

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3569766号

(P3569766)

(45) 発行日 平成16年9月29日(2004.9.29)

(24) 登録日 平成16年7月2日(2004.7.2)

(51) Int.Cl.⁷

F I

C 1 2 N 15/09

C 1 2 N 15/00

Z N A A

C O 7 K 14/16

C O 7 K 14/16

C 1 2 P 21/02

C 1 2 P 21/02

C

G O 1 N 33/53

G O 1 N 33/53

M

G O 1 N 33/566

G O 1 N 33/566

発明の数 6 (全 55 頁) 最終頁に続く

(21) 出願番号 特願2002-137039 (P2002-137039)
 (22) 出願日 平成14年5月13日(2002.5.13)
 (62) 分割の表示 特願平6-98231の分割
 原出願日 昭和60年10月11日(1985.10.11)
 (65) 公開番号 特開2003-38190 (P2003-38190A)
 (43) 公開日 平成15年2月12日(2003.2.12)
 審査請求日 平成14年5月13日(2002.5.13)
 (31) 優先権主張番号 659339
 (32) 優先日 昭和59年10月10日(1984.10.10)
 (33) 優先権主張国 米国(US)
 (31) 優先権主張番号 693866
 (32) 優先日 昭和60年1月23日(1985.1.23)
 (33) 優先権主張国 米国(US)

(73) 特許権者 594078353
 セントコア インコーポレーテッド
 アメリカ合衆国 ペンシルバニア州 マル
 バーン グレート バリー パークウェイ
 200番
 (74) 代理人 100089705
 弁理士 社本 一夫
 (74) 代理人 100075236
 弁理士 栗田 忠彦
 (72) 発明者 ナンシー・ティー・チャン
 アメリカ合衆国ペンシルバニア州1930
 1, パオリ, シュガータウン・ロード 1
 504

最終頁に続く

(54) 【発明の名称】 HTLV-III DNAのクローニングおよび発現

(57) 【特許請求の範囲】

【請求項1】

HTLV-IIIのpol DNA及びsor DNA由来であり、且つ、図8～図10のヌクレオチド42
 28-5323のEcoRI-EcoRI断片

【化 1】

gaat tccctacaat

ccccaaagtc aaggagtagt agaattctatg aataaagaat taaagaaaat tataggacag
 gtaagagatc aggctgaaca tcttaagaca gcagtacaaa tggcagtatt catccacaat
 tttaaaagaa aaggggggat tggggggtac agtgcagggg aaagaatagt agacataata
 gcaacagaca tacaaactaa agaattacaa aaacaaatta caaaaattca aaattttcgg
 gtttattaca gggacagcag aaatccactt tggaaaggac cagcaaagct cctctggaaa
 ggtgaagggg cagtagtaat acaagataat agtgacataa aagtagtgcc aagaagaaaa
 gcaaagatca ttagggatta tggaaaacag atggcaggtg atgatttgtt ggcaagtaga
 caggatgagg attagaacat ggaaaagttt agtaaaacac catatgtatg tttcagggaa
 agctagggga tggttttata gacatcacta tgaaagccct catccaagaa taagttcaga
 agtacacatc ccactagggg atgctagatt ggtaataaca acatattggg gtctgcatac
 aggagaaaaga gactggcatt tgggtcaggg agtctccata gaatggagga aaaagagata
 tagcacacaa gtagaccctg aactagcaga ccaactaatt catctgtatt actttgactg
 tttttcagac tctgctataa gaaaggcctt attaggacac atagttagcc ctaggtgtga
 atatcaagca ggacataaca aggtaggatc tctacaatac ttggcactag cagcattaat
 aacaccaaaaa aagataaagc caccitttgc tagtgttacg aaactgacag aggatagatg
 gaacaagccc cagaagacca agggccacag agggagccac acaatgaatg gacactagag
 ctttttagagg agcttaagaa tgaagctgtt agacattttc ctaggatttg gctccatggc
 ttagggcaac atatctatga aacttatggg gatactiggg caggagtgga agccataata
 ag

10

20

30

によりコードされる、後天性免疫不全症候群の患者の血清またはHTLV-IIIに対する抗体を含む血清に免疫反応性である、単離された組換えHTLV-IIIポリペプチド。

【請求項 2】

40

HTLV-III DNA 断片を含む組換えベクターで形質転換された細胞により発現される、請求項 1 記載のポリペプチド。

【請求項 3】

組換えベクターがOmpAまたはpMR100である、請求項 2 記載のポリペプチド。

【請求項 4】

β - ガラクトシダーゼとの融合タンパク質として生産される、請求項 3 記載のポリペプチド。

【請求項 5】

以下の工程：

(a) 固相に固定した請求項 1 乃至 4 の何れか 1 項に記載の単離されたHTLV-IIIポリペプ

50

チドを含む免疫吸着剤を、単離された体液と接触させることにより、体液中のHTLV-IIIポリペプチドに対する抗体を固定化HTLV-IIIポリペプチドと結合させ；

(b) 体液から該免疫吸着剤を分離し；

(c) 該免疫吸着剤を標識HTLV-IIIポリペプチドまたはヒト免疫グロブリンに対する標識抗体と接触させ；そして

(d) HTLV-IIIに対する抗体の存在の指標として、免疫吸着剤に結合した工程(c)の標識ポリペプチドまたは標識抗体の存在を検出すること；

からなる、単離されたHTLV-IIIポリペプチドに対する体液中の抗体を検出する方法。

【請求項6】

請求項1乃至4の何れか1項に記載の単離されたHTLV-IIIポリペプチドを試料と接触させ；そして、前記ポリペプチドと結合した抗体の存在を検出する工程からなる、HTLV-IIIに対する抗体を検出するための免疫アッセイ。

10

【請求項7】

請求項1乃至4の何れか1項に記載の単離されたHTLV-IIIポリペプチドを含む、診断試薬。

【請求項8】

ポリペプチドが標識されている、請求項7に記載の診断試薬。

【請求項9】

ポリペプチドが固定化されている固相をさらに含む、請求項7に記載の診断試薬。

【請求項10】

20

請求項1乃至4の何れか1項に記載の単離されたHTLV-IIIポリペプチドと特異的に反応する固相固定化抗体、および該HTLV-IIIポリペプチドと特異的に反応する標識可溶性抗体を含む、アッセイキット。

【請求項11】

(i) 固相に固定化した請求項1乃至4の何れか1項に記載の単離されたHTLV-IIIポリペプチドを含む、免疫吸着剤、および

(ii) 標識HTLV-IIIポリペプチドまたはヒト免疫グロブリンに対する標識抗体を含む、体液中のHTLV-IIIに対する抗体の存在を測定するためのアッセイキット。

【発明の詳細な説明】

【0001】

30

【産業上の利用分野】

本発明は分子生物学およびウイルス学の分野、特にヒトT細胞白血病ウイルス - III型 (HTLV - III) に関する。

【0002】

【従来の技術】

ヒトT細胞白血病 - リンパ腫ウイルス (HTLV) という用語は、T細胞トロピックレトロウイルスの特定の属を意味する。これらのウイルスはある種のT細胞新生物の発生において重要な役割を演ずる。現在、3つの型のHTLVが知られている。この属の1つの亜属であるHTLV - I型 (HTLV - I) は、日本、カリブ海沿岸およびアフリカのある地域に発生する成人T細胞白血病 - リンパ腫 (ATLL) の原因と関係がある。HTLV - II型 (HTLV - II) は毛状細胞白血病のT細胞変異型をもつ患者から単離された。M. ポポビック (Popovic) らの「エイズおよびエイズ患者からの細胞変性レトロウイルス (HTLV - III) の検出、単離および連続生産」、サイエンス, 224: 497 - 500 (1984) を参照されたい。

40

【0003】

HTLV - III型 (HTLV - III) は後天性免疫不全症候群 (AIDS, エイズ) にかかった多数の患者から単離された。HTLV - IIIはエイズ患者から単離されたプロトタイプのウイルスである。このエイズに対して最も危険な立場にあると報告された人々には、同性愛または両性愛の男性、静脈内薬物使用者および米国のハイチ人移民が含まれる。提供者から集めた血液製剤を受容する血友病患者および頻繁に輸血を受ける人々も

50

危険な立場にある。エイズの臨床上的症状は、一般にヘルパーＴリンパ球の減少をとともなう今だ解明されていない恐ろしい免疫不全を包含する。これらは悪性腫瘍および感染症をとともなうことがある。エイズ患者の死亡率は高い。あまり恐ろしくない型のエイズも存在し、これはリンパ腫および低下したヘルパーＴ細胞数をともなうことがあるが、成熟したエイズの破壊的症状を示さない。これらの徴候を呈する初期エイズ（プレエイズ）患者も多数存在する。これらの人々のなかで誰が一層恐ろしい症状へ進行するかを予言することは今のところ不可能である。

【 0 0 0 4 】

伝染性エイズの病因としてHTLV-IIIが含まれることは多くのことから証明されている。第一に、首尾一貫した疫学が存在し、すなわちエイズ患者の95%以上がHTLV-IIIに対して特異的な抗体を有する。第二に、この病気において再現できるウイルスの同定および単離がなされ、HTLV-IIIの100種以上の変異型がエイズ患者から単離された。第三に、エイズに感染した血液提供者から輸血を受けた正常な健康人がエイズに感染する。

10

【 0 0 0 5 】

HTLV-IIIはHTLV-IおよびHTLV-IIといくつかの特性を共有するが、形態学的、生物学的および抗原的に区別し得ることがわかった。R.C.ガロ(Gallo)らの「危険な状態にあるエイズ患者からの細胞変性レトロウイルスの検出および単離」、サイエンス、224:500-503(1984)を参照されたい。例えば、HTLV-IおよびHTLV-IIのコアタンパク質p24およびp19、ならびにエンベロープ抗原に対する抗体との交差反応性を証明するか、あるいはクローニングしたHTLV-IおよびHTLV-IIのDNAとの核酸交差ハイブリダイゼーション実験を行うことにより、HTLV-IIIがHTLV-IおよびHTLV-IIと抗原的に関係のあることがわかった。しかしながら、HTLV-IおよびHTLV-IIと異なって、HTLV-IIIは正常人の臍帯血液および骨髓からのＴ細胞にインビトロで感染しかつそれを形質転換する能力を欠いており、感染細胞に対してのみ細胞変性作用を有していた。

20

【 0 0 0 6 】

他のレトロウイルスのRNAゲノムと同様に、HTLV-IIIのRNAゲノムはウイルスタンパク質をコードする3つの遺伝子を含んでいる：すなわち1)内部構造(ヌクレオカプシドまたはコア)タンパク質をコードするgag遺伝子；2)RNA依存性DNAポリメラーゼ(逆転写酵素)をコードするpol遺伝子；および3)ウイルス粒子のエンベロープ糖タンパク質をコードするenv遺伝子。さらに、HTLV-IIIゲノムはenv遺伝子と3'側LTRとの間に位置するPxと呼ばれる領域を含んでおり、これはウイルスの機能的死亡と関連があるように思われる。

30

【 0 0 0 7 】

現在のところ、エイズの症状が現われる以前にエイズを診断することはまだ困難である。この病気の予防に利用できる方法も目下のところ存在しない。一般にエイズ患者の治療は成功せず、そしてHTLV-IIIが身体に及ぼす破壊的作用に多数の犠牲者が出ている。

【 0 0 0 8 】

40

【発明が解決しようとする問題点】

本発明は、免疫反応性HTLV-IIIポリペプチドを発現できる組換え体/ベクター宿主系内でのHTLV-III DNAのクローニングに基づいている。免疫反応性ポリペプチドを発現する系内でのHTLV-III DNAのクローニングに基づいて、本出願人はエイズの診断、治療および予防に有用な方法を開発した。本出願人は体液(例えば血液、唾液、精液)中のHTLV-IIIおよびその抗体を検出する方法、ならびに免疫療法(例えばエイズの予防接種や受動免疫)において有用な方法を開発した。さらに、本出願人は体液中のHTLV-IIIの検出に有用なHTLV-III DNAプローブおよびRNAプローブの作製方法を開発した。

【 0 0 0 9 】

50

HTLV-I IIIゲノムのセグメントによってコードされるポリペプチドは、これらの組換えDNA法を使って生産した。例えば、HTLV-I IIIゲノムの3つの領域（env遺伝子配列、env-lor遺伝子配列、およびHTLV-I III cDNAからの1.1 KbのEcoRI制限断片）によってコードされるポリペプチドを生産した。発現されたポリペプチドは単離した。これらのポリペプチドはエイズ患者の血清およびHTLV-I IIIに対する抗体と免疫反応性であり、従ってHTLV-I IIIに対する抗体の有無について血液およびその他の体液をスクリーニングするのに有用である。それ故、本発明はエイズの診断方法ばかりでなく、HTLV-I IIIを含む血液または血液成分を通じて他の人々へのこの病気が伝染するのを予防する方法をも提供するものである。特に後者は、血液成分（例えば血友病の治療のための第VIII因子；第IX因子）を得るために提供された血液を輸血または使用するに先立って、その血液をスクリーニングするのに価値がある。

10

【0010】

組換えDNA法によって生産されたポリペプチドは、このウイルスに対する抗体（モノクローナル抗体を含む）の生産に利用できる。この種の抗体は血液、唾液、精液などの体液中のHTLV-I IIIを直接検出するためのイムノアッセイおよび診断技術の基礎をなすものである。このウイルスに対する中和抗体は、この病気に対して受動免疫を与えるのに使用できる。

【0011】

このような組換え体ノベクター宿主系内でのHTLV-I III DNAのクローニングは、さらにHTLV-I III DNAのヌクレオチド塩基配列を決定するための基礎を提供する。DNAプローブはHTLV-I IIIゲノムに特異的なDNA領域と相同であり、これらのDNAプローブは血液、唾液または他の体液中のHTLV-I IIIを検出する別の方法を提供する。HTLV-I IIIゲノムに特異的な領域を含むRNAプローブも作製され、体液中のHTLV-I IIIの検出に使用される。

20

【0012】

【問題点を解決するための手段】

HTLV-I IIIとHTLV-ウシ白血病ウイルス（BLV）属の他のウイルスとの間の類似性にもかかわらず、HTLV-I IIIの生物学および病理学は実質的に異なっている。例えば、比較的乏しい相同がHTLV-Iまたは-I IIIゲノムのそれと比較したときHTLV-I IIIゲノムに見られた。HTLV-I IIIによる感染はしばしば強い免疫抑制（エイズ）をもたらし、その結果OKT4（+）細胞集団の減少をきたす。この作用はインビトロでのリンパ球培養において、OKT4（+）細胞へのHTLV-I III感染の、形質転換作用というよりむしろ、著しい細胞変性作用によって反映される。これとは対照的に、HTLV-Iによる感染はT細胞白血病-リンパ腫（OKT4（+）細胞の悪性化）の低い発生率をもたらす。HTLV-I患者の場合もある程度の免疫不全を示す。HTLV-Iおよび-I IIIによる一次リンパ球培養物の感染は、主にOKT4（+）細胞のインビトロ形質転換を引き起こす。リンパ球へのHTLV-I感染の細胞変性作用は明らかに起こるが、HTLV-I IIIの場合に観察されるものほど顕著な細胞変性作用ではない。

