similarity between two sequences. BESTFIT is more suited to comparing two polynucleotide or two polypeptide sequences that are dissimilar in length, the program assuming that the shorter sequence represents a portion of the longer. In comparison, GAP aligns two sequences, finding a "maximum similarity", according to the algorithm of Neddleman and Wunsch (J Mol Biol, 48, 443-453, 1970). GAP is more suited to comparing sequences that are approximately the same length and an alignment is expected over the entire length. Preferably, the parameters "Gap Weight" and "Length Weight" used in each program are 50 and 3, for polynucleotide sequences and 12 and 4 for polypeptide sequences, respectively. Preferably, % identities and

similarities are determined when the two sequences being compared are optimally aligned.

Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Altschul S F et al, J Mol Biol, 215, 403-410, 1990, Altschul S F et al, Nucleic Acids Res., 25:389-3402, 1997, available from the National Center for Biotechnology Information (NCBI), Bethesda, Maryland, USA and accessible through the home page of the NCBI at www.ncbi.nlm.nih.gov) and FASTA (Pearson W R, Methods in Enzymology, 183, 63-99, 1990; Pearson W R and Lipman D J, Proc Nat Acad Sci USA, 85, 2444-2448,1988, available as part of the Wisconsin Sequence Analysis Package).

Preferably, the BLOSUM62 amino acid substitution matrix (Henikoff S and Henikoff J G, Proc. Nat. Acad Sci. USA, 89, 10915-10919, 1992) is used in polypeptide sequence comparisons including where nucleotide sequences are first translated into amino acid sequences before comparison.

Preferably, the program BESTFIT is used to determine the % identity of a query polynucleotide or a polypeptide sequence with respect to a reference polynucleotide or a polypeptide sequence, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value, as hereinbefore described.

"Identity Index" is a measure of sequence relatedness which may be used to compare a candidate sequence (polynucleotide or polypeptide) and a reference sequence. Thus, for instance, a candidate polynucleotide sequence having, for example, an Identity Index of 0.95 compared to a reference polynucleotide sequence is identical to the reference sequence except that the candidate polynucleotide sequence may include on average up to five differences per each 100 nucleotides of the reference sequence. Such differences are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion. These differences may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between these terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or

more contiguous groups within the reference sequence. In other words, to obtain a polynucleotide sequence having an Identity Index of 0.95 compared to a reference polynucleotide sequence, an average of up to 5 in every 100 of the nucleotides of the in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

Similarly, for a polypeptide, a candidate polypeptide sequence having, for example, an Identity Index of 0.95 compared to a reference polypeptide sequence is identical to the reference sequence except that the polypeptide sequence may include an average of up to five differences per each 100 amino acids of the reference sequence. Such differences are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion. These differences may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between these terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. In other words, to obtain a polypeptide sequence having an Identity Index of 0.95 compared to a reference polypeptide sequence, an average of up to 5 in every 100 of the amino acids in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

The relationship between the number of nucleotide or amino acid differences and the Identity Index may be expressed in the following equation:

$$n_a \leq x_a - (x_a \bullet I),$$

in which:

n_a is the number of nucleotide or amino acid differences,

x_a is the total number of nucleotides or amino acids in SEQ ID NO:1 or SEQ ID NO:2, respectively,

I is the Identity Index,

• is the symbol for the multiplication operator, and

in which any non-integer product of x_a and I is rounded down to the nearest integer prior to subtracting it from x_a.

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a reference sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the two sequences as hereinbefore defined. Falling within this generic term are the terms "ortholog", and "paralog". "Ortholog" refers to a polynucleotide or polypeptide that is the functional equivalent of the polynucleotide or polypeptide in another species. "Paralog" refers to a polynucleotideor polypeptide that within the same species which is functionally similar.

"Fusion protein" refers to a protein encoded by two, unrelated, fused genes or fragments thereof. Examples have been disclosed in US 5541087, 5726044. In the case of Fc-SIP-1 ligand, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for performing the functional expression of Fc-SIP-1 ligand or fragments of the ligand, to improve pharmacokinetic properties of such a fusion protein when used for therapy and to generate a dimeric SIP-1 ligand. The Fc-SIP-1 ligand DNA construct comprises in 5' to 3' direction, a secretion cassette, i.e. a signal sequence that triggers export from a mammalian cell, DNA encoding an immunoglobulin Fc region fragment, as a fusion partner, and a DNA encoding SIP-1 ligand or fragments thereof. In some uses it would be desirable to be able to alter the intrinsic functional properties (complement binding, Fc-Receptor binding) by mutating the functional Fc sides while leaving the rest of the fusion protein untouched or delete the Fc part completely after expression.

All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set forth. Any patent application to which this application claims priority is also incorporated by reference herein in

its entirety in the manner described above for publications and references.

Discription of figures

Figure 1
Yeast-2-Hybrid detection if the interaction of Survivin with SIP-1 in *S. cerevisiae* EGY48/pSH18-32

Figure legend: *S. cerevisiae* EGY48/18-32 (Clontech) was transformed using plasmids encoding LexA- and Gal4-activation domain fusion proteins. Single colonies were isolated and streaked onto Media lacking Histidine, Tryptophane and Uracil (-WHU) and onto Medium lacking Histidine, Tryptophane, Leucine and Uracil containing 40 mg/l X-Gal (-WHLU-XGal). Growth of all strains on -WHU-Medium indicates presence of plasmids. Blue colour of yeast and growth on Medium lacking Leucine indicates interaction of baits and preys since expression of reportergenes depends on a Two-Hybrid selection scheme. Lanes 1 and 2 are used as unspecific controls. Lane 3 indicates the specific interaction of Survivin and SIP-1 in this Yeast-Two-Hybrid System.