40

【0013】

HTLV-I IIIはさらにインビボおよびインビトロでの伝染性ウイルス粒子生産の程度においてもHTLV-Iおよび-I IIIと異なっている。細胞不含の高力価伝染性ウイルス粒子はエイズ患者の精液および唾液、ならびにHTLV-I IIIを感染させた培養物の上清から得ることができる。成人T細胞白血病-リンパ腫（ATLL）患者またはHTLV-Iもしくは-I IIIを感染させた培養物からは、もし存在するとしても、きわめて少数の細胞不含伝染性ウイルス粒子を回収できるにすぎない。

【0014】

エンベロープ糖タンパク質はエイズ患者の抗血清によって認識される主な抗原である。この点で、HTLVは他のレトロウイルスと類似しており、そのためにエンベロープ糖タン

50

パク質は一般に最も抗原的なウイルスポリペプチドである。さらに、中和抗体は一般にレトロウイルスのエンベロープ糖タンパク質に対して向けられる。エイズ患者からの血清試料の88～100%はHTLV-IIIの抗原と反応する抗体を含むことが見出され、主な免疫反応はHTLV-IIIの推定上のエンベロープ抗原であるp41に対して向けられた。コアタンパク質に対する抗体もエイズ患者の血清に見出されたが、エンベロープ抗原に対する抗体の存在ほど感染指示剤として有効であるとは思われない。

【0015】

HTLV-IIIのp41抗原は、ウイルスエンベロープがウイルスの不活化および精製過程で部分的に破壊されるので、その性状決定が難しかった。本発明はHTLV-IIIウイルスの抗原成分の性状を決定しかついくつかの方法で他のウイルス抗原成分の存在および特性を決定するという大いなる要求に答えるものである。それはHTLV-IIIポリペプチド、そのポリペプチドに対する抗体、RNAプローブおよびDNAプローブのような産物を提供し、またそれらの生産方法も提供する。これらはスクリーニング、診断および治療を行う際に役立つ。

10

【0016】

本発明は、HTLV-IIIタンパク質をコードする組換えDNA塩基配列の翻訳によって生産されるHTLV-IIIポリペプチドに関する。この方法で生産され、エイズ患者からの血清またはHTLV-IIIに対する抗体と免疫反応性のポリペプチドは、組換えDNA-生産による免疫反応性HTLV-IIIポリペプチドと称される。これらにはHTLV-IIIコアタンパク質およびエンベロープ糖タンパク質をそれぞれコードするgagおよびenv DNA配列に特異的な組換えDNA配列の翻訳によって生産される抗原HTLV-IIIコアポリペプチドおよびエンベロープポリペプチドが含まれるが、これらに限定されない。それらはまたHTLV-III cDNAの1.1 KbのEcoRI制限断片に含まれる組換えDNA配列の翻訳、およびHTLV-IIIのsor遺伝子およびPx遺伝子に特異的な組換えDNA配列の翻訳によって生産されるポリペプチドをも包含する。sor DNA配列は複製感応HTLV-IIIウイルスに共通している。Px遺伝子はHTLV-IIIゲノムのenv遺伝子と3'末端との間に位置し、1つの大きな読み取り枠を有するコーディング配列(lor)を含む。env DNA配列とlor DNA配列とは共にHTLV-IIIゲノムの同じ読み取り枠内に位置するため、この遺伝子領域はenv-lorと称される。

20

30

【0017】

HTLV-IIIのこれらの領域によりコードされるポリペプチドは、HTLV-III感染およびHTLV-IIIに対する抗体を検出する免疫化学検定法において使用することができる。これらの方法はエイズを診断する際に役立つ。さらに、これらは血液成分(例えば血友病の治療のための第VIII因子)の輸血または生産のために血液を使用するのに先立って、その血液をスクリーニングする際にも利用できる。スクリーニング技術の利用可能性はエイズ伝染の危険を減ずるであろう。

【0018】

ポリペプチドと反応する抗体の検出は多くの確立された方法によって実施できる。例えば、免疫反応性HTLV-IIIポリペプチドをポリスチレンビーズや他の固体支持体のような固相に付着させる。次にその固相をHTLV-IIIに対する抗体について試験しようとする血液試料とインキュベートする。適当なインキュベーション期間の経過後固相と血液試料とを分離する。固相に結合した抗体は、標識ポリペプチドまたはヒト免疫グロブリンに対する標識抗体を用いて検出できる。

40

【0019】

HTLV-IIIポリペプチドはエイズのワクチン予防にも使用される。このウイルスに対する予防接種のためには中和抗体を誘導する免疫原性ポリペプチドが利用されるであろう。その第一の候補はウイルスエンベロープポリペプチドである。

【0020】

これらのポリペプチドはHTLV-IIIポリペプチドに対する抗体(モノクローナル抗

50

体を含む)の生産にも使用できる。これらの抗体は体液(血液, 唾液, 精液など)中のウイルスを直接検出する免疫化学検定法において使用される。特異的H T L V - I I I 抗原決定基に対するモノクローナル抗体を使用する検定法は、偽陽性の結果を減少させることによってウイルス検定の正確さを改善するであろう。また、ウイルスに対する抗体は免疫療法にも利用でき、例えばウイルスに対する受動免疫を与えることができる。

【0021】

ポリペプチドの生産方法、ならびにこれらのポリペプチドに基づく診断方法も本発明の主題である。

【0022】

さらに、本発明は免疫反応性ポリペプチドをコードするH T L V - I I I の遺伝子の単離方法; これらの遺伝子のヌクレオチド塩基配列の同定方法; ウイルスR N Aを生産させるべく適当なベクター内へのウイルスD N Aに特異的なD N A配列の導入方法、およびD N Aプローブの作製方法を提供する。これらのプローブはH T L V - I I I D N Aに特異的な塩基配列で構成され、例えば血液のような体液中の相補的H T L V - I I I D N A配列を検定するのに有用である。

H T L V - I I I ポリペプチド

免疫反応性H T L V - I I I ポリペプチドをコードするH T L V - I I I D N Aのセグメントを単離するために、遺伝子操作法が使用される。これらのポリペプチドはエイズ患者からの血清またはH T L V - I I I に対する抗体と免疫反応性であり、これらにはコアタンパク質、H T L V - I I I D N Aの1.1 K b E c o R I制限断片によりコードされる15 K dのペプチド、およびエンベロープ糖タンパク質が含まれる。これらの方法はまたポリペプチドをコードする断片の塩基配列決定にも使用できる。宿主細胞D N A内に組み込まれたプロウイルス遺伝子は分子的にクローニングされ、そしてクローニングされたプロウイルスのヌクレオチド塩基配列を決定する。

【0023】

H T L V - I I I D N Aの大腸菌発現ライブラリーを作製する。H T L V - I I I ゲノムをクローニングし、その後制限酵素を用いてクローニングしたH T L V - I I I ゲノムを切断し、それによりD N A断片を得る。(図1および図2を参照されたい。)約200 ~ 500 b pのH T L V - I I I D N A断片をアガロースゲルから単離し、T₄ポリメラーゼで末端修復し、そしてリンカーD N Aに連結させる。リンカー連結D N Aを次に制限酵素で処理し、アガロースゲルから精製して発現ベクター内でクローニングする。使用する発現ベクターの例はo m p A, p I N (A, BおよびC)、ラムダp L, T7, l a c, T r p, O R Fおよびラムダg t 11である。さらに、p S V 28 p t, p S V 2 n e o, p S V d h f rおよびV P Vベクターのような哺乳動物細胞、ならびにG A L IおよびG A L 10のよう酵母ベクターも使用できる。

【0024】

細菌ベクターはl a cコーディング配列を含み、この中にβ - ガラクトシダーゼ融合タンパク質を生成させるためにH T L V - I I I D N Aを挿入する。次に、この組換えベクターを細菌(例えば大腸菌)内に導入する(H T L V - I I I D N Aを含むベクターを取り込んだこれらの細胞は形質転換されたと言われる)。その後、融合タンパク質を発現する形質転換細胞を同定するために、これらの細胞をスクリーニングする。例えば、細菌をマッコンキー(M a c C o n k e y)寒天平板に置いてクローンの表現型を確かめる。もしβ - ガラクトシダーゼが生産されているならば、コロニーが赤く見えるだろう。

【0025】

また、細菌コロニーはH T L V - I I I D N Aプローブを用いてスクリーニングすることにより、興味あるD N A領域(例えば、H T L V - I I I g a g, p o lおよびe n v D N A配列)を含むクローンを同定できる。D N Aプローブを用いたスクリーニングにおいて陽性であり、かつマッコンキー寒天平板において陽性であるクローンを単離する。

【0026】

10

20

30

40

50

HTLV-III DNA配列を含む細胞の同定は、HTLV-III特異的抗体と免疫反応性であるHTLV-IIIポリペプチドの生産を可能にする。選択されたコロニーからの細胞は、ハイブリッドタンパク質を発現させる条件下に培養して増殖させる。その後、当分野で知られた方法により細胞タンパク質を得る。例えば、培養物を遠心分離し、得られた細胞沈殿物を破壊する。宿主細胞によって分泌されたポリペプチドを細胞培養物の上清から（細胞を破壊することなく）得ることもできる。

【0027】

全細胞タンパク質はSDSポリアクリルアミドゲル電気泳動を行うことにより分析する。他のタンパク質を全く含まないゲル上の一箇所で融合タンパク質を同定する。また、ウェスタンブロット分析が陽性のクローンに対して行われる。このような分析はエイズ患者からの血清を用いて行われ、その結果HTLV-III特異的抗体と交差反応するHTLV-III-β-ガラクトシダーゼ融合タンパク質（抗原）を発現するクローンを同定することが可能になる。

10

【0028】

HTLV-III DNAを含むラムダ₁₀クローンはウイルスの複製型からクローニングされる。レトロウイルスが複製するとき、二重鎖DNAが生産される。クローニングされたHTLV-III DNAは制限酵素SstIで消化する。（図1・a参照）HTLV-III DNAのLTR内に2つのSstI認識部位が存在するので、ラムダ₁₀ベクターから取り出されたクローン化DNA配列には一方のLTR領域が存在しない。その結果、HTLV-III DNAの小さな（約200bp）断片が欠失する。

20

【0029】

得られたDNAを線状化し、env遺伝子領域を含む線状ゲノムDNAを制限酵素で消化していくつかの断片を得る。例えば、図1・bに示すようにKpnIまたはEcoRIプラスHindIIIを使って断片をつくる。得られた2.3Kb KpnI-KpnI断片；1.0Kb EcoRI-EcoRI断片；および2.4Kb EcoRI-HindIII断片はゲル電気泳動および電気溶出により単離する。これらの断片をランダムにせん断してさらに小さい断片をつくる。こうして得られた断片はアガロースゲルから分離し、そして約200～500bpのDNA断片を溶出する。

【0030】

溶出された200～500bp DNA断片は大腸菌T₄ポリメラーゼの使用により末端修復し、そして平滑末端を読み取り枠発現（ORF）ベクター（例えばpMR100）に連結させる。この連結はpMR100ベクターのSmaI部位で起こり、pMR100ベクターは2つのプロモーター領域、ラムダCI遺伝子のハイブリッドコーディング配列およびlacI-lacZ遺伝子融合配列を含む。このベクターにおいて、これらは読み取り枠の塩基配列からはずれており、その結果このベクターは非生産性である。HTLV-III DNAをこのベクターに挿入することにより、正しいDNA断片が読み取り枠を修正し、その結果CI-HTLV-III-β-ガラクトシダーゼ融合タンパク質が生産されるだろう。このハイブリッドの発現はlacプロモーターの制御下にある。pMR100の塩基配列に基づくと、SmaI部位にクローニングされたDNA断片挿入物がラムダCI遺伝子断片とlac-Z断片との間に適当な読み取り枠を生ずる場合、その挿入DNAはラムダCI遺伝子の枠により定められる読み取り枠に停止コドンを含むべきでないと考えられる。

30

40

【0031】

次いで、組換えpMR100ベクターを大腸菌の中に導入する。細菌はマッコンキー寒天平板上に置いてクローンの表現型を確かめる。β-ガラクトシダーゼが生産されている場合は、コロニーが赤く見えるだろう。また、DNA挿入物を含むクローンを同定するために、HTLV-III DNAプローブを用いてコロニーをスクリーニングする。DNAプローブを用いてスクリーニングしたとき陽性でありかつマッコンキー寒天平板上で陽性であるクローンを単離する。

【0032】

50

選択されたコロニーからの細胞を培養して増殖させる。この培養物は遠心分離にかけ、そして細胞沈殿物を破壊する。全細胞タンパク質はSDSポリアクリルアミドゲルによる電気泳動を行って分析する。融合タンパク質は他のタンパク質を全く含まないゲルの上の一箇所で同定される。(図15参照)

ウエスタンブロット分析もスクリーニングしたとき陽性であったクローンに対して行われる。エイズ患者からの血清が使用され、こうしてHTLV-III特異的抗体と交差反応するHTLV-III-β-ガラクトシダーゼ融合タンパク質を発現するクローンを同定することが可能になる。この方法によって1000個のクローンがスクリーニングされ、そして6個のクローンが陽性であった。

【0033】

pMR100クローニングベクターの性質ゆえに、生産性DNA挿入物は大きな融合ポリペプチドの一部として発現されるべきである。HTLV-III env遺伝子を含む組換えクローンは、コロニーハイブリダイゼーションによって同定した。β-ガラクトシダーゼ活性を有する大きな融合ポリペプチドの生産は、マッコンキー寒天平板上での表現型同定、β-ガラクトシダーゼ酵素検定および75% SDS-ポリアクリルアミドゲルの分析によって確かめた。この大きな融合タンパク質とHTLV-IIIに対する抗体との免疫反応は、エイズ患者からの血清を用いるウエスタンブロット分析によって評価した。また、大きな融合タンパク質は抗β-ガラクトシダーゼおよび抗CI抗血清とも反応した。この発見はそれらがCI-HTLV-III-lacZのタンパク質であるという仮説と一致する。

【0034】

HTLV-IIIの読み取り枠挿入物断片は、さらにDNA塩基配列決定によって分析する。pMR100のSmaIクローニング部位の両側にある2つのBamHI部位のうち一方はクローニング段階で破壊されるので、陽性クローンはHindIIIおよびClaI制限酵素で消化して挿入HTLV-III DNA断片を遊離させる。融合組換え体からHTLV-III ORF挿入物を単離し、そしてHindIIIおよびAccIで消化したM13クローニングベクターmp18およびmp19内でクローニングする。その後、陽性クローンのDNA塩基配列を決定する。

【0035】

約200~500bpのHTLV-III DNA断片をアガロースゲルから単離し、T₄ポリメラーゼで末端修復し、そしてEcoRIリンカーへ連結させる。次にEcoRIリンカーへ連結させたDNAをEcoRIで処理し、1%アガロースゲルから精製して発現ベクターラムダgt11内でクローニングする。このベクターはlacZ遺伝子コーディング配列を含み、このベクター内にβ-ガラクトシダーゼ融合タンパク質を生成させるべく外来性DNAを挿入することができる。ハイブリッド遺伝子の発現はlacリプレッサーの制御下にある。lacリプレッサー遺伝子lacIは宿主細胞の大腸菌Y1090内の別のプラスミドpMC9によって運ばれる。1.5×10⁴の組換えファージを含むHTLV-IIIゲノムDNAのラムダgt11ライブラリーを調べるためにエイズ患者の血液を使用した。5000の組換え体のスクリーニングにおいて、強力なシグナルを発する100の独立クローンを単離した。陽性の組換えDNAクローンはさらにそれらの特定の遺伝子発現について明らかにされた。gag遺伝子特異的クローンを同定するために、P24に対するウサギ高度免疫血清を使用した。特定のHTLV-III遺伝子(詳細にはgag遺伝子, env遺伝子およびP_x遺伝子)のニックトランスレーションされたDNAプローブは、陽性の免疫反応性クローンを特定の遺伝子領域にグループ分けするのに使用した。

【0036】

エイズ血清によって強力なシグナルを発しかつHTLV-IIIのgag, pol, sorおよびenv-lor遺伝子領域を有する挿入物DNAを含む組換えクローンは、制限酵素によるマッピングおよびDNA塩基配列決定によって詳細に調べた。

HTLV-III DNAのヌクレオチド配列決定

10

20

30

40

50

HTLV-III DNAのヌクレオチド配列決定には遺伝子操作法を使用した。この配列決定に使用することができる1つの技術は、ショットガン/ランダム配列決定法である。HTLV-III DNAを約300~500bpの大きさの断片にランダムにせん断する。これらの断片は例えばM13を使ってクローニングし、次いでコロニーはHTLV-III DNA断片挿入物を含むものを同定するためにスクリーニングする。その後塩基配列に重複部分が生ずるような多数の分析を行って、そのヌクレオチド配列を決定する。HTLV-III DNAの両鎖の塩基配列を決定してその向きを定める。制限酵素マッピングを用いて得られた塩基配列決定データを調べる。

【0037】

1つのクローニングされたHTLV-IIIゲノム(BH10)のヌクレオチド配列を図3~図14に示す。ここにはgagタンパク質p17およびgagp24のN末端およびgagp15のC末端(これはpolタンパク質のN末端と重なり合う)をコードする塩基配列位置が示されている。また、pol, sorおよびenv-lorの読み取り枠(ORF)も示される。クローンBH10(Rの一部, U5, tRNAプライマー結合部位およびリーダー配列の一部を含む)に存在しないHTLV-III DNAの残りの182塩基対の配列はクローンHXB2から誘導された。別の2つのクローン(BH8およびBH5)の塩基配列も示される。制限酵素部位はヌクレオチド配列の上に示し、クローンBH8に存在するがクローンBH10に存在しない部位はカッコ内に示す。ヌクレオチド251, 254, 5671および6987-7001に決失が見られる(〔 〕)。ヌクレオチドの位置(各線の右側)は転写開始部位からスタートしている。アミノ酸残基は4つの最も大きな読み取り枠に対して番号が付けられており(各線の右側)、各々の場合先行する終止コドンの後からスタートするが、gagの場合は最初のメチオニンコドンから数えている。提案されたペプチド切断部位(V)および起こりうるアスパラギン-結合グリコシル化部位(*)をenv-lor読み取り枠に対して示す。図3~図14の最初に示したクローンBH8およびBH10から誘導されるLTRの塩基配列は各クローンの3'部分から誘導され、これらのウイルスゲノムの組み込まれたコピーの5'-LTRに存在するものと同一であると推定される。

【0038】

クローンHXB2はラムダJ1内でクローニングされたHTLV-III感染H9細胞からのXbaI消化DNAの組換えファージライブラリーから誘導された。H9細胞はエイズ患者の血液から得られたHTLV-IIIのプールによって感染させたヒト白血病細胞である(F. ワング-スタール(Wong-Staal)のネイチャー, 312, 1984年11月を参照されたい)。クローニングベクタークローンBH10, BH8およびBH5はランダムgtWees. ラムダB内でクローニングされたHTLV-III感染H9細胞のHirt上清分画からのSstI消化DNAのライブラリーより誘導された。両方のライブラリーはプライマーとしてオリゴdTを使ってウイルス粒子RNAから合成されたcDNAプローブを用いてスクリーニングした。クローンBH8, BH5およびHXB2の一部はマクサム-ギルバード法(A. M. MaxamおよびGilbert, Methods in Enzymology, 65: 499-560 (1980)を参照)によって塩基配列を決定した。クローンBH10はプライマーとしてM13挿入物の塩基配列に相補的なオリゴヌクレオチドを使用しかつポリメラーゼとしてDNAポリメラーゼIのクレノウ(Klenow)断片または逆転写酵素を使用することにより変更したサンガー(Sanger)の方法によって塩基配列を決定した。

HTLV-IIIに特異的なRNA, RNAプローブおよびDNAプローブの作製

HTLV-IIIからの全遺伝子または遺伝子セグメントであるDNA配列をT7ベクターのようなベクターに挿入する。この具体例は、ベクターがT細胞遺伝子10プロモーターからのTceupロモーターおよびT細胞遺伝子10タンパク質からの11個のアミノ酸をコードするDNA配列を含む。

【0039】

次に、このベクターを用いて大腸菌のような細胞を形質転換する。T7ベクターはT7ポリ

10

20

30

40

50

メラゼを利用する。T7ポリメラーゼはRNA形成を触媒し、RNAポリメラーゼが転写開始のために結合する部位であるT7プロモーターのみを認識する。このT7ポリメラーゼは大腸菌プロモーターを認識しない。その結果、HTLV-II DNA配列をT7ベクターのプロモーターおよびポリメラーゼ遺伝子の後に挿入し、かつターミネーターをHTLV-II DNA配列の直後に配置する場合、T7ベクターはHTLV-II DNA挿入物に相補的なRNAを直接製造するのである。

【0040】

また、HTLV-II DNAのヌクレオチド配列の決定は、DNAプローブを作るための基礎を提供する。RNAプローブとDNAプローブは共に、体液中のHTLV-IIの検出に利用できるようにHTLV-IIゲノムの特定領域を含まねばならない。HTLV-IIゲノムとHTLV-Iおよび-IIゲノムとの間の相同は比較的少なく、プローブはHTLV-IIに独特の領域（すなわちHTLV-Iまたは-IIと共有しない領域）を含む。例えば、HTLV-IIのenv遺伝子領域のヌクレオチド配列が使用される。

10

【0041】

ウイルスRNAまたはDNAは、非常に高い濃度のウイルスを含むことが知られている。例えば唾液中のHTLV-IIを検出するのに用いられる。これは、例えば唾液試料を変性し、これを濾紙に移し、どちらかの型のプローブを用いてスクリーニングするドットプロット（dot blot）の手段によって行うことができる。試験体液として唾液を使用する場合、HTLV-IIの検出は血液を試験する場合よりも迅速かつ容易であると考えられる。

20

HTLV-IIポリペプチドと反応するモノクローナル抗体の生産

HTLV-IIポリペプチドと反応するモノクローナル抗体は、抗体産生細胞株によって生産される。抗体産生細胞株は普通ハイブリドーマとして知られるハイブリッド細胞株でありうる。このハイブリッド細胞は、HTLV-IIポリペプチドに対する抗体を生産する細胞と、不滅性細胞（すなわちハイブリッド細胞に組織培養の長期安定性を付与する細胞）との融合によって作られる。ハイブリッド細胞株を作る場合、その第一の融合パートナー（抗体産生細胞）はHTLV-IIポリペプチドに対して免疫化した動物の脾臓細胞でありうる。これとは別に、抗体産生細胞はHTLV-II抗原に対する抗体を生産するところの単離されたBリンパ球でありうる。このリンパ球は脾臓、末梢血液、リンパ節またはその組織から得ることができる。第二の融合パートナー（不滅細胞）はリンパ芽細胞または形質細胞腫細胞（例えば、それ自身抗体産生能を有する悪性の骨髓腫細胞）でありうる。

30

【0042】

HTLV-IIポリペプチドに対するモノクローナル抗体を生産するネズミハイブリドーマは、マウス骨髓腫細胞とポリペプチドに対して免疫化したマウスからの脾臓細胞との融合によって作られる。マウスを免疫化するために、種々の異なる免疫感作方法を行うことができる。例えば、マウスは精製ポリペプチドの一次および二次免疫感作を受ける。融合は標準方法において行われる。コーラー（Kohler）およびミルスタイン（Milstein）のネイチャー（ロンドン）256, 495-497（1975）；R. ケネット（Kennet）のモノクローナル抗体（ケネットら編集，365-367頁，プレナム・プレス，ニューヨーク，1980年）を参照されたい。

40

【0043】

次いで、ハイブリドーマをポリペプチドと反応する抗体の産生についてスクリーニングする。これは当分野で知られたスクリーニング方法により行われる。

【0044】

抗体産生細胞株を作る別の方法は、抗体産生細胞の形質転換による。例えば、HTLV-IIポリペプチドに対して免疫化した動物から得られたBリンパ球を、ヒトBリンパ球の場合にはエプスタイン-バールウイルス（Epstein-Barr virus）のようなウイルスを感染させて形質転換し、それにより不滅性抗体産生細胞を得る。例えば

50

、コズボー (K o z b o r) およびロドー (R o d o r) の Immunology Today 4 (3) , 72 - 79 (1983) を参照されたい。これとは別に、形質転換遺伝子または形質転換遺伝子産物を用いてBリンパ球を形質転換してもよい。

【 0 0 4 5 】

H T L V - I I I ポリペプチドに対するモノクローナル抗体は、抗体産生ハイブリドーマをマウスの腹腔内に注入し、適当な時間の経過後非常に高力価の均一抗体を含む腹水を採取し、そして腹水からモノクローナル抗体を単離することにより大量に生産することができる。異種ハイブリドーマは照射マウスまたは胸腺欠損ヌードマウスに注入すべきである。また、H T L V - I I I ポリペプチドを生産する細胞をインビトロで培養し、その細胞培養基から分泌されたモノクローナル抗体を単離する方法によっても抗体を得ることができる。これらの方法により産生された抗体は診断検定法（例えば体液中のH T L V - I I I を検出する）および受動免疫療法において使用できる。H T L V - I I I ポリペプチドと反応する抗体は、エイズまたは体液（例えば血液、精液、唾液）中のH T L V - I I I の存在を検出するための診断試験、ならびに受動免疫療法のための基礎を提供する。例えば、抗 p 4 1 を生産し、慣用技術によりこれを固相に付着させ、そして試験すべき体液を固定化抗体と接触させることが可能である。このようにして、体液中のH T L V - I I I （抗原）を検出することができ、この方法はH T L V - I I I に対する抗体を検出する試験よりもはるかに少ない偽陽性の試験結果をもたらすだろう。

【 0 0 4 6 】

本発明はさらに次の実施例により説明されるであろう。

実施例 1

超音波処理されたDNA断片の作製

ゲルにより精製したH T L V - I I I 制限酵素断片 10 μ g を超音波処理して、平均 500 b p の大きさに断片化した。超音波処理後、容量を減らすためにDNAを 0.1 $\times$ T B E 中のD E A E - セルロースカラムに通した。D E A E 結合DNAは 0.2 M N a C l - T E (2 M N a C l , 10 m m トリス - H C l p H 7.5、1 m M E D T A) 5 m l で洗い、次に 1 M N a C l - T E で溶出してエタノール沈殿させた。超音波処理したDNAの大きさを 1.2 % アガロースゲルにより測定した。所望の長さ (200 ~ 500 b p) のDNA断片をゲルから溶出した。T 4 DNA ポリメラーゼを使って、超音波処理法によって生じた一本鎖DNA末端を修復および/または切り取った。DNA断片をT 4 ポリメラーゼと共にヌクレオチドを添加することなく 37 ° で 5 分間インキュベートして 3' 末端からヌクレオチドを除き、次に 4 種のヌクレオチド前駆体を最終濃度が 100 μ M となるように加え、この反応混合物をさらに 30 分間インキュベートして 5' 末端の一本鎖のつき出た部分を修復した。68 ° で 10 分間酵素を熱不活化することによりこの反応を停止させた。DNAは 1 度フェノール抽出し、エタノール沈殿させ、そしてT E 中に再懸濁した。

実施例 2

ランダムにせん断したDNA断片のクローニング

超音波処理しかつ平滑末端に修復したH T L V - I I I DNA断片を、O R F 発現ベクター p M R 100 のS m a I 部位に連結し、そして標準形質転換法を使って宿主細胞 L G 90 を形質転換した。形質転換細胞の β - ガラクトシダーゼ陽性表現型は、その形質転換細胞をアンピシリン (25 μ g / m l) 含有マッコンキー寒天平板上に採置して、表現型を 37 ° で 20 時間後に評価することによって同定した。

実施例 3

ハイブリッドタンパク質分析

アンピシリン (25 μ g / m l) 含有L ブイヨン中で一晚増殖させた飽和培養物からの細胞 10 m l 試料を遠心分離し、細胞沈殿物を 1.2 倍に濃縮したレムリ (L a e m m l i) サンプル緩衝液 500 μ l に再懸濁した。細胞は渦巻混合し、100 ° で 3 分間沸騰させることにより再懸濁した。次に、この溶菌液を 22 ゲージ針を使って繰り返しせん断し、溶菌液の粘度を低下させた。タンパク質試料約 10 μ l は 7.5 % S D S - P A G E

10

20

30

40

50

(S D S - ポリアクリルアミド) ゲルによる電気泳動処理を行った。

【 0 0 4 7 】

タウビン (T o w b i n) らの方法に従って、 S D S - P A G E ゲルからニトロセルロース紙へタンパク質を移行させた。移行後、このフィルターを 0 . 1 % 消泡剤 A および 0 . 0 0 0 1 % メルチオレート (m e r t h i o l a t e) を含有する P B S 中の 5 % (w / v) 脱脂牛乳の溶液中で 3 7 °C において 2 時間インキュベートし、それにより全ての利用しうるタンパク質結合部位を飽和させた。エイズ抗血清との反応は、大腸菌溶菌液をあらかじめ吸収させた 1 % エイズ患者抗血清含有の上記牛乳緩衝液中で行った。この反応は回転攪拌機上の密閉プラスチック袋中で 4 °C において 1 8 ~ 2 4 時間実施した。このインキュベーション後、 0 . 5 % デオキシコール酸、 0 . 1 M N a C l 、 0 . 5 % トリトン X - 1 0 0 、 1 0 m m 燐酸塩緩衝液 p H 7 . 5 および 0 . 1 m M P M S F を含有する溶液で室温において 2 0 分づつ 3 回フィルターを洗浄した。