Figure 2
SIP-1 Northern blot

Figure legend: Random primed ^{32}P-radiolabelled SIP-1 probe was used for hybridisation consisting of 590 pase pairs corresponding to position 553-1143 of the SIP-1 cds. The blot was loaded with 2µg Poly (A) RNA per lane (Blot from Invitrogen: „real human tumor panel blot 3# D3803-60).
A 2,8 kb SIP-1 mRNA is detected in mRNA-samples of kidney normal tissue, kidney tumor tissue, lung normal tissue and lung tumor tissue as indicated.

Figure 3
Cell cycle regulation of Survivin and SIP-1 mRNA expression

Figure legend: Multiplex PCR of mRNA derived from human HeLa cells was amplified using the „Gene Specific Relative RT-PCR Kit„ (Protocol according to producer of this kit system: Ambion).
Human HeLa cells were cell phase arrested using the following treatment (G1-Phase: L-Mimosine [400µM]; S-Phase: Thymidine [2mM]; G2/M Phase: Nocodazole [40 ng/ml]). The cell cycle status was controlled by FACS analysis; control cells were treated with 1µl DMSO/ ml medium (lane1 control; lane 2 G1-Phase; lane 3 S-phase; lane 4 G2/M-Phase)
SIP-1 mRNA expression appears to be upregulated during S-phase and G2/M phase of the cell cycle.

Figure 4
SIP-1/Survivin coimmunoprecipitation: SIP-1 and Survivin interact in human HeLa cells and after expression as in vitro translated proteins

Figure legend: The analysis of the SIP-1/Survivin interaction was done via V5-antibody mediated immunoprecipitation (IP) assays and detected by anti-V5 and anti-GFP antibodies as indicated.

Lanes:
A: Hela Transfection SIP-1-V5-His
B: Hela Transfection SIP-1-V5-His + GFP-Survivin
C: Hela Transfection SIP-1-V5-His + GFP-vector
1: In vitro translation/transcription SIP-1V5-His
2: In vitro translation/transcription SIP-1-V5-His + GFP-vector
3: In vitro translation/transcription SIP-1-V5-His + GFP-Survivin
4: In vitro translation/transcription without DANN
5: In vitro translation/transcription SIP-1-V5-His + GFP-Survivin
6: Anti-V5 Antibody (2µg)
7: Hela cell extract; transfection GFP-Survivin
Note: lanes 1-4: anti V5 IPs but lane 5*: anti-GFP IP

Further examples

Plasmid constructs
Cloning of pDBLeu – survivin: The survivin gene was cloned from cDNA of human PC-3 cells by using gene specific PCR primer pairs with in-built restriction sites for NheI and StuI, amplifying the coding sequence of

survivin. The PCR product was cloned via TA-cloning into the pCRII vector (Invitrogen). The insert was isolated via NheI/StuI restriction and cloned into the Y2H vector pDBLeu (PROQUEST™ Two-Hybrid System, GibcoBRL LIFE TECHNOLOGIES) creating pDBLeu-Survivin. This results in generation of a fusion gene construct for the expression of the Gal4 DB domain fused to the full length survivin peptid. All vectors were confirmed by sequencing.

PCR primers for the cloning of the survivin gene were the following:

NheI-5´survivin primer: Primer 1

StuI-3´survivin primer: Primer 2

Selection of Survivin interacting proteins

A yeast Two Hybrid screen (Proquest, Life Technologies) using pDBLeu-Survivin as the bait construct was performed using a human HeLa cell cDNA-Y2H-library and a cDNA library derived from human fetal brain tissue (selection was performed on –Trp, -Leu, -His Minimalmedium containing 25 mM 3-Aminotriazole). SIP-1 clones were isolated from both libraries. Interaction was confirmed by ß-Galactosidase-filter assay and Uracile, Tryptophane and Leucine lacking medium as described in the manufacturers protocoll.

One positive clone was isolated and sequenced. The corresponding interacting protein ligand was termed SIP-1 (Survivin-Interacting-Protein-1) ligand and the sequences have been disclosed in SEQ ID NO: 1 and NO: 2

Claims

1. A polypeptide selected from the group consisting of:

(a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1;

(b) a polypeptide comprising a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

c) a polypeptide having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

d) the polypeptide sequence of SEQ ID NO:2 and

(e) fragments and variants of such polypeptides in (a) to (d).

2. The polypeptide of claim 1 comprising the polypeptide sequence of SEQ ID NO:2.

3. The polypeptide of claim 1 which is the polypeptide sequence of SEQ ID NO:2.

4. A polynucleotide selected from the group consisting of:

(a) a polynucleotide comprising a polynucleotide sequence having at least 95% identity to the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide having at least 95% identity to the polynucleotide of SEQ ID NO:1;

(c) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

(d) a polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

(e) a polynucleotide with a nucleotide sequence of at least 100 nucleotides obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof having at least 15 nucleotides;

(f) a polynucleotide which is the RNA equivalent of a polynucleotide of (a) to (e);

(g) a polynucleotide sequence complementary to said polynucleotide of any one of (a) to (f), and

(h) polynucleotides that are variants or fragments of the polynucleotides of any one of (a) to (g) or that are complementary to above mentioned polynucleotides, over the entire length thereof.

5. A polynucleotide of claim 4 selected from the group consisting of:

(a) a polynucleotide comprising the polynucleotide of SEQ ID NO:1;

(b) the polynucleotide of SEQ ID NO:1;

(c) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2; and

(d) a polynucleotide encoding the polypeptide of SEQ ID NO:2.

6. An expression system comprising a polynucleotide capable of producing a polypeptide of any one of claim 1-3 when said expression vector is present in a compatible host cell.

7. A recombinant host cell comprising the expression vector of claim 6 or a membrane thereof expressing the polypeptide of any one of claim 1-3.

8. A process for producing a polypeptide of any one of claim 1-3 comprising the step of culturing a host cell as defined in claim 7 under conditions sufficient for

the production of said polypeptide and recovering the polypeptide from the culture medium.

9. A fusion protein consisting of the Immunoglobulin Fc-region and a polypeptide any one one of claims 1-3.

10. An antibody immunospecific for the polypeptide of any one of claims 1 to 3.

11. A method for screening to identify compounds that stimulate or inhibit the function or level of the polypeptide of any one of claim 1-3 comprising a method selected from the group consisting of:

(a) measuring or, detecting, quantitatively or qualitatively, the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

(b) measuring the competition of binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof in the presence of a labeled competitior;

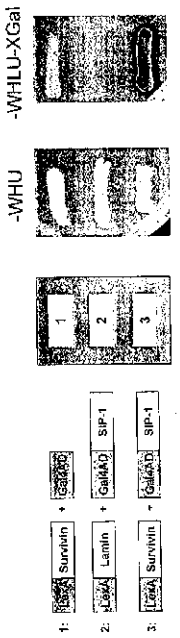
(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes expressing the polypeptide;

(d) mixing a candidate compound with a solution containing a polypeptide of any one of claims 1-3, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a control mixture which contains no candidate compound; or

(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide or said polypeptide in cells, using for instance, an ELISA assay, and

(f) producing said compound according to biotechnological or chemical standard techniques.

Figure1


-WHLU-XGal
-WHU
1
2
3
1: LexA Survivin + GalAD
2: LexA Lamin + GalAD SIP-1
3: LexA Survivin + GalAD SIP-1

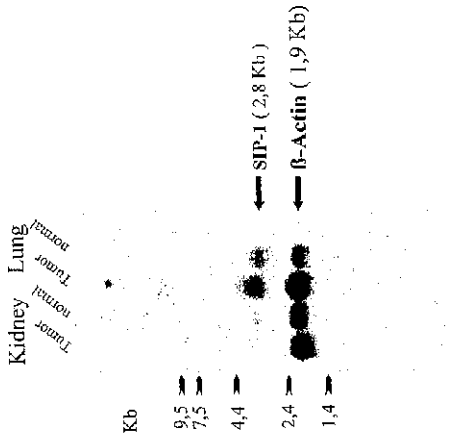

Kidney
Tumor
normal
Lung
Tumor
normal
SIP-1 (2,8 Kb)
ß-Actin (1,9 Kb)
Kb
9,5
7,5
4,4
2,4
1,4

Figure 2

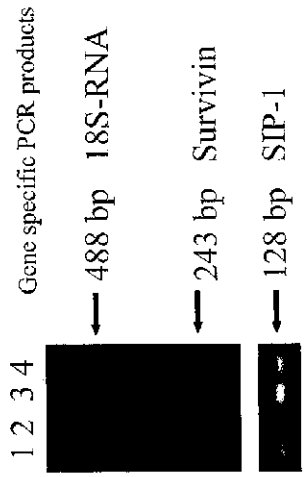

Gene specific PCR products
488 bp 18S-RNA
243 bp Survivin
128 bp SIP-1
1 2 3 4

Figure 3

Figure 4

SEQUENCE LISTING

<110> Merck Patent GmbH
Merck Patent GmbH

<120> Survivin Interacting Protein I

<130> SIPDHNS

<140>
<141>

<160> 4

<170> PatentIn Ver. 2.1

<210> 1
<211> 3263
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (233)..(2320)

<400> 1
tgacgcacgg aggggccggg acgcgatttg ccgcgagtga ccggagccgc aaccgtcgcc 60

gctgctgccg ccgccgcccg ggcccgctcc gcccccagcc cgccttcctc ccgccgcgcc 120

ctccgcctcc gcccgcaatt cggctacagt tcccctctct tccccaacgc cccggagccg 180

ccggccgcta gcgtcagcga cagccagaat taaggaagtt cactggagta aa atg gag 238
Met Glu
1

gcc tca gta ata tta ccc att ctg aag aaa aaa cta gcc ttc ctt tca 286
Ala Ser Val Ile Leu Pro Ile Leu Lys Lys Lys Leu Ala Phe Leu Ser
5 10 15

gga gga aag gac aga cgg agt ggc ctc att ttg aca att cca cta tgc 334
Gly Gly Lys Asp Arg Arg Ser Gly Leu Ile Leu Thr Ile Pro Leu Cys
20 25 30