10

【 0 0 4 8 】

抗原 - 抗体反応を視覚化するために、¹²⁵I でヨウ素化した第二のヤギ抗ヒト抗体とニトロセルロースとをインキュベートした。ヨウ素化抗体との反応は、第一の抗体のときに使用したものと同一牛乳緩衝液中室温で 3 0 分間実施した。その後先に述べたようにしてニトロセルロースを洗浄し、コダック X A R 5 フィルムおよび増感スクリーンを使って - 7 0 °C で感光した。

実施例 4

コロニーハイブリダイゼーションによる H T L V - I I I O R F ライブラリーのスクリーニング

20

大腸菌 L G 9 0 形質転換体は意図する D N A 領域 (例えば H T L V - I I I g a g , e n v または P x 遺伝子の特定配列) を含む H T L V - I I I D N A プローブを用いてスクリーニングした。ニトロセルロースフィルター上でコロニーを増殖させ、ハイブリダイゼーションプローブとしてニックトランスレーションされた H T L V - I I I D N A を使い、グルンスタイン (G r u n s t e i n) およびホグネス (H o g n e s s) の方法に従ってスクリーニングした。

【 0 0 4 9 】

一般には、制限エンドヌクレアーゼ消化によって D N A 断片を切り取り、これをゲル精製し、そしてニックトランスレーションにより 0 . 5 × 1 0 ⁸ c p m / μ g の比活性に ³²P で標識した。〔 P . W . J . リグビー (R i g b y) らの J . M o l . B i o l . , 1 1 3 , 2 3 7 (1 9 7 7) を参照〕 D N A が固定されたニトロセルロースフィルターは 6 × S S C (0 . 9 M N a C l / 0 . 0 9 M クエン酸ナトリウム、p H 7 . 0) 、 5 × デンハート溶液 (D e n h a r d t ' s s o l u t i o n : ポリビニルピロリドン、フィコルおよびウシ血清アルブミン各々 0 . 0 2 %) 、 1 0 μ g / m l の変性した超音波処理大腸菌 D N A を用いて 5 5 °C で 3 ~ 5 時間プレハイブリダイゼーションを行った。次いで、変性ハイブリダイゼーションプローブを加えた同じ溶液の新しい試料中にこのフィルターを入れた。ハイブリダイゼーションは 6 8 °C で 1 6 時間行った。その後フィルターを 5 5 °C において 0 . 3 × S S C で繰り返し洗浄し、X 線フィルムを使用して感光した。

30

実施例 5

40

エイズ患者からの血清と免疫反応する H T L V - I I I の組換え D N A によって生産されたペプチド

発現ベクターとして p I N - I I I - o m p A (o m p A) を使用した。 o m p A はリボタンパク質 (大腸菌中で最も豊富なタンパク質) 遺伝子プロモーター (l p p) および l a c U V 5 プロモーター - オペレーターを有する (図 1 6 参照) 。 o m p A ベクターはさらに l a c リプレッサーをコードする D N A セグメントを含み、これは挿入 D N A の発現を I P T G のような l a c オペロン誘導物質により調節させる。 o m p A クローニングベクターは 3 つの読み取り枠の全てに 3 つの特異な制限酵素部位 E c o R I 、 H i n d I I I および B a m H I を含み、そしてこれらの制限酵素部位のいずれかに D N A が挿入される。

50

【 0 0 5 0 】

ベクターラムダ g t W E S ラムダ B の S s t I 部位に 9 K b の長い H T L V - I I I D N A 挿入物を含む組換えクローンのラムダ B H 1 0 から、種々の制限断片を切り取った。次に、これらの制限断片を 3 つの読み取り枠の全てで o m p A ベクターに挿入し、そして大腸菌 J A 2 2 1 細胞を形質転換するのに使用した。形質転換細胞は、ニックトランスレーションされた H T L V - I I I D N A プローブを使用する i n s i t u コロニーハイブリダイゼーションにより、H T L V - I I I D N A についてスクリーニングした。陽性クローンは H T L V - I I I に特異的な抗体を使用して H T L V - I I I 抗原ペプチドの発現についてスクリーニングした。このために、H T L V - I I I D N A 組換えプラスミドを含む大腸菌細胞の溶菌液を 1 2 . 5 % S D S - ポリアクリルアミドゲルによる電気泳動にかけ、そしてニトロセルロースフィルター上へ移行した。その後、このフィルターを最初にエイズ患者からの十分な性状決定された血清とインキュベートし、次に^{1 2 5}I で標識したヤギ抗ヒト I g G 抗体とインキュベートした。洗浄したフィルターをオートラジオグラフィーにかけ、抗 H T L V - I I I 抗体と反応性のペプチドを同定した。

10

【 0 0 5 1 】

抗 H T L V - I I I 抗体と免疫反応性であるペプチドをコードするいくつかの遺伝子セグメントが見出された。これらのうちの 1 つに 1 . 1 K b の E c o R I 制限断片がある。この断片は 3 つの読み取り枠の全てで o m p A ベクター内に挿入した (図 1 6 参照) 。細胞は 1 0 0 m g / m l のアンピシリンを含む L ブイヨン中 3 7 で O D₆₀₀ = 0 . 2 へと増殖させた。この時点で細胞培養物を 2 つのアリコートに分割した。一方のアリコートに最終濃度が 2 m M となるように I P T G を加えた (誘導) 。他方のアリコートには I P T G を加えなかった (非誘導) 。 I P T G 誘導の際に、3 つのプラスミド構成物 [O m p A₁ - R - 6 (O 1 R 6) , O m p A₂ - R - 7 (O 2 R 7) および O m p A₃ - R - 3 (O 3 R 3) と命名される] の全ての形質転換細胞は、エイズ患者の血清中の抗 H T L V - I I I 抗体と強く反応する 1 5 K d のペプチドを生産した (図 1 7 参照、レーン 1 は精製した H T L V - I I I ウイルス粒子 ; レーン 2 および 3 は非誘導および誘導の O 1 R 6 ; レーン 4 および 5 は非誘導および誘導の O 2 R 7 ; レーン 6 および 7 は非誘導および誘導の O 3 R 3 である) 。この反応性は正常人から採取した血清を用いた場合検出されない。

20

【 0 0 5 2 】

H T L V - I I I ゲノムの D N A 配列データは、E c o R I 断片の 5 ' 末端に位置する p o l 遺伝子内に 1 つの読み取り枠が存在することを示す。3 つの組換え構成物 O 1 R 6 , O 2 R 7 および O 3 R 3 の D N A 塩基配列決定は、これらの組換え体の各々が各ベクターのコーディング配列に結合した H T L V - I I I プラス D N A 鎖の異なる読み取り枠を有するということを証明した。O 3 R 3 の場合のみ、挿入 D N A の読み取り枠が O m p A ベクター内のシグナルペプチドによって定められるそれと一致するが、O 1 R 6 および O 2 R 7 の場合 p o l 遺伝子セグメント D N A の読み取り枠がベクターにより定められるそれと不一致である。(図 1 8 参照)

30

ヌクレオチド位置 2 4 - 2 9 には 6 b p のリボソーム結合部位 A A G G A G (シャイン - ダルガルノの配列、S h i n e - D a l g a r n o s e q u e n c e) があり、これにより 1 1 b p 下流 (位置 4 1 - 4 3) には開始コドン A T G が存在する。3 つの組換え体の全てによって合成される 1 5 K d のペプチドは、この内部開始コドンを使用して転写物から翻訳されると考えられる。このことが真実である場合には、ペプチドは位置 4 1 - 4 3 に存在する A T G から出発して、位置 4 4 6 - 4 4 8 に存在する停止コドンで終り、それにより H T L V - I I I の p o l 遺伝子の 3 ' 末端セグメントによりコードされる 1 3 5 アミノ酸残基から成るペプチドが生産される。

40

【 0 0 5 3 】

H T L V - I I I D N A p o l 遺伝子の読み取り枠がベクターにより定められるそれと一致する O 3 R 3 構成物は、1 5 K d ペプチドの他に、1 9 K d と 1 6 . 5 K d の大きさの 2 つの別のペプチドを生産した (図 1 7 参照) 。1 9 K d ペプチドはさらに 3 5 個のアミノ酸残基を含み、そのうちの 2 1 個は O m p A₃ ベクターによってコードされるシグナ

50

ルペプチドからのものであり、残りの14個は挿入HTLV-III DNAそれ自体によってコードされると思われる。16.5 Kdペプチドはプロセッシングを受けた19 Kdペプチド(シグナルペプチドが開裂している)でありうる。

【0054】

O1R6およびO2R7構成物もエイズ患者の血清と弱く反応する約17.5 Kdの別のペプチドを生産する(図17参照)。このペプチドの開始点は明らかでない。1.1 KbのEcoRI断片は、ヌクレオチド位置360-965の間に延びるSOR(短い読み取り枠)と呼ばれる第二のコーディング領域を含む(図16参照)。この領域内の5つのAUGメチオニンコドンのうち4つがこの読み取り枠の5'末端の近くに存在する。このDNAセグメントは192, 185, 177または164個のアミノ酸残基から成るペプチドをコードできるだろう。しかしながら、この読み取り枠の5'末端にははっきりと認めうるリボソーム結合部位が存在しない。

10

【0055】

さらに、15 Kdペプチドは実際にpol遺伝子から誘導されるという決定が多くの証拠により支持された。第一に、O1R6、O2R7およびO3R3(図16)からの1.1 Kb EcoRI挿入物からの3'末端StuI/EcoRI断片の欠失は15 Kdペプチドの合成に影響を与えない。第二に、5'末端EcoRI/NdeI断片のみを含むクローンがまだ同じ15 Kdペプチドを生産する。最後に、読み取り枠クローニングベクターpMR100内に適切に挿入されたSORコーディング配列をもつ種々のDNA断片を含む組換えクローンは、エイズ患者の血清中に存在する抗HTLV-III抗体との免疫反応性が非常に乏しいラムダCI-HTLV-III-β-ガラクトシダーゼ3分節融合タンパク質を生産した。

20

【0056】

エイズ患者の血清中でウイルスpol遺伝子から誘導された15 Kdペプチドに対する有意な免疫反応性を検出した。この免疫反応性ペプチドの特性は、SDS-ポリアクリルアミドゲル電気泳動におけるHTLV-IIIウイルス粒子抗原の結合パターンに関して、拮抗阻害免疫検定法を用いて決定した。精製HTLV-IIIウイルス粒子をSDSで処理し、電気泳動にかけ、そしてニトロセルロースフィルターの上に移した。破壊したHTLV-IIIウイルス粒子を含む同一のフィルター細片を、O1R6、O2R7または対照細菌クローンの溶菌液の存在下または不存在下に、十分に性状決定がなされたエイズ患者の血清とインキュベートした。エイズ患者の血清に含まれる抗HTLV-III抗体とフィルターにしみ込ませたウイルス粒子のタンパク質との間の特異的免疫反応は、¹²⁵Iで標識したヤギ抗ヒト抗体によって明らかにした。図19に示すように、O1R6の溶菌液はウイルスp31タンパク質とエイズ血清との免疫反応を阻害するが、対照細胞の溶菌液は阻害しない。この結果は、ウイルスpol遺伝子の3'末端によってコードされる組換え15 Kdタンパク質が他のウイルス粒子タンパク質p31の一部であるということを示唆しており、p31がHTLV-IIIウイルス粒子と一緒に同時精製される細胞タンパク質であるというある人々によって共有された考え方と対照をなしている。

30

【0057】

15 Kdペプチドに対する抗体がエイズ患者の血清中に広く含まれるかについても評価した。抗原の源としてO1R6の溶菌液を使用するウェスタンブロット分析において、エイズ患者と正常人々からのコード化血清のパネルを試験した。20のエイズ血清の全てが15 Kdペプチドと反応し、8の正常対照はどれも15 Kdペプチドと反応しなかった。代表的な結果を図20に示す。これらのデータは、全部ではないにしても、大部分のエイズ患者がウイルスp31タンパク質に対する抗体を産生することを示している。

40

実施例 6

HTLV-IIIの読み取り枠遺伝子セグメントの大腸菌内での発現

HTLV-III DNAをラムダBH10から切り取った。このラムダBH10はベクターラムダgtWESラムダB内に挿入したHTLV-III DNAの9 Kbセグメントを含む先に作製した組換えラムダファージである。このHTLV-III DNAを超

50

音波処理し、約 0.5 Kb の DNA 断片をゲル電気泳動により精製し、末端修復し、そして ORF ベクター pMR100 の SmaI 部位に挿入した (図 21 参照)。このベクターはハイブリッドコーディング配列 [バクテリオファージラムダのラムダ C I 遺伝子の N 末端 (5' セグメント) が N 末端欠失 lacI Z 遺伝子 (3' セグメント) に融合される] を含む第二 DNA 断片に結合した細菌 lac プロモーター DNA セグメントを含む。SmaI クローニング部位を含む短いリンカー DNA 断片を、lacI Z コーディング DNA の上流でフレームシフト変異が起こるような方法で、これらの 2 つの断片の間に挿入した。その結果、pMR100 は lac⁻ 宿主大腸菌 LG90 の細胞内に導入したとき検出可能な β - ガラクトシダーゼ活性を示さない。SmaI クローニング部位に読み取り枠を含む外来性 DNA (この場合は HTLV - III DNA) を挿入すると、その挿入コーディング配列がラムダ C I リーダーおよび lacI Z 遺伝子の双方に関して正しい読み取り枠内にある場合、フレームシフト変異を逆転させることができる。形質転換細胞はマッコンキー平板上でスクリーニングして、β - ガラクトシダーゼ酵素活性をその場で発現する個々のクローンを検出した。

【0058】

スクリーニングされた 6000 のアンピシリン耐性形質転換細胞のうち、約 300 が β - ガラクトシダーゼ活性を発現するとわかった。^{3 2} P で標識し、ニックトランスレーションされた HTLV - III DNA をプローブとして使用するコロニーハイブリダイゼーションは、これら lac⁺ クローンの全てが HTLV - III DNA を含むことを明らかにした。lac⁺ クローンにおいて、pMR100 の SmaI 部位に挿入された HTLV - III 断片はラムダ C I リーダーセグメントによって定められる読み取り枠に停止コードンを含んではならず、また lacI Z 遺伝子も正しい翻訳読み取り枠で存在しなければならない。3 要素融合遺伝子は、N 末端にラムダ C I タンパク質の一部をもち、中間に HTLV - III セグメントをもち、そして C 末端に lacI Z ポリペプチドをもつ 3 分節融合タンパク質として発現された。