ctc gaa cag aca aat atg gat gag ctg agt gtc acc tta gac tac cta 382
Leu Glu Gln Thr Asn Met Asp Glu Leu Ser Val Thr Leu Asp Tyr Leu
35 40 45 50

ctc agc att cca agt gag aag tgt aag gct aga gga ttt acc gtg att 430
Leu Ser Ile Pro Ser Glu Lys Cys Lys Ala Arg Gly Phe Thr Val Ile
55 60 65

gtg gat ggc aga aaa tca cag tgg aat gtg gtg aaa aca gta gtc gta 478
Val Asp Gly Arg Lys Ser Gln Trp Asn Val Val Lys Thr Val Val Val
70 75 80

atg cta cag aat gtt gtt cca gct gag gtg tcc ctt gtt tgt gtg gta 526
Met Leu Gln Asn Val Val Pro Ala Glu Val Ser Leu Val Cys Val Val
85 90 95

```
aag cca gat gaa ttc tgg gat aag aaa gta acg cat ttt tgt ttt tgg   574
Lys Pro Asp Glu Phe Trp Asp Lys Lys Val Thr His Phe Cys Phe Trp
    100                 105                 110

aag gag aag gat aga ctt ggc ttt gag gtt att tta gtg tcc gcc aac   622
Lys Glu Lys Asp Arg Leu Gly Phe Glu Val Ile Leu Val Ser Ala Asn
115                 120                 125                 130

aaa ttg act cgt tat ata gaa cca tgc caa tta aca gaa gat ttt ggt   670
Lys Leu Thr Arg Tyr Ile Glu Pro Cys Gln Leu Thr Glu Asp Phe Gly
                135                 140                 145

ggg agt ctc acc tat gat cac atg gac tgg tta aat aag agg ctg gtt   718
Gly Ser Leu Thr Tyr Asp His Met Asp Trp Leu Asn Lys Arg Leu Val
            150                 155                 160

ttt gag aag ttt aca aag gaa tct aca tca tta tta gat gaa ctt gct   766
Phe Glu Lys Phe Thr Lys Glu Ser Thr Ser Leu Leu Asp Glu Leu Ala
        165                 170                 175

ttg att aac aat ggc agt gat aaa gga aat cag caa gag aaa gaa agg   814
Leu Ile Asn Asn Gly Ser Asp Lys Gly Asn Gln Gln Glu Lys Glu Arg
    180                 185                 190

tct gtg gat tta aac ttt ctt ccc tcg gtt gat cct gaa aca gtt ctt   862
Ser Val Asp Leu Asn Phe Leu Pro Ser Val Asp Pro Glu Thr Val Leu
195                 200                 205                 210

cag aca ggg cat gaa ttg ttg tcc gta tta cag cag cgt aga ttt aat   910
Gln Thr Gly His Glu Leu Leu Ser Glu Leu Gln Gln Arg Arg Phe Asn
                215                 220                 225

ggc tca gac gga ggg gtt tca tgg tct cct atg gat gat gaa ctt ctt   958
Gly Ser Asp Gly Gly Val Ser Trp Ser Pro Met Asp Asp Glu Leu Leu
            230                 235                 240

gca cag cca cag gtt atg aaa tta tta gat tca ctc cga gag caa tat   1006
Ala Gln Pro Gln Val Met Lys Leu Leu Asp Ser Leu Arg Glu Gln Tyr
        245                 250                 255

acc cgc tac cag gaa gtt tgt agg caa cgt agc aag cgc aca cag tta   1054
Thr Arg Tyr Gln Glu Val Cys Arg Gln Arg Ser Lys Arg Thr Gln Leu
    260                 265                 270

gaa gag att caa cag aag gta atg cag gtg gtg aac tgg cta gaa ggg   1102
Glu Glu Ile Gln Gln Lys Val Met Gln Val Val Asn Trp Leu Glu Gly
275                 280                 285                 290

cct gga tca gaa caa cta aga gcc cag tgg ggc att gga gac tcc att   1150
Pro Gly Ser Glu Gln Leu Arg Ala Gln Trp Gly Ile Gly Asp Ser Ile
                295                 300                 305

agg gcc tcc cag gcc cta cag cag aaa cac gaa gag att gag agc cag   1198
Arg Ala Ser Gln Ala Leu Gln Gln Lys His Glu Glu Ile Glu Ser Gln
            310                 315                 320

cac agt gaa tgg ttt gca gtg tat gtg gaa ctt aat cag caa att gca   1246
His Ser Glu Trp Phe Ala Val Tyr Val Glu Leu Asn Gln Gln Ile Ala
        325                 330                 335
```

```
gca ctc ttg aat gct ggc gat gag gaa gat ctt gtg gaa cta aag tca  1294
Ala Leu Leu Asn Ala Gly Asp Glu Glu Asp Leu Val Glu Leu Lys Ser
    340                 345                 350

ctg cag caa caa ctt agt gat gtt tgt tat cga cag gcc agt cag ctg  1342
Leu Gln Gln Gln Leu Ser Asp Val Cys Tyr Arg Gln Ala Ser Gln Leu
355                 360                 365                 370

gaa ttt agg caa aat ctc tta caa gca gct ctt gaa ttt cat ggt gtt  1390
Glu Phe Arg Gln Asn Leu Leu Gln Ala Ala Leu Glu Phe His Gly Val
                375                 380                 385

gcc caa gat ttg tct cag cag ttg gat ggc tta tta ggg atg ttg tgc  1438
Ala Gln Asp Leu Ser Gln Gln Leu Asp Gly Leu Leu Gly Met Leu Cys
            390                 395                 400

gta gat gta gca cca gct gat gga gca tcg att cag caa act tta aaa  1486
Val Asp Val Ala Pro Ala Asp Gly Ala Ser Ile Gln Gln Thr Leu Lys
        405                 410                 415

cag ctt gaa gag aag ctg aaa agt gtt gat gtg gga ttg caa cgt ttg  1534
Leu Leu Glu Glu Lys Leu Lys Ser Val Asp Val Gly Leu Gln Gly Leu
    420                 425                 430

cgt gaa aaa gct cca ggt ctc ctg gat cag atc tcc aat cag gca tcc  1582
Arg Glu Lys Gly Gln Gly Leu Leu Asp Gln Ile Ser Asn Gln Ala Ser
435                 440                 445                 450

tgg gcc tat gga aag gat gta acc att gaa aat aaa gaa aat gtg gat  1630
Trp Ala Tyr Gly Lys Asp Val Thr Ile Glu Asn Lys Glu Asn Val Asp
                455                 460                 465

cac ata caa gga gtg atg gaa gat atg cag ctt aga aaa caa aga tgt  1678
His Ile Gln Gly Val Met Glu Asp Met Gln Leu Arg Lys Gln Arg Cys
            470                 475                 480

gaa gac atg gta gat gtg cga agg tta aag atg ctt cag atg ctg cag  1726
Glu Asp Met Val Asp Val Arg Arg Leu Lys Met Leu Gln Met Leu Gln
        485                 490                 495

ttg ttt aaa tgt gaa gaa gat gct gcc cag gca gta gaa tgg cta agt  1774
Leu Phe Lys Cys Glu Glu Asp Ala Ala Gln Ala Val Glu Trp Leu Ser
    500                 505                 510

gaa ctt ctg gat gct ctg ctt aag act cac atc aga ttg ggc gat gat  1822
Glu Leu Leu Asp Ala Leu Leu Lys Thr His Ile Arg Leu Gly Asp Asp
515                 520                 525                 530

gct caa gaa acg aaa gtt ttg ctg gaa cag cat aga aaa ttt gtt gat  1870
Ala Gln Glu Thr Lys Val Leu Leu Glu Gln His Arg Lys Phe Val Asp
                535                 540                 545

gtt gca cag agc act tat gac tat ggc agg cag ttg cta cag gcc aca  1918
Val Ala Gln Ser Thr Tyr Asp Tyr Gly Arg Gln Leu Leu Gln Ala Thr
            550                 555                 560

gtt gtc tta tgc caa tct ttg cgc tgc act tct cgg tca tct ggg gat  1966
Val Val Leu Cys Gln Ser Leu Arg Cys Thr Ser Arg Ser Ser Gly Asp
        565                 570                 575
```

```
aca ctt cct cga ctg aac aga gta tgg aaa caa ttt aca ata gca tct  2014
Thr Leu Pro Arg Leu Asn Arg Val Trp Lys Gln Phe Thr Ile Ala Ser
    580                 585                 590

gaa gag aga gta cat aga ttg gaa atg gct att gca ttt cac tca aat  2062
Glu Glu Arg Val His Arg Leu Glu Met Ala Ile Ala Phe His Ser Asn
595                 600                 605                 610

gct gaa aag att ttg cag gac tgt cca gaa cag cct gaa gct att aat  2110
Ala Glu Lys Ile Leu Gln Asp Cys Pro Glu Gln Pro Glu Ala Ile Asn
                615                 620                 625

gat gag gag caa ttt gat gaa att gaa gca gtt ggg aaa tca ctt ttg  2158
Asp Glu Glu Gln Phe Asp Glu Ile Glu Ala Val Gly Lys Ser Leu Leu
            630                 635                 640

gat aga tta act gtt cca gta gtt tat cct gat gga acc gaa caa tat  2206
Asp Arg Leu Thr Val Pro Val Val Tyr Pro Asp Gly Thr Glu Gln Tyr
        645                 650                 655

ttt ggg agt cca agt gac atg gct tct act gca caa aac atc aga gac  2254
Phe Gly Ser Pro Ser Asp Met Ala Ser Thr Ala Glu Asn Ile Arg Asp
    660                 665                 670

agg atg aaa cta gtt aat ccc aaa agg cag cag ctg aga cat cct gaa  2302
Arg Met Lys Leu Val Asn Leu Lys Arg Gln Gln Leu Arg His Pro Glu
675                 680                 685                 690

atg gtg acc aca gag agc taatagctac gagctacota cagatttgca         2350
Met Val Thr Thr Glu Ser
                695

gttcataatc ccgcatgttg tcaacataot acngcattag ccaccacacc ttaagatgca 2410

tttcacagcc aaaataagtc tcnttttctt tcatgacaca tttcictita catgttaaca 2470

ccttgctact accaaggcat anttacttaa catgcttcga ggctgtagat tccaagtatc 2530

ttaaaagaag gaactataaa cattgcactg aaaacttgct ttaaagcttt acctgacctg 2590

tcagtttgta gacaaacaac tgataataac ctttgaatgg tgctaataag agtaggaatt 2650

cnctntattn aaaaaaaaan aaaaagttgc ccttcctcca cagglgattt agtaaatttn 2710

gacagtagtt aaactcnttgt tagtagacag tggtgtcctc aaaattttac tttgtaatnc 2770

ttcagaattg attatttttn ttgtgtcaat acagagaaag cctttcagat ctttgatata 2830

tcatagtcat taaaagacct tttcctattt gtattgataa tgtattaaaa gttgtttgtg 2890

cttaataaaa gacttctttn aacattttat ttaatttagt agttacatcc tatttccaan 2950

catgagtgcc ttatttaaaa gggcattctt aggactgtga ggatggttta atatttgttt 3010

ttcatggtgg ttgcatgtat tttagacagg aaatacatat gtaagcatgt gtntataaca 3070

aataancatg cttttatcatg aaaaaattat tgtgnacaat ttaaatcttt aagaacttat 3130

taetaatgga atactattct aattttcctc ttttcnactt gaaaatttcc caaattatta 3190

acttcccaan aanatttgtn tttaggggaa gaaggtnaag aaaagggttc ttattcctan 3250
```