【0059】

lac⁺ クローンによって生産されたタンパク質は、ラムダ C I - β - ガラクトシダーゼ融合タンパク質を生産する対照 lac⁺ クローン pMR200 の溶菌液と共に 7.5 % SDS - ポリアクリルアミドゲル上でその溶菌液を分割することにより分析した。pMR200 の lacI Z 遺伝子は、pMR100 の lacI Z 遺伝子が活性 β - ガラクトシダーゼを生産するためにラムダ C I 遺伝子と同じ読み取り枠になるべく 1 個の塩基対を欠失している点を除いて pMR100 のそれと同じである。β - ガラクトシダーゼおよびその融合タンパク質の非常に大きな寸法に基づいて、それらは SDS - ポリアクリルアミドゲル上で溶菌液中の大部分のタンパク質から分離され、そして図 22 . A に示すようにクーマシーブリリアントブルー染色によって簡単に同定することができる。HTLV - III DNA を含む lac⁺ クローンのいくつかは、ラムダ C I - lacI Z 融合タンパク質よりも大きな (15000 ~ 27000 ダルトン) ポリペプチドを生産する。これらの発見は DNA 挿入物の長さが 700 bp 以下であるというデータと一致する。β - ガラクトシダーゼ融合タンパク質は全細胞タンパク質の約 1 ~ 2 % を占めた。

【0060】

lac⁺ クローンにより生産されたペプチドは、エイズ患者の血清との免疫反応についてウエスタンブロット分析により試験した。lac⁺ クローンの溶菌液を SDS - ポリアクリルアミドゲルで電気泳動処理した後、それらをニトロセルロースフィルターへ移行させた。これらのタンパク質ブロットを最初にエイズ患者の血清と反応させ、次に ^{1 2 5} I で標識したヤギ抗ヒト IgG と反応させた。図 22 . B のオートラジオグラフは代表的な融合タンパク質とエイズ患者血清との免疫反応性を示す。組換えペプチドはこの他に抗 β - ガラクトシダーゼ抗血清とも反応し、それらが一般構造式ラムダ C I - HTLV - III ペプチド - lacI Z を有するという提案と一致した。HTLV - III DNA 挿入物を含まない陰性対照の pMR100 および pMR200 の免疫反応パターンから、この特別のエイズ血清は宿主大腸菌のいくつかの細菌タンパク質と反応する抗体を含むことが

10

20

30

40

50

明白である。このことはエイズ患者が通常多数の細菌に感染していることからして驚くにあたらない。大腸菌抽出物を結合させたセファロース 4 B を用いてエイズ患者の血清を吸着させると、バックグランド免疫反応がある程度減少するが、それは完全には消失しなかった。

【0061】

約 300 の独立した HTLV - III DNA 含有 Lac⁺ コロニーは、クーマシーブリリアントブルー染色およびウエスタンブロッティングを使用して SDS - ポリアクリルアミドゲルにより分析した。それらのうち約半数は長さが 300 ~ 600 bp の DNA 挿入物に対応して、約 100 ~ 200 個のアミノ酸からなる付加的ペプチドを含む融合タンパク質を発現することがわかった。これらの融合タンパク質のうち 20 はエイズ患者からの血清と特異的に反応することが見出された。非反応性クローンは恐らく、抗体と反応しないような状態に折り重なったペプチドを含むか、あるいはエイズ患者の体内で免疫原性がない HTLV - III タンパク質分子の領域に相当するペプチドを含むであろう。Lac⁺ クロンの残りの半数は、大きさがラムダ CI - β - ガラクトシダーゼタンパク質と明らかに異なる融合タンパク質を発現した。このグループの融合タンパク質からはエイズ患者の血清と反応するものが全く見出せなかった。

【0062】

Lac⁺ ORF クローンからの HTLV - III DNA 挿入物は、サザンブロッティング法を使って HTLV - III ゲノムの特定セグメントへマッピングされた。これらの実験において、各プラスミドクローンはニクトランレーションにより^{3 2}P で標識し、そして一群の HTLV - III DNA 制限断片とハイブリダイゼーションを行った。このハイブリダイゼーション分析は、DNA 塩基配列決定データと一致する ORF - A, ORF - B, ORF - C および ORF - D (図 2 . A 参照) と名づけた 4 つの読み取り枠セグメントへと全ての Lac⁺ ORF クローンをマッピングした。gag および pol 遺伝子のコーディング領域に相当する読み取り枠 ORF - A および - B は、それぞれ 1 . 5 Kb および 3 . 0 Kb の長さである。ORF - C は約 0 . 6 Kb の長さであり、ORF - B 領域とわずかに重なっており、そして 21 Kd のポリペプチドをコードすることができる。ORF - C の位置および pol 遺伝子とそれとの重複部分は、HTLV - I および - II の env 遺伝子構造を暗示する。しかしながら、sor (short open reading frame, 短い読み取り枠) と命名した ORF - C は完全なエンベロープタンパク質をコードするにはあまりに短かすぎる。第四の読み取り枠の ORF - D はその長さが 2 . 5 Kb であり、主要エンベロープ糖タンパク質の大型前駆体と 3' 末端から誘導される別のタンパク質 (HTLV - I および - II の lor 産物に類似している) との両方をコードできるだろう。env - lor と命名した HTLV - III のこの遺伝子領域は、HTLV - I および HTLV - II の lor の少なくとも 2 倍の長さであり、この領域に 1 つまたは 2 つ以上のタンパク質がコードされているかどうか現在のところはっきりしない。

【0063】

多数のクローンを分析するために、サザンブロッティングおよび DNA 塩基配列決定の両方法を使用した。図 2 . B に示すように、エイズ患者の血清と免疫反応する融合タンパク質を発現する Lac⁺ ORF クローンは、ORF - A (例えば # 175 および # 191)、ORF - B (例えば # 13, 31 および 162) または ORF - D (例えば # 113, 121 および 127) の領域に位置し、sor 領域には存在しなかった。これらの領域の全てのタンパク質が免疫反応性であるとは限らない (例えば ORF - D に位置する ORF クローン # 76)。

【0064】

HTLV - III の読み取り枠構造の分析は、どの読み取り枠が env 遺伝子に相当するかについての疑問を提起した。HTLV - III の env - lor 領域が推定される lor 遺伝子の他に env 遺伝子の全部または一部を含むと考えられる。最近になって、HTLV - I の lor 遺伝子はウイルス活性化および形質転換の過程に関連した 42 Kd タン

10

20

30

40

50

パク質をコードすることが示唆された。ORFクローンの1つ(図2・Bの#127)の溶菌液を、ウエスタンブロット法に基づくストリッブラジオイムノアッセイで、20人のエイズ患者と12人の正常人からの血清に対して試験した場合、ラムダC I - H T L V - I I I - β - ガラクトシダーゼ融合タンパク質に対する免疫反応は19人のエイズ患者血清において検出され、正常対照からは全く検出されなかった。この結果は、ORFクローン#127に含まれるenv-lor領域の部分によりコードされるタンパク質がH T L V - I I I 感染細胞内で生産され、そしてエイズ患者の、全部ではないにしても、大部分に抗体を産生させるということを示している。

産業上の適用可能性

本発明は、体液中のH T L V - I I I DNAの存在に関するスクリーニングおよびエイズの診断に利用される。 10

均等物

本明細書で述べた特定の物質および方法に均等の多数の物質および方法が、単に日常的な実験法を使用することにより当業者によって認識され、また確認されるであろう。このような均等物は本発明の範囲内であると考えられ、先の特許請求の範囲に包含される。

【図面の簡単な説明】

【図1】図1はH T L V - I I I DNAを表わす。図1・aはゲノムが制限酵素S s t Iによって切断される部位を示し、図1・bは制限酵素K p n , E c o R IおよびH i n d I I Iの作用により作られたH T L V - I I Iゲノムの断片を示す。

【図2】図2はH T L V - I I I DNAを示す。図2・Aはゲノム内の制限酵素部位を示し、図2・Bは読み取り枠クローン内のDNA挿入物のH T L V - I I Iゲノム中での位置を示す。(+)および(-)は、それぞれORFベクターを用いて形質転換した細胞により発現された融合タンパク質とエイズ患者の血清との反応性および非反応性を示す。 20

【図3】図3はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その1)を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図4】図4はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その2)を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図5】図5はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その3)を示す。ヌクレオチド配列の上に制限酵素部位を示す。 30

【図6】図6はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その4)を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図7】図7はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その5)を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図8】図8はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その6)を示す。ヌクレオチド配列の上に制限酵素部位を示す。 40

【図9】図9はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その7)を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図10】図10はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その8)を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図11】図11はH T L V - I I I DNAのヌクレオチド配列、および4つの最も長い読み取り枠の推定上のアミノ酸配列(その9)を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図 1 2】図 1 2 は H T L V - I I I DNA のヌクレオチド配列、および 4 つの最も長い読み取り枠の推定状のアミノ酸配列 (その 1 0) を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図 1 3】図 1 3 は H T L V - I I I DNA のヌクレオチド配列、および 4 つの最も長い読み取り枠の推定状のアミノ酸配列 (その 1 1) を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図 1 4】図 1 4 は H T L V - I I I DNA のヌクレオチド配列、および 4 つの最も長い読み取り枠の推定状のアミノ酸配列 (その 1 2) を示す。ヌクレオチド配列の上に制限酵素部位を示す。

【図 1 5】図 1 5 は H T L V - I I I e n v - β - ガラクトシダーゼ融合タンパク質の S D S - ポリアクリルアミドゲル上での位置を示すイムノプロットの写真である。 10

【図 1 6】図 1 6 はゲノムが制限酵素 E c o R I によって切断される部位、および H T L V - I I I DNA を保有する組換えプラスミドの作製を示す。

【図 1 7】図 1 7 は組換え構成物 o m p A 1 - R - 6 , o m p A 2 - R - 7 および o m p A 3 - R - 3 (この中に 1 . 1 K b の E c o R I H T L V - I I I c D N A 制限断片が挿入された) を用いて形質転換した細菌細胞により生産されたペプチドのニトロセルロースフィルター上での位置を示すイムノプロットの写真である。

【図 1 8】図 1 8 は o m p A シグナルペプチドのヌクレオチド配列、および組換えプラスミド o m p A 1 - R - 6 , o m p A 2 - R - 7 , o m p A 3 - R - 3 の関連領域を示す。

【図 1 9】図 1 9 は H T L V - I I I DNA 組換えプラスミド o m p A 1 - R - 6 を含む大腸菌の溶菌液による H T L V - I I I 抗原とエイズ血清との間の反応の阻害 (レーン 1 - 5) 、および大腸菌対照細胞の溶菌液による上記反応の非阻害 (レーン 6 - 1 0) を示すイムノプロットの写真である。 20

【図 2 0】図 2 0 は正常人 (レーン 1 - 3) およびエイズ患者 (レーン 4 - 1 1) からの血清中での、H T L V - I I I c D N A の 1 . 1 K b E c o R I H T L V - I I I 制限断片によりコードされるペプチドに対する抗体の有無を示すイムノプロットの写真である。精製した H T L V - I I I ウイルス (パネル A) 、または細菌クローン o m p A 1 - R - 6 (O 1 R 6) の全細胞溶菌液 (パネル B) を血清試料と反応させた。

【図 2 1】図 2 1 は H T L V - I I I DNA を有する読み取り枠発現ベクター p M R 1 0 0 を示す。 30

【図 2 2】図 2 2 はラムダ C I - H T L V - I I I - β - ガラクトシダーゼ融合タンパク質を示す。図 2 2 . A はラムダ C I - H T L V - I I I - β - ガラクトシダーゼ融合タンパク質の S D S - ポリアクリルアミドゲル上での位置を示すイムノプロットの写真であり、図 2 2 . B は上記タンパク質とエイズ患者血清との免疫反応を示す写真である。

【配列表】

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 11

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 492 base pairs

(B) TYPE: nucleic acid

10

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HTLVIII

(ix) FEATURE:

20

(A) NAME/KEY: misc.sub. feature

(B) LOCATION: 1..492

(D) OTHER INFORMATION: standard.sub.name = "Clone BH10"

note = "Corresponds to nucleotide
positions 453 to 39 in figure 3 of
US 06/693,866 (parent)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

30

```

tggaagggt aattactcc caacgaagac aagatatect tgatcigtgg atctaccaca 60
cacaaggcta ctccctgat tagcagaact acacaccagg gccagggtac agatatccac 120
tgaccttigg atggtgctac aagctagtag cagttgagcc agagaagtta gaagaagcca 180
acaaaggaga gaacaccagc ttgttacacc ctgtgagcct gcatggaatg gatgacccgg 240
agagagaagt gttagagtgg aggtttgaca gccgcctagc atticatcac atggcccagag 300
agctgcatcc ggagtacttc aagaactgct gacatcgagc ttgctacaag ggactttccg 360
ctggggactt tccaggaggg cgtggcctgg gcgggactgg ggagtggcga gccctcagat 420
cctgcatata agcagctgct ttttgccgtg actgggtctc tctggttaga ccagatctga 480

```

40

gcctgggagc tc

492

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 492 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HTLVIII

(ix) FEATURE:

(A) NAME/KEY: misc.sub. feature

(B) LOCATION: 1..492

20

(D) OTHER INFORMATION: standard.sub. name= "Clone BH8"

note = "Corresponds to nucleotide
positions 453 to 39 in figure 3 of
US 06/693,866 (parent)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

tggaagggct aattcactcc caacgaagac aagatatacct tgatcigtgg atccaccaca   60
cacaaggcta ctccctgat tggcagaact acacaccagg gccaggagtc agatatccac   120
tgacctttgg atgggtgtac aagctagtac cagttgagcc agagaagtaa gaagaagcca   180
ataaaggaga gaacaccagc ttgttacacc ctgtgagcct gcatggaatg gatgacctig   240
agagagaagt gttagagtgg aggtttgaca gccgcctagc attcatcac atggcccagag   300
agctgcatcc ggagtacttc aagaactgct gatatcgagc ttgctacaag ggactttccg   360
ctggggactt tccagggagg cgtggcctgg gcgggactgg ggagtggcga gccctcagat   420
cctgcatata agcagctgct ttttgccgtg actgggtctc tctggttaga ccagatctga   480
gcctgggagc tc                                         492

```

40

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 182 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

10

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HTLVIII

(ix) FEATURE:

(A) NAME/KEY: misc.sub. feature

(B) LOCATION: 1..182

(D) OTHER INFORMATION: standard.sub. name= "Clone HXB2"

note = "Corresponds to nucleotide
positions 40 to 221 in figure 3 of
US 06/693,866 (parent)"

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
tctggctaac tagggaaccc actgcttaag cctcaataaa gcttgccctg agtgcttcaa   60
gtagtgtgtg cccgtctgtt gttgactctt ggtaactaga gatccctcag acccttttag   120
tcagtgtgga aaatctctag cagtggcgcc cgaacagggg cctgaaagcg aaagggaaac   180
ca                                                                    182
```