Lyttlecott eat 3263

<210> 2
<211> 696
<212> PRT
<213> Homo sapiens

<400> 2
Met Glu Ala Ser Val Ile Leu Pro Ile Leu Lys Lys Lys Leu Ala Phe
1 5 10 15

Leu Ser Gly Gly Lys Asp Arg Arg Ser Gly Leu Ile Leu Thr Ile Pro
20 25 30

Leu Cys Leu Glu Gln Thr Asn Met Asp Glu Leu Ser Val Thr Leu Asp
35 40 45

Tyr Leu Leu Ser Ile Pro Ser Glu Lys Cys Lys Ala Arg Gly Phe Thr
50 55 60

Val Ile Val Asp Gly Arg Lys Ser Gln Trp Asn Val Val Lys Thr Val
65 70 75 80

Val Val Met Leu Gln Asn Val Val Pro Ala Glu Val Ser Leu Val Cys
85 90 95

Val Val Lys Pro Asp Glu Phe Trp Asp Lys Lys Val Thr His Phe Cys
100 105 110

Phe Trp Lys Glu Lys Asp Arg Leu Gly Phe Glu Val Ile Leu Val Ser
115 120 125

Ala Asn Lys Leu Thr Arg Tyr Ile Glu Pro Cys Gln Leu Thr Glu Asp
130 135 140

Phe Gly Gly Ser Leu Thr Tyr Asp His Met Asp Trp Leu Asn Lys Arg
145 150 155 160

Leu Val Phe Glu Lys Phe Thr Lys Glu Ser Thr Ser Leu Leu Asp Glu
165 170 175

Leu Ala Leu Ile Asn Asn Gly Ser Asp Lys Gly Asn Gln Gln Glu Lys
180 185 190

Glu Arg Ser Val Asp Leu Asn Phe Leu Pro Ser Val Asp Pro Glu Thr
195 200 205

Val Leu Gln Thr Gly His Glu Leu Leu Ser Glu Leu Gln Gln Arg Arg
210 215 220

Phe Asn Gly Ser Asp Gly Gly Val Ser Trp Ser Pro Met Asp Asp Glu
225 230 235 240

Leu Leu Ala Gln Pro Gln Val Met Lys Leu Leu Asp Ser Leu Arg Gln
245 250 255

Gln Tyr Thr Arg Tyr Gln Glu Val Cys Arg Gln Arg Ser Lys Arg Thr
260 265 270

Gln Leu Glu Glu Ile Gln Gln Lys Val Met Gln Val Val Asn Trp Leu
275 280 285

Glu Gly Pro Gly Ser Glu Gln Leu Arg Ala Gln Trp Gly Ile Gly Asp
290 295 300

Ser Ile Arg Ala Ser Gln Ala Leu Gln Gln Lys His Gln Gln Ile Glu
305 310 315 320

Ser Gln His Ser Glu Trp Phe Ala Val Tyr Val Glu Leu Asn Gln Gln
325 330 335

Ile Ala Ala Leu Leu Asn Ala Gly Asp Glu Glu Asp Leu Val Glu Leu
340 345 350

Lys Ser Leu Gln Gln Gln Leu Ser Asp Val Cys Tyr Arg Gln Ala Ser
355 360 365

Gln Leu Glu Phe Arg Gln Asn Leu Leu Gln Ala Ala Leu Glu Phe His
370 375 380

Gly Val Ala Gln Asp Leu Ser Gln Gln Leu Asp Gly Leu Leu Gly Met
385 390 395 400

Leu Cys Val Asp Val Ala Pro Ala Asp Gly Ala Ser Ile Gln Gln Thr
405 410 415

Leu Lys Leu Leu Glu Glu Lys Leu Lys Ser Val Asp Val Gly Leu Gln
420 425 430

Gly Leu Arg Glu Lys Gly Gln Gly Leu Leu Asp Gln Ile Ser Asn Gln
435 440 445

Ala Ser Trp Ala Tyr Gly Lys Asp Val Thr Ile Glu Asn Lys Glu Asn
450 455 460

Val Asp His Ile Gln Gly Val Met Glu Asp Met Gln Leu Arg Lys Gln
465 470 475 480

Arg Cys Glu Asp Met Val Asp Val Arg Arg Leu Lys Met Leu Gln Met
485 490 495

Val Gln Leu Phe Lys Cys Glu Glu Asp Ala Ala Gln Ala Val Glu Trp
500 505 510

Leu Ser Glu Leu Leu Asp Ala Leu Leu Lys Thr His Ile Arg Leu Gly
515 520 525

Asp Asp Ala Gln Glu Thr Lys Val Leu Leu Glu Lys His Arg Lys Phe
530 535 540

Val Asp Val Ala Gln Ser Thr Tyr Asp Tyr Gly Arg Gln Leu Leu Gln
545 550 555 560

Ala Thr Val Val Leu Cys Gln Ser Leu Arg Cys Thr Ser Arg Ser Ser
565 570 575

Gly Asp Thr Leu Pro Arg Leu Asn Arg Val Trp Lys Gln Phe Thr Ile