30

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8933 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

40

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HTLVIII

(ix) FEATURE:

(A) NAME/KEY: misc.sub. feature

(B) LOCATION: 1..8933

(D) OTHER INFORMATION: standard.sub. name = "Clone BH10"

note = "Corresponds to nucleotide
positions 222 to 9154 in figure 3 of
EP 85307260"

10

(ix) FEATURE:

(A) NAME/KEY: mat.sub. peptide

(B) LOCATION: 113..1648

(D) OTHER INFORMATION: /product = "gag"

(ix) FEATURE:

20

(A) NAME/KEY: mat.sub. peptide

(B) LOCATION: 1408..4452

(D) OTHER INFORMATION: product = "pol"

(ix) FEATURE:

(A) NAME/KEY: mat.sub. peptide

(B) LOCATION: 4367..4975

(D) OTHER INFORMATION: product = "sor"

30

(ix) FEATURE:

(A) NAME/KEY: mat.sub. peptide

(B) LOCATION: 5560..8148

(D) OTHER INFORMATION: /product = "env"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

gagctctctc gacgcaggac tcggcttgct gaagcgcgca cggcaagagg cgaggggagg 60
cgactggatga gtacgccaaa aattttgact agcggaggct agaaggagag agatgggtgc 120
gagagcgta gtattaagcg ggggagaatt agatcgatgg gaaaaaattc ggtaaggcc 180

40

agggggaaag aaaaaatata aattaaaaca tatagiatgg gcaagcaggg agctagaacg 240
 attcgcagtt aatcctggcc tgttagaaac atcagaaggc tgtagacaaa tactgggaca 300
 gctacaacca tcccttcaga caggatcaga agaactitaga tcattatata atacagtagc 360
 aacctctiat tgtgtgcac aaaggataga gataaaagac accaaggaag ctttagacaa 420
 gatagaggaa gagcaaaaca aaagtaagaa aaaagcacag caagcagcag ctgacacagg 480
 acacagcagt caggtcagcc aaaattaccc tatagtgcag aacatccagg ggcaaattgtt 540
 acatcaggcc atatcaccta gaactttaaa tgcattgggt aaagtagtag aagagaaggc 600
 tttcagccca gaagtaatac ccatgttttc agcattatca gaaggagcca cccacaaga 660
 tttaaacc accatgctaaaca cagtgggggg acatcaagca gccatgcaaa tgttaaaaga 720
 gaccatcaat gaggaagctg cagaatggga tagagtacat ccagtgcatt cagggcctat 780
 tgcaccaggc cagatgagag aaccaagggg aagtgcata gcaggaacta ctagtacctt 840
 tcaggaacaa ataggatgga tgacaaataa tccacctatc ccagtaggag aaatttataa 900
 aagatggata atcctgggat taaataaaat agtaagaatg tatagcccta ccagcattct 960
 ggacataaga caaggaccaa aagaaccttt tagagactat gtagaccggg tctataaaac 1020
 tctaagagcc gagcaagctt cacaggaggt aaaaaattgg atgacagaaa cctgtgtgtt 1080
 ccaaaatgct aacctcagatt gtaagactat tttaaaagca ttgggaccag cggctacact 1140
 agaagaaatg atgacagcat gtcaggaggt aggaggaccc ggccataagg caagagtttt 1200
 ggctgaagca atgagccaag taacaaatac agctaccata atgatgcaga gaggcaattt 1260
 taggaaccaa agaaagatgg ttaagtgttt caattgtggc aaagaagggt acacagccag 1320
 aaattgcagg gcccctagga aaaagggtct ttggaaatgt ggaaagggaag gacaccaaatt 1380
 gaaagattgt actgagagac aggttaattt tttagggaag atctggccct cctacaaggg 1440
 aaggccaggg aattttcttc agagcagacc agagccaaca gcccaccat ttcttcagag 1500
 cagaccagag ccaacagccc caccagaaga gagcttcagg tctggggtag agacaacaac 1560
 tccccctcag aagcaggagc cgatagacaa ggaactgtat cctttaactt cctcagatc 1620
 actctttggc aacgacctt cgtcacaata aagatagggg ggcaactaaa ggaagctcta 1680
 ttagatacag gagcagatga tacagtatta gaagaaatga gtttgccagg aagatggaaa 1740
 ccaaaaatga tagggggaat tggaggtttt atcaaagtaa gacagiatga tcagatactc 1800
 atagaaatct gtggacataa agctataggt acagtattag taggacctac acctgtcaac 1860
 ataattggaa gaaatctgtt gactcagatt ggttgcactt taaattttcc cattagccct 1920

10

20

30

40

attgagactg taccagttaa attaaagcca ggaatggatg gcccaaaagt taaacaatgg 1980
 ccattgacag aagaaaaaat aaaagcatta gtagaaattt gtacagaaat ggaaaaggaa 2040
 gggaaaattt caaaaatigg gcctgagaat ccatacaata ciccagtatt tgccataaag 2100
 aaaaaagaca gtactaaatg gagaaaatta gtagattica gagaacttaa taagagaact 2160
 caagactict gggaagtica attaggaata ccacatcccg cagggitaaa aaagaaaaaa 2220
 tcagtaacag tactggatgt gggtagatga tatttttcag ticccttaga tgaagacttc 2280
 aggaagtata ctgcatttac catacctagt ataaacaatg agacaccagg gattagatat 2340
 cagtacaatg tgcctccaca gggatggaaa ggatcaccag caatattcca aagtagcatg 2400
 aaaaaaatct tagagccttt taaaaaaca aatccagaca tagttatcta tcaatacatg 2460
 gatgatttgt atgtaggatc tgacttagaa atagggcagc atagaacaaa aatagaggag 2520
 ctgagacaac atcigttag gtaggggacti accacaccag acaaaaaaca tcagaaagaa 2580
 cctccattcc tttaggatggg ttatgaactc catcctgata aatggacagt acagcctata 2640
 gtgctgccag aaaaagacag ctggacigtc aatgacatac agaagttagt ggggaaatig 2700
 aattgggcaa gtcagattta cccagggatt aaagtaaggc aattatgtaa actccttaga 2760
 ggaaccaaag cactaacaga agtaatacca ctaacagaag aagcagagct agaactggca 2820
 gaaaacagag agattctaaa agaaccagta catggagigt attatgacct atcaaaagac 2880
 ttaatagcag aaatacagaa gcaggggcaa ggccaatgga catatcaaat ttatcaagag 2940
 ccatttaaaa atcigaaaac aggaaaatat gcaagaatga ggggtgcca cactaatgat 3000
 gtaaaacaat taacagaggc agtgcaaaaa ataaccacag aaagcatagt aatatgggga 3060
 aagactccta aattttaact acccatacaa aaggaaacat gggaaacatg gtagacagag 3120
 tattggcaag ccacctggat tcttgagtgg gagtttgta atacctctc tttagtgaaa 3180
 ttatggtacc agttagagaa agaaccata gtaggagcag aaaccttcta tgtagatggg 3240
 gcagctaaca gggagactaa attaggaaaa gcaggatatg ttactaaca aggaagacaa 3300
 aaggttgtcc cctaactaa cacaacaaat cagaaaactg agttacaagc aatttatcta 3360
 gctttgcagg attcaggatt agaagtaaac atagtaacag actcacaata tgcattagga 3420
 atcattcaag cacaaccaga taaaagigaa tcagagttag tcaatcaaat aatagagcag 3480
 ttaataaaaa aggaaaaggc ctatctggca tgggtaccag cacacaaagg aattggagga 3540
 aatgaacaag tagataaatt agtcagtgct ggaatcagga aaatactatt tttagatgga 3600
 atagataagg cccaagatga acatgagaaa tatcacagta attggagagc aatggctagt 3660

10

20

30

40

gattttaacc tgcacctgt agtagcaaaa gaaatagtag ccagctgtga taaatgtcag 3720
 ctaaaaggag aagccatgca tggacaagta gactgtagtc caggaatatg gcaactagat 3780
 tgtacacatt tagaaggaaa agttatccig gtagcagttc atgtagccag tggatatata 3840
 gaagcagaag ttattccagc agaaacaggg caggaaacag catatittct tttaaaatta 3900
 gcaggaagat ggccagttaa aacaatacat acagacaatg gcagcaattt caccagtgct 3960
 acggttaagg ccgctgttig gtgggcggga atcaagcagg aatttggat tcctacaat 4020
 ccccaaagtc aaggagtagt agaattctatg aataaagaat taaagaaaat tataggacag 4080
 gtaagagatc aggttgaaca tcttaagaca gcagtacaaa tggcagtatt catccacaat 4140
 tttaaaagaa aaggggggat tggggggtac agtgcagggg aaagaatagt agacataata 4200
 gcaacagaca tacaaactaa agaattacaa aaacaaatta caaaaattca aaattttcgg 4260
 gtttattaca gggacagcag aaatccactt tggaaaggac cagcaaagct cctctggaaa 4320
 ggtgaagggg cagtagtaat acaagataat agtgacataa aagtagtgcc aagaagaaaa 4380
 gcaaagatca ttagggttta tggaaaacag atggcagggtg atgattgtgt ggcaagtiga 4440
 caggatgagg attagaacat ggaaaagttt agtaaaacac catatgtatg tttcagggaa 4500
 agctaggggg tggttttata gacatcacta tgaagccct catccaagaa taagttcaga 4560
 agtacacatc ccactagggg atgctagatt ggtaataaca acatattggg gcttcgatac 4620
 aggagaaaga gactggcatt tgggtcaggg agtctccata gaatggagga aaaagagata 4680
 tagcacacaa gtagacctig aactagcaga ccaactaatt catctgtatt actttgacig 4740
 ttttcagac tctgtataa gaaaggcctt attaggacac atagttagcc ctagggtgtga 4800
 atatcaagca ggacataaca aggtaggatc tctacaatac ttggcactag cagcattaat 4860
 aacacaaaa aagataaagc caccittgcc tagtgttacg aaactgacag aggatagatg 4920
 gaacaagccc cagaagacca agggccacag agggagccac acaatgaatg gacactagag 4980
 cttttagagg agcttaagaa tgaagctgtt agacattttc ctaggatttg gctccatggc 5040
 ttagggaac atatctatga aacttatggg gatacttggg caggagtggg agccataata 5100
 agaattctgc aacaactgct gtttatccat tticagaatt ggggtgtcgc atagcagaat 5160
 aggcgttact cgacagagga gagcaagaaa tggagccagt agatcctaga ctagagccct 5220
 ggaagcatcc aggaagtcag cctaaaactg ctigtaccaa ttgctattgt aaaaagtgtt 5280
 gctttcattg ccaagtttgt ttcataacaa aagccttagg catctectat ggcaggaaga 5340
 agcggagaca gcgacgaaga cctctcaag gcgtcagac tcatcaagtt tctctatcaa 5400

10

20

30

40

agcagtaagt agtacaatgta atgcaacctt tacaatatgc aatagtagca ttagtagtag 5460
 caataataat agcaatagtt gtgtgggtcca tagtaatcat agaatatagg aaaatatata 5520
 gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa gacagtggtca 5580
 atgagagiga aggagaaata tcagcacttg tggagatggg ggtggagatg gggcaccatg 5640
 ctcttggga tgttgatgat ctgtatgct acagaaaaat tgtgggtcac agtctattat 5700
 ggggtacctg tgtggaagga agcaaccacc actctatit tgtcatcaga tgcataagca 5760
 tatgatacag aggtacataa tgtttgggcc acacatgctt gtgtacccac agaccccaac 5820
 ccacaagaag tagtattggt aaatgtgaca gaaaatttta acatgtggaa aaatgacatg 5880
 gtagaacaga tgcattgagga tataatcagt ttatgggata aaagcctaaa gccatgtgta 5940
 aaattaaccc cactctgtgt tagtttaaaag tgcactgatt tgaagaatga tactaatacc 6000
 aatagtagta gcgggagaat gataatggag aaaggagaga taaaaaacgt ctctttcaat 6060
 atcagcacia gcataagagg taagggtcag aaagaatatg catTTTTTT taaacttgat 6120
 ataatacaa tagataatga tactaccagc tatacgtiga caagtigtaa caccicagtc 6180
 attacacagg cctgtccaaa ggtatccttt gagccaattc ccatacatta ttgtgccccg 6240
 gctggttttg cgattctaaa atgtaataat aagacgtica atggaacagg accatgtaca 6300
 aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac tcaactgcig 6360
 ttaaattggca gtciggcaga agaagaggta gtaattagat ctgccaattt cacagacaat 6420
 gctaaaacca taatagtaca gctgaaccaa tcigtatgaaa ttaattgtac aagacccaac 6480
 aacaatacaa gaaaaagtat ccgtatccag agaggaccag ggagagcatt tgttacaata 6540
 ggaaaaatag gaaatatgag acaagcacat tgtaacatta gtagagcaaa atggaataac 6600
 actttaaac agatagatag caaattaaga gaacaatttg gaaataataa aacaataatc 6660
 ttaagcagt cctcaggagg ggaccagaa atigtacgc acagtittaa ttgtggaggg 6720
 gaattttct actgtaatc aacacaactg ttaatatgta ctgtgttaa tagtacttgg 6780
 agtactaaag ggtcaataa cactgaagga agtgacacia tcacctccc atgcagaata 6840
 aaacaaatta taaacatgtg gcaggaagta ggaaaagcaa tgtatgcccc tccatcagt 6900
 ggacaaatta gatgttcac aaatatataa gggctgctat taacaagaga tgggtgtaat 6960
 agcaacaatg agtccgagat cticagacct ggaggaggag atatgaggga caattggaga 7020
 agtgaattat ataaatataa agtagtaaaa atigaacct taggagtagc accaccaag 7080
 gcaaagagaa gagtgggtga gagagaaaaa agagcagttg gaataggagc ttgtttcctt 7140