580 585 590

Ala Ser Glu Glu Arg Val His Arg Leu Glu Met Ala Ile Ala Phe His
595 600 605

Ser Asn Ala Glu Lys Ile Leu Gln Asp Cys Pro Glu Glu Pro Glu Ala
610 615 620

Ile Asn Asp Glu Glu Gln Phe Asp Glu Ile Glu Ala Val Gly Lys Ser
625 630 635 640

Leu Leu Asp Arg Leu Thr Val Pro Val Val Tyr Pro Asp Gly Thr Glu
645 650 655

Gln Tyr Phe Gly Ser Pro Asn Asp Met Ala Ser Thr Ala Glu Asn Ile
660 665 670

Arg Asp Arg Met Lys Leu Val Asn Leu Lys Arg Gln Gln Leu Arg His
675 680 685

Pro Glu Met Val Thr Thr Glu Ser
690 695

<210> 3
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer 1
(NheI-5' survivin primer)

<400> 3
ctagctagca tgggtgcccc gacgttg 27

<210> 4
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer 2
(StuI-3' survivin primer)

<400> 4
taggccttca atccatgcca gccag 25

【国際調査報告】

INTERNATIONAL SEARCH REPORT

Int... ...l Application No
PCT/EP 01/03204

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 C12N15/62 C07K16/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

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)
EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL, HEIDELBERG, FRG [Online] 20 October 1999 (1999-10-20) BLOECKER, H. ET AL.: "Homo sapiens mRNA, cDNA DKFZp434o0515 (from clone DKFZp434o0515)" Database accession no. AL122046 XP002174514 the whole document ---	1,4,6-9
X	DATABASE EMBL, HEIDELBERG, FRG [Online] 18 April 1997 (1997-04-18) ADAMS, M.D. ET AL.: "EST177826 Jurkat T-cells VI Homo sapiens cDNA 5' end" Database accession no. AA306831 XP002174515 the whole document --- -/--	1,4,6-9

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 August 2001	23 8 01
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Fuchs, U

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int. Application No
PCT/EP 01/03204

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 00 58473 A (CURAGEN CORPORATION) 5 October 2000 (2000-10-05) abstract page 1, line 1 -page 5, line 28 page 12, line 1 -page 56, line 29 page 72, line 19 -page 74, line 27 page 91, line 7 - line 15 ORF ID NO: 3026 page 389; table 1 page 410 -page 420; table 2 SEQ ID NOS: 6051 and 6052 page 5230 -page 5233 page 5503 -page 5507; claims 1-32	1,4,6-11
A	AMBROSINI, G. ET AL.: "A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma" NATURE MEDICINE, vol. 3, no. 8, August 1997 (1997-08), pages 917-921, XP002074968 the whole document	1-11

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int. Application No
PCT/EP 01/03204

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0058473 A	05-10-2000	AU 3774500 A	16-10-2000

Form PCT/ISA/210 (patent family annex) (July 1992)

フロントページの続き

(51)Int.Cl.[7]	FI		テーマコード（参考）
Ｃ１２Ｎ 1/19	Ｃ１２Ｎ 1/19		４Ｈ０４５
Ｃ１２Ｎ 1/21	Ｃ１２Ｎ 1/21		
Ｃ１２Ｎ 5/10	Ｃ１２Ｐ 21/02	Ｃ	
Ｃ１２Ｐ 21/02	Ｃ１２Ｑ 1/02		
Ｃ１２Ｑ 1/02	Ｇ０１Ｎ 33/15	Ｚ	
Ｇ０１Ｎ 33/15	Ｇ０１Ｎ 33/50	Ｚ	
Ｇ０１Ｎ 33/50	Ｇ０１Ｎ 33/53	Ｄ	
Ｇ０１Ｎ 33/53	Ｇ０１Ｎ 33/53	Ｍ	
Ｇ０１Ｎ 33/566	Ｇ０１Ｎ 33/566		
	Ｃ１２Ｎ 5/00	Ａ	

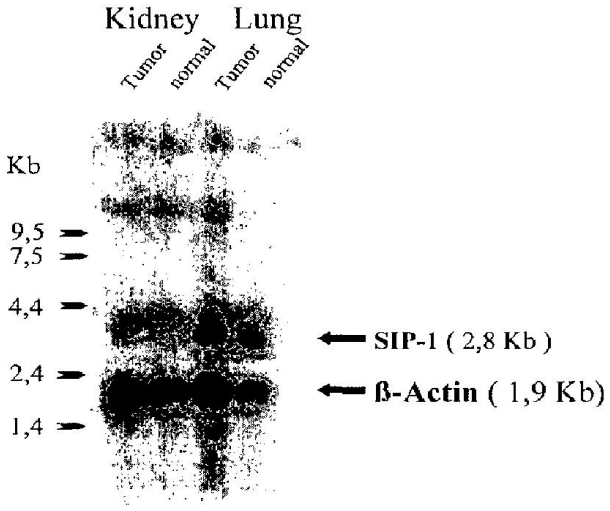
(74)代理人 100088328
弁理士 金田 暢之
(74)代理人 100106297
弁理士 伊藤 克博
(74)代理人 100106138
弁理士 石橋 政幸
(72)発明者 ヘンチュ、 ベルント
ドイツ連邦共和国 ６４２８９ ダルムシュタット カンプシュトラーセ ４８
(72)発明者 デュッカー、 クラウス
ドイツ連邦共和国 ６４２９１ ダルムシュタット エテシュテルシュトラーセ ５
(72)発明者 ホック、 ビューン
ドイツ連邦共和国 ６３４７７ マインタル ノルトシュトラーセ ２アー

Ｆターム(参考) 2G045 AA40 DA12 DA13 DA14 DA36 FB02 FB03
4B024 AA01 AA11 BA44 BA80 CA01 CA04 CA07 CA11 DA01 DA02
DA05 DA11 EA01 EA02 EA03 EA04 FA02 GA01 GA11 HA01
HA03 HA11
4B063 QA01 QA18 QQ05 QQ13 QQ42 QQ52 QR08 QR33 QR42 QR55
QR59 QR62 QR80 QS05 QS36 QX02
4B064 AG01 AG27 CA01 CA19 CA20 CC24 DA01 DA13
4B065 AA01X AA57X AA87X AA93Y AB01 AB02 BA01 BA08 CA24 CA25
CA44 CA46
4H045 AA10 AA11 AA20 AA30 BA10 BA41 CA40 DA00 DA76 EA20
EA50 FA72 FA74

专利名称(译)	人类存活蛋白相互作用蛋白1（SIP-1）		
公开(公告)号	JP2004500111A	公开(公告)日	2004-01-08
申请号	JP2001570731	申请日	2001-03-21
申请(专利权)人(译)	默克专利GESELLSCHAFT手套Beshurenkuteru Hafutongu		
[标]发明人	ヘンチュベルント デュッカークラウス ホックビューン		
发明人	ヘンチュ、 ベルント デュッカー、 クラウス ホック、 ビューン		
IPC分类号	G01N33/50 C07K14/47 C07K16/18 C07K19/00 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12N15/09 C12N15/12 C12N15/62 C12P21/02 C12Q1/02 G01N33/15 G01N33/53 G01N33/566		
CPC分类号	C07K14/4747 C07K2319/00		
FI分类号	C12N15/00.ZNA.A C07K14/47 C07K16/18 C07K19/00 C12N1/15 C12N1/19 C12N1/21 C12P21/02.C C12Q1/02 G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 C12N5/00.A		
F-TERM分类号	2G045/AA40 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/FB02 2G045/FB03 4B024/AA01 4B024/AA11 4B024/BA44 4B024/BA80 4B024/CA01 4B024/CA04 4B024/CA07 4B024/CA11 4B024/DA01 4B024/DA02 4B024/DA05 4B024/DA11 4B024/EA01 4B024/EA02 4B024/EA03 4B024/EA04 4B024/FA02 4B024/GA01 4B024/GA11 4B024/HA01 4B024/HA03 4B024/HA11 4B063/QA01 4B063/QA18 4B063/QQ05 4B063/QQ13 4B063/QQ42 4B063/QQ52 4B063/QR08 4B063/QR33 4B063/QR42 4B063/QR55 4B063/QR59 4B063/QR62 4B063/QR80 4B063/QS05 4B063/QS36 4B063/QX02 4B064/AG01 4B064/AG27 4B064/CA01 4B064/CA19 4B064/CA20 4B064/CC24 4B064/DA01 4B064/DA13 4B065/AA01X 4B065/AA57X 4B065/AA87X 4B065/AA93Y 4B065/AB01 4B065/AB02 4B065/BA01 4B065/BA08 4B065/CA24 4B065/CA25 4B065/CA44 4B065/CA46 4H045/AA10 4H045/AA11 4H045/AA20 4H045/AA30 4H045/BA10 4H045/BA41 4H045/CA40 4H045/DA00 4H045/DA76 4H045/EA20 4H045/EA50 4H045/FA72 4H045/FA74		
代理人(译)	宫崎昭雄 伊藤 克博		
优先权	2000106408 2000-03-24 EP		
外部链接	Espacenet		

摘要(译)

公开了存活蛋白相互作用蛋白1和多核苷酸以及通过重组技术产生这种多肽的方法。还公开了在诊断测定中使用SIP-1配体多肽和多核苷酸的方法。


Kidney
Lung
Tumor
normal
Tumor
normal
Kb
9,5
7,5
4,4
2,4
1,4
SIP-1 (2,8 Kb)
ß-Actin (1,9 Kb)