10

20

30

40

gggttcttgg gagcagcagg aagcactatg ggcgcagcgt caatgacgct gacgggtacag 7200
 gccagacaat tatigtctgg tatagtgcag cagcagaaca attigtctgag ggctatttag 7260
 gcgcaacagc atcigtigca atcacagtc iggggcatca agcagctcca ggcaagaatc 7320
 ctggctgtgg aaagatacct aaaggatcaa cagctccctgg ggatttgggg ttgctctgga 7380
 aaactcattt gcaccactgc tigtccctgg aatgctagtt ggagtaataa atctctggaa 7440
 cagatttggg ataacatgac ctggatggag tgggacagag aaattaacaa ttacacaagc 7500
 ttaatacact ccttaatiga agaatcgcaa aaccagcaag aaaagaatga acaagaatia 7560
 ttggaattag ataaatgggc aagtttgtgg aattggttta acataacaaa ttggctgtgg 7620
 tatataaaat taticataat gatagtagga ggcttggtag gttaaagaat agtttttgc 7680
 gtactttctg tagtgaatag agttaggcag ggatattcac cattatcgtt tcagaccac 7740
 ctcccaatcc cgaggggacc cgacaggccc gaaggaatag aagaagaagg tggagagaga 7800
 gacagagaca gatccattcg attagigaac ggatccttag cacctatctg ggacgatctg 7860
 cggagccigt gccctttcag ctaccaccgc ttgagagact tactcttgat tgiaacgagg 7920
 atttgtgaac ttctgggacg caggggggtgg gaagccctca aatatiggig gaatctccia 7980
 cagtattgga gtcaggagct aaagaatagt gcigttagct tgcicaatgc cacagctata 8040
 gcagtagctg aggggacaga tagggttata gaagtagtac aaggagctta tagagctatt 8100
 cgccacatac ctagaagaat aagacagggc ttggaaagga ttttgctata agatgggtgg 8160
 caagtggica aaaagtatig tggttggatg gccctgctgta agggaaagaa tgagacgagc 8220
 tgagccagca gcagatgggg tgggagcagc atctcgagac ctagaaaaac atggagcaat 8280
 cacaagtatg aacacagcag ctaacaatgc tgattgtgcc tggctagaag cacaagagga 8340
 ggaggaggig ggttttccag tcacacctca ggtaccttta agaccaatga ctacaaggc 8400
 agctgttagat cttagccact ttttaaaaga aaagggggga ctggaagggc taattcacic 8460
 ccaacgaaga caagataatc ttgatctgtg gatctaccac acacaaggct acttccctga 8520
 ttagcagaac tacacaccag ggccagggat cagatatcca ctgaccttg gatggtgcta 8580
 caagctagta ccagttaggc cagagaagtt agaagaagcc aacaaaggag agaacaccag 8640
 ctgtttacac cctgtgagcc tgcattggaat ggatgaccgc gagagagaag tgttagagtg 8700
 gaggtttgac agccgcctag catticatca catggccga gagctgcac cggagtactt 8760
 caagaactgc tgacatcgag ctigtctaaa gggactttcc gctggggact ticcaggag 8820
 gcgtggccctg ggcgggactg gggagtggcg agccctcaga tcttgcataa aagcagctgc 8880

10

20

30

40

ttttgacctg tactgggtct cctgggttag accagatctg agcctgggag ctc 8933

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5362 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HTLVIII

(ix) FEATURE:

(A) NAME/KEY: misc.sub. feature

(B) LOCATION: 1..5362

20

(D) OTHER INFORMATION: standard.sub. name = "Clone BH5"

note = "Corresponds to nucleotide
positions 222 to 5585 in figure 3 of
US 06/693,866 (parent)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

gagctctctc gacgcaggac tcggcttgcg agcgcgcacg gcaagaggcg aggggcggcg 60
actggtagat acgccccaaa tttgacttag cggaggctag aaggagagag atgggtgcga 120
gagcgtcagt attaagcggg ggagaattag atcgatggga aaaaattcgg ttaaggccag 180
ggggaaagaa aaaatataaa ttaaaacata tagtatgggc aagcaggag ctagaacgat 240
tcgcagttaa tcttggcctg ttagaaacat cagaaggctg tagacaaata ctgggacagc 300
tacaaccatc ccttcagaca ggatcagaag aacttagatc attatataat acagtagcaa 360
ccctctattg tgtgcatcaa aggatagaga taaaagacac caaggaagct ttagacaaga 420
tagaggaaga gcaaaacaaa agtaagaaaa aagcacagca agcagcagct gacacaggac 480
acagcagica ggtcagccaa aattacccta tagtgcagaa catccagggg caaatggtag 540
atcaggccat atcacctaga actttaaatg catgggtaaa agtagtagaa gagaaggctt 600

30

40

tcagcccaga agtgataccc atgttticag cattatcaga aggagccacc ccacaagatt 660
 taaacaccat gctaaacaca gtggggggac atcaagcagc catgcaaag ttaaaagaga 720
 ccatcaatga ggaagctgca gaatgggata gagtgcatcc agtgcatgca gggcciatcg 780
 caccaggcca gatgagagaa ccaaggggaa gtgacatagc aggaactact agtacccttc 840
 aggaacaaat aggatggatg acaaataatc cacctatccc agtagggagaa atttataaaa 900
 gatggataat cctgggattt aataaaatag taaggatgta tagtccctacc agcattctgg 960
 acataagaca aggaccaaag gaacccttta gagactatgt agaccgggtc tataaaacac 1020
 taagagccga gcaagcttca caggaagtaa aaaattggat gacagaaacc ttgttgggtc 1080
 aaaatgcgaa cccagattgt aagactatit taaaagcatt gggaccagcg gctacactag 1140
 aagaaatgat gacagcatgt cagggagtag gaggaccggg ccataaggca agagttttgg 1200
 ctgaagcaat gagccaagta acaaattcaa ctaccataat gatgcaaaga ggcaatttta 1260
 ggaaccaaag aaaaattgtt aagtgtttca atgttggcaa agaagggcac atagcaagaa 1320
 attgcaaggc ccttagaaaa aagggcgttt ggaaatgttg aaaggaagga caccaaatga 1380
 aagattgtac tgagagacag gctaattttt tagggaagat ctggcccttc tacaagggaa 1440
 ggccagggaa tttcttcag agcagaccag agccaacagc cccaccattt cttagagca 1500
 gaccagagcc aacagcccca ccagaagaga gcttcaggtc tgggtagag acaacaacac 1560
 cccctcagaa gcaggagccg atagacaagg aactgtatcc ttttaacttc cttagatcac 1620
 tctttggcaa cgacccctcg tcacaataaa gatagggggg caactaaagg aagctctatt 1680
 agatacagga gcagatgata cagtattaga agaaatgagt ttgccaggaa gatggaaacc 1740
 aaaaatgata gggggaattg gaggttttat caaagtaaga cagtatgac agatactcat 1800
 agaaatcgtt ggacataaag ctataggtag agtattagta ggacctacac ctgtcaacat 1860
 aattggaaga aatctgttga cttagattgg ttgcacttta aattttccca ttagtccat 1920
 tgaaactgta ccagtataat taaagccagg aatggatggc ccaaaagta aacaatggcc 1980
 attgacagaa gaaaaataa aagcattagt agaaattgtt acagaaatgg aaaaggaagg 2040
 gaaaattica aaaattgggc ctgaaaatcc atacaatact ccagtatttg ccataaagaa 2100
 aaaagacagt actaaatgga gaaaattagt agatttcaga gaacttaata ggagaactca 2160
 agacttcagg gaagtccaat tgggaatacc acatcccga gggttaaaaa agaaaaaatc 2220
 agtaacagta ctggatgttg gtgatgcata ttttcagtt ccttagatg aagacttcag 2280
 gaagtatact gcatttacca tacttagtat aaataatgag acaccaggga gtggatatca 2340

10

20

30

40

gtacaatgtg ctccacagg gatggaaagg atcaccagca atattccaaa giagcatgac 2400
 aaaaatctta gagcctttta gaaaacaaaa tccagacata gttatttata aatacatgga 2460
 tgatttgtat gtaggatctg acttagaaat agggcagcat agaacaaaaa tagaggagct 2520
 gagacaacat ctgttgaggt ggggatitac cacaccagac aaaaaacatc agaaagaacc 2580
 tccattcctt tggatgggtt atgaaticca tcttgataaa tggacgatac agcctatagt 2640
 gctgccagaa aaagacagct ggactigcaa tgacatacag aagttagtgg gaaaattgaa 2700
 ttgggcaagt cagatttata cagggaataa agtaaggcaa ttatgtaaac tcttagagg 2760
 aaccaaagca ctaacagaag taataccact aacagaagaa gcagagctag aactggcaga 2820
 aaacagagag attctaaaag aaccagtaca tggagtgtat tatgacccat caaaagacit 2880
 aatagcagaa atacagaagc aggggcaagg ccaatggaca tatcaaattt atcaagagcc 2940
 atttaaaaat ctgaaaacag gaaaataatgc aagaatgagg ggtgcccaca ctaatgatgt 3000
 aaaacaatta acagaggcag tgcaaaaaat aaccacagaa agcatagtaa tatggggaaa 3060
 gactcctaaa tttaaactac ccatacaaaa agaaacatgg gaaacatggt ggacagagta 3120
 ttggcaagcc acctggattc ctgagtggga gtttgttaat accctcctt tagtgaaatt 3180
 atggtaccag ttagagaaag aacccatagt aggagcagaa accttctatg tagatggggc 3240
 agctagcagg gagactaaat taggaaaagc aggatatgtt actaatagag gaagacaaaa 3300
 agttgtacc ctaactaca caacaaatca gaagactgaa ttacaagcaa tcatctagc 3360
 ttgacaggat tgggattag aagttaaata agtaacagac tcacaatatg cattaggaat 3420
 cattcaagca caaccagata aaagtgaatc agagttagtc aatcaaataa tagagcagtt 3480
 aataaaaaag gaaaaggctt atctggcatg ggtaccagca cacaaaggaa tiggaggaaa 3540
 tgaacaagta gataaattag tcagtgcctg aatcaggaaa atactatttt tagatggaat 3600
 agataaggcc caagaagaac atgagaaata tcacagtaat tggagagcaa tggctagtga 3660
 ttttaacctg ccacctgtag tagcaaaaga aatagtagcc agctgtgata aatgtcagct 3720
 aaaaggagaa gccatgcatg gacaagtaga ctgtagtcca ggaatatggc aactagattg 3780
 tacacatita gaaggaaaag ttatccigtg agcagttcat gtagccagtg gatatataga 3840
 agcagaagtt attccagcag aaacagggca ggaaacagca tattttctt taaaattagc 3900
 aggaagaatg ccagtaaaaa caatacatac agacaatggc agcaatttca ccagtgtac 3960
 ggttaaggcc gccigtgtgt gggcgggaat caagcaggaa ttiggaattc cctacaatcc 4020
 ccaaagtcaa ggagttagt aatctatgaa taaagaatta aagaaaatta taggacaggt 4080

10

20

30

40

aagagatcag gctgaacatc ttaagacagc agtacaaatg gcagtattca tccacaattt 4140
taaaagaaaa ggggggattg ggggggtacag tgcaggggaa agaatagtag acataatagc 4200
aacagacata caaactaaag aattacaaaa acaaatitaca aaaatitaaa attttcgggt 4260
ttattacagg gacagcagaa atccacittg gaaaggacca gcaaagctcc tctggaaagg 4320
tgaaggggca gtagtaatac aagataatag tgacataaaa gtagtgccaa gaagaaaagc 4380
aaagatcatt agggattatg gaaaacagat ggcaggtagat gattgigtgg caagtagaca 4440
ggatgaggat tagaacaatg aaaagttag taaaacaccg tatgtatgtt tcagggaag 4500
ctaggggatg gttttataga catcactatg aaagccctca tccaagaata agttcagaag 4560
tacacatccc actaggggat gctagattgg taataacaac atattgggtt ctgcatacag 4620
gagaaagaga ctggcatttg ggtcaggag tctccataga atggaggaaa aggagatata 4680
gcacacaagt agaccctgaa ctagcagacc aactaattca tctgcattac ttgatgtgtt 4740
tttcagactc tgcataaga aaggccctat taggacacat agttagccct aggtgtgaat 4800
atcaagcagg acataacaag gtaggatctc tacaatactt ggcactagca gcattaataa 4860
caccaaaaaa ggtaaagcca cctttgcta gttttacgaa actgacagag gatagatgga 4920
acaagcccca gaagaccaag ggccacagag gaagccacac aatgaatgga cactagagct 4980
tttagaggag cttagaatg aagctgttag acattttcct aggatitggc tccatggctt 5040
agggaacat atctatgaaa cttatgggga tacttgggca ggagtgggaag ccataataag 5100
aattctgcaa caactgctgt ttatccattt tcagaattgg gtgtcgacat agcagaatag 5160
gcgttactca acagaggaga gcaagaaatg gagccagtag atcctagact agagccctgg 5220
aagcatccag gaagtcagcc taaaactgct tgiaccactt gctatigtaa aaagtgttgc 5280
tttcattgcc aagttgttt cataacaaaa gccttaggca tctcctatgg caggaagaag 5340
cggagacagc gacgaagagc tc 5362

10

20

30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3563 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

40

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HTLVIII

(ix) FEATURE:

(A) NAME/KEY: misc.sub. feature

(B) LOCATION: 1..3563

(D) OTHER INFORMATION: standard.sub. name = "Clone BH8"

10

note = "Corresponds to nucleotide
positions 5580 to 9154 in figure 3 of
US 06/693,866"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

gagctcatcg aagcagtcag actcatcaag ttctctatc aaagcagtaa gtagtacatg 60
taacgcaacc tataccaata gtaacaatag tagccitagc agtagcaata ataatagcaa 120
tagttgtgtg gtccatagta atcatagaat ataggaaaat attaagacaa agaaaaatag 180
acaggtaaat tgatagacta atagaaagag cagaagacag tggcaatgag agtgaaggag 240
aaatatcagc acttgtggag atgggggtgg agatggggca ccatgtcctt tgggatgttg 300
atgatctgta gtgctacaga aaaatgttgg gtcacagtct attttggggg acctgtgttg 360
aaggaagcaa ccaccactct attttgtgca tcagatgcta aagcatatga tacagaggta 420
cataatgitt gggccacaca tgccigtgta cccacagacc ccaaccaca agaagtagta 480
ttggtaaatg tgacagaaaa ttttaacatg tggaaaaatg acatggtaga acagatgcat 540
gaggatataa tcagtttatg ggatcaaagc ctaaagccat gtgtaaaatt aacccctc 600
tgtgttagtt taaagtgcac tgatttgaag aatgatacta ataccaatag tagtagcggg 660
agaatgataa tggagaaagg agagataaaa aactgtctct tcaatatcag cacaagcaaa 720
agaggtaagg tgcagaaaga atatgcattt tttataaac ttgatataat accaatagat 780
aatgatacta ccagctatac gtigacaagt tgiaacacct cagtcattac acaggccctgt 840
ccaaaggtaa cctttgagcc aattcccata cattattgtg ccccggtctg ttttgcgatt 900
ctaaaatgta ataataagac gtccaatgga acaggacat gtacaaatgt cagcacagta 960
caatgtacac atggaattag gccagtagta tcaactcaac tgcgtttaa tggcagtcgt 1020
gcagaagaag aggtagtaat tagatctgtc aatttcacgg acaatgctaa aaccataata 1080

```

20

30

40

gtacagctgg acacatctgt agaaattaat tgiacaagac ccaacaacaa tacaagaaaa 1140
 aaaatccgta tccagagggg accagggaga gcatttgita caataggaaa aataggaaat 1200
 atgagacaag cacatigtia cattagiaga gcaaaatgga atgccacttt aaaacagata 1260
 gatagcaaat taagagaaca atttggaaat aataaaacaa taatcittaa gcagtcctca 1320
 ggaggggacc cagaaatigt aacgcacagt tttaattigt gaggggaatt ttcttactgt 1380
 aattcaacac aacigtittaa tagtacttgg agtactaaag ggtcaaataa cactgaagga 1440
 agtgacacaa tcacctccc atgcagaata aaacaaatta taaacatgtg gcaggaagia 1500
 ggaaaagcaa tgtatgcccc tccatcagt ggacaaatta gatgttcatc aaatattaca 1560
 gggctgctat taacaagaga tggitggiaat agcaacaatg agtccgagat cticagacct 1620
 ggaggaggag ataigagggg caatiggaga agigaattat ataaatataa agtagtaaaa 1680
 attgaacat taggagtagc accaccaag gcaaagagaa gagtggtagc gagagaaaaa 1740
 agagcagtgg gaataggagc ttigtctctt gggttcttgg gagcagcagg aagcactatg 1800
 ggcgcagcgt caatgacgt gacggtacag gccagacaat tattgtctgg tatagtgcag 1860
 cagcagaaca attigtctgag ggctattgag ggccaacagc atctgttgc actcacagtc 1920
 tggggcatca agcagctcca ggcaagaatc ctggctgtgg aaagatacct aaaggatcaa 1980
 cagctcctgg ggatttgggg ttgtcttggg aaactcattt gcaccactgc tgtgccctgg 2040
 aatgctagt ttggagtaata atctctggaa cagatttggg ataacatgac ctggatggag 2100
 tgggacagag aaattaacaa ttacacaagc ttaatacact ccttaattga agaatcgcaa 2160
 aaccagcaag aaaagaatga acaagaatta ttggaattag ataaatgggc aagtittgtg 2220
 aattggitta acataacaaa ttggctgttg tatataaaat tattcataat gatagtagga 2280
 ggcttggtag gtttaagaat agtttttct gtactttcta tagtgaatag agttaggcag 2340
 ggatattcac cattatcgtt tcagaccac ctcccaaacc cgaggggacc cgacaggccc 2400
 gaaggaatag aagaagaagg tggagagaga gacagagaca gatccattcg attagtgaac 2460
 ggatccttag cacttatctg ggacgatctg cggagcctgt gccctttcag ctaccaccgc 2520
 ttgagagact tactcttgat tgaacgagg attgtggaac ttctgggacg caggggttgg 2580
 gaagccctca aatattggig gaatctctta cagtattgga gtcaggaact aaagaatagt 2640
 gctgttaact tgcicaatgc cacagctata gcagtagctg aggggacaga tagggttata 2700
 gaattagtac aagcagctta tagagccatt cgccacatac ctagaagaat aagacagggc 2760
 ttggaaagga ttttgcata agatgggttg caagtgttca aaaagtagtg tggttggatg 2820

10

20

30

40

gcctgctgta agggaaagaa tgagacgagc tgagccagca gcagatggg tgggagcagt 2880
 atctcgagac ctagaaaaac atggagcaat cacaagtagc aatacagcag ctaccaatgc 2940
 cgattgtgct tggctagaag cacaagagga ggaggaggig ggttticcag tcacacctca 3000
 ggtaccttta agaccaatga cttaacaaggc agctgtagat cttagccact ttttaaaaga 3060
 aaagggggga ctggaagggc taattcatic ccaacgaaga caagatatcc tigtatctig 3120
 gatccaccac acacaaggct acttccctga ttggcagaac tacacaccag ggccaggagt 3180
 cagatatcca ctgaccttig gatggigtca caagctagta ccagttagc cagagaagia 3240
 agaagaagcc aataaaggag agaacaccag ctigttagac cctgtgagcc tgcattggaat 3300
 ggatgacct gagagagaag ttttagagtg gaggtttgac agccgcctag catticatca 3360
 catggcccga gagctgcatc cggagtactt caagaactgc tgatatcgag ctgtctacaa 3420
 gggactttcc gctggggact ttccaggag gcgtggcctg ggcgggactg gggagtggcg 3480
 agccctcaga tctgcatat aagcagctgc ttttggcctg tactgggtct ctctggttag 3540
 accagatctg agcctgggag ctc 3563

10

(2) INFORMATION FOR SEQ ID NO:7:

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

30

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: HTLV-III

(ix) FEATURE:

- (A) NAME/KEY: misc.sub. feature
- (B) LOCATION: 1..142

- (D) OTHER INFORMATION: standard.sub. name = "Clone HXB2"

note = "Corresponds to nucleotide

positions 9155 to 9296 in figure 3 of

40

US 06/693,866"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

tctggctagc tagggaaccc actgcttaag cctcaataaa gcttgccttg agtgcctcaa   60
gtagtgtgtg cccgtctgtt gtgtgactct ggtaactaga gatccctcag acccttttag   120
tcagtgtgga aaatctctag ca                                           142

```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 512 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

20

(A) ORGANISM: HTLV-III

(ix) FEATURE:

(A) NAME/KEY: Protein

(B) LOCATION: 1..512

(D) OTHER INFORMATION: note = "gag protein of HTLV-III"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Met Gly Ala Arg Ala Ser Val Leu Ser Gly Gly Glu Leu Asp Arg Trp   30
 1           5           10           15
Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys Tyr Lys Leu Lys
          20          25          30
His Ile Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Val Asn Pro
          35          40          45
Gly Leu Leu Glu Thr Ser Glu Gly Cys Arg Gln Ile Leu Gly Gln Leu
          50          55          60
Gln Pro Ser Leu Gln Thr Gly Ser Glu Glu Leu Arg Ser Leu Tyr Asn

```

30

40

65 70 75 80
 Thr Val Ala Thr Leu Tyr Cys Val His Gln Arg Ile Glu Ile Lys Asp
 85 90 95
 Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Glu Gln Asn Lys Ser Lys
 100 105 110
 Lys Lys Ala Gln Gln Ala Ala Ala Asp Thr Gly His Ser Ser Gln Val
 115 120 125
 Ser Gln Asn Tyr Pro Ile Val Gln Asn Ile Gln Gly Gln Met Val His
 130 135 140
 Gln Ala Ile Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Val Glu
 145 150 155 160
 Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu Ser
 165 170 175
 Glu Gly Ala Thr Pro Gln Asp Leu Asn Thr Met Leu Asn Thr Val Gly
 180 185 190
 Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu Glu
 195 200 205
 Ala Ala Glu Trp Asp Arg Val His Pro Val His Ala Gly Pro Ile Ala
 210 215 220
 Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr
 225 230 235 240
 Ser Thr Leu Gln Glu Gln Ile Gly Trp Met Thr Asn Asn Pro Pro Ile
 245 250 255
 Pro Val Gly Glu Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys
 260 265 270
 Ile Val Arg Met Tyr Ser Pro Thr Ser Ile Leu Asp Ile Arg Gln Gly
 275 280 285
 Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Tyr Lys Thr Leu
 290 295 300

10

20

30

40

Arg Ala Glu Gln Ala Ser Gln Glu Val Lys Asn Trp Met Thr Glu Thr
 305 310 315 320
 Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Thr Ile Leu Lys Ala
 325 330 335
 Leu Gly Pro Ala Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
 340 345 350
 Val Gly Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
 355 360 365
 Gln Val Thr Asn Thr Ala Thr Ile Met Met Gln Arg Gly Asn Phe Arg
 370 375 380
 Asn Gln Arg Lys Met Val Lys Cys Phe Asn Cys Gly Lys Glu Gly His
 385 390 395 400
 Thr Ala Arg Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys
 405 410 415
 Gly Lys Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn
 420 425 430
 Phe Leu Gly Lys Ile Trp Pro Ser Tyr Lys Gly Arg Pro Gly Asn Phe
 435 440 445
 Leu Gln Ser Arg Pro Glu Pro Thr Ala Pro Pro Phe Leu Gln Ser Arg
 450 455 460
 Pro Glu Pro Thr Ala Pro Pro Glu Glu Ser Phe Arg Ser Gly Val Glu
 465 470 475 480
 Thr Thr Thr Pro Pro Gln Lys Gln Glu Pro Ile Asp Lys Glu Leu Tyr
 485 490 495
 Pro Leu Thr Ser Leu Arg Ser Leu Phe Gly Asn Asp Pro Ser Ser Gln
 500 505 510

10

20

30

(2) INFORMATION FOR SEQ ID NO:9:

(1) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 1015 amino acids

- (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: protein
 (iii) HYPOTHETICAL: NO
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: HTLV-III
 (ix) FEATURE:
 (A) NAME/KEY: Protein
 (B) LOCATION: 1..1015
 (D) OTHER INFORMATION: note = "pol protein of HTLV-III"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Phe Phe Arg Glu Asp Leu Ala Phe Leu Gln Gly Lys Ala Arg Glu Phe
 1 5 10 15
 Ser Ser Glu Gln Thr Arg Ala Asn Ser Pro Thr Ile Ser Ser Glu Gln
 20 25 30
 Thr Arg Ala Asn Ser Pro Thr Arg Arg Glu Leu Gln Val Trp Gly Arg
 35 40 45
 Asp Asn Asn Ser Pro Ser Glu Ala Gly Ala Asp Arg Gln Gly Thr Val
 50 55 60
 Ser Phe Asn Phe Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr
 65 70 75 80
 Ile Lys Ile Gly Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala
 85 90 95
 Asp Asp Thr Val Leu Glu Glu Met Ser Leu Pro Gly Arg Trp Lys Pro
 100 105 110
 Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp
 115 120 125
 Gln Ile Leu Ile Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu

10

20

30

40

130 135 140
 Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln
 145 150 155 160
 Ile Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro
 165 170 175
 Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro
 180 185 190
 Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met
 195 200 205
 Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn
 210 215 220
 Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys
 225 230 235 240
 Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu
 245 250 255
 Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser
 260 265 270
 Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp
 275 280 285
 Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn
 290 295 300
 Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp
 305 310 315 320
 Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu
 325 330 335
 Pro Phe Lys Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp
 340 345 350
 Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys
 355 360 365

10

20

30

40

Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro
 370 375 380
 Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu
 385 390 395 400
 Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys
 405 410 415
 Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn
 420 425 430
 Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys
 435 440 445
 Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu
 450 455 460
 Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro
 465 470 475 480
 Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile
 485 490 495
 Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro
 500 505 510
 Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His
 515 520 525
 Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr
 530 535 540
 Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile
 545 550 555 560
 Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr
 565 570 575
 Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu
 580 585 590
 Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr

10

20

30

40

595	600	605
Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr		
610	615	620
Val Thr Asn Lys Gly Arg Gln Lys Val Val Pro Leu Thr Asn Thr Thr		
625	630	635
Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser		
645	650	655
Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile		
660	665	670
Ile Gln Ala Gln Pro Asp Lys Ser Glu Ser Glu Leu Val Asn Gln Ile		
675	680	685
Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro		
690	695	700
Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser		
705	710	715
Ala Gly Ile Arg Lys Ile Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln		
725	730	735
Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp		
740	745	750
Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp		
755	760	765
Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser		
770	775	780
Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile		
785	790	795
Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile		
805	810	815
Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala		
820	825	830

10

20

30

40

Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe
 835 840 845
 Thr Ser Ala Thr Val Lys Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln
 850 855 860
 Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser
 865 870 875 880
 Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala
 885 890 895
 Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe
 900 905 910
 Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val
 915 920 925
 Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile
 930 935 940
 Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro
 945 950 955 960
 Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val
 965 970 975
 Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala
 980 985 990
 Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val
 995 1000 1005
 Ala Ser Arg Gln Asp Glu Asp
 1010 1015

10

20

30

(2) INFORMATION FOR SEQ ID NO:10:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 203 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

40

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HTLV-III

(ix) FEATURE:

(A) NAME/KEY: Protein

10

(B) LOCATION: 1..203

(D) OTHER INFORMATION: note = "sor protein of HTLV-III"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Cys Gln Glu Glu Lys Gln Arg Ser Leu Gly Ile Met Glu Asn Arg Trp

1 5 10 15

Gln Val Met Ile Val Trp Gln Val Asp Arg Met Arg Ile Arg Thr Trp

20 25 30

20

Lys Ser Leu Val Lys His His Met Tyr Val Ser Gly Lys Ala Arg Gly

35 40 45

Trp Phe Tyr Arg His His Tyr Glu Ser Pro His Pro Arg Ile Ser Ser

50 55 60

Glu Val His Ile Pro Leu Gly Asp Ala Arg Leu Val Ile Thr Thr Tyr

65 70 75 80

Trp Gly Leu His Thr Gly Glu Arg Asp Trp His Leu Gly Gln Gly Val

85 90 95

30

Ser Ile Glu Trp Arg Lys Lys Arg Tyr Ser Thr Gln Val Asp Pro Glu

100 105 110

Leu Ala Asp Gln Leu Ile His Leu Tyr Tyr Phe Asp Cys Phe Ser Asp

115 120 125

Ser Ala Ile Arg Lys Ala Leu Leu Gly His Ile Val Ser Pro Arg Cys

130 135 140

40

Glu Tyr Gln Ala Gly His Asn Lys Val Gly Ser Leu Gln Tyr Leu Ala

145 150 155 160
 Leu Ala Ala Leu Ile Thr Pro Lys Lys Ile Lys Pro Pro Leu Pro Ser
 165 170 175
 Val Thr Lys Leu Thr Glu Asp Arg Trp Asn Lys Pro Gln Lys Thr Lys
 180 185 190
 Gly His Arg Gly Ser His Thr Met Asn Gly His
 195 200

10

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 863 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

20

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HTLV-III

(ix) FEATURE:

(A) NAME/KEY: Protein

(B) LOCATION: 1..863

(D) OTHER INFORMATION: note = "env protein of HTLV-III"

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Lys Glu Gln Lys Thr Val Ala Met Arg Val Lys Glu Lys Tyr Gln His
 1 5 10 15
 Leu Trp Arg Trp Gly Trp Arg Trp Gly Thr Met Leu Leu Gly Met Leu
 20 25 30
 Met Ile Cys Ser Ala Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly
 35 40 45
 Val Pro Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp

40

50	55	60	
Ala Lys Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala			
65	70	75	80
Cys Val Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val			
85	90	95	
Thr Glu Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His			
100	105	110	
Glu Asp Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys			
115	120	125	
Leu Thr Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Lys Asn Asp			
130	135	140	
Thr Asn Thr Asn Ser Ser Ser Gly Arg Met Ile Met Glu Lys Gly Glu			
145	150	155	160
Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg Gly Lys Val			
165	170	175	
Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile Pro Ile Asp			
180	185	190	
Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr Ser Val Ile			
195	200	205	
Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr			
210	215	220	
Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn Lys Thr Phe			
225	230	235	240
Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr His			
245	250	255	
Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu			
260	265	270	
Ala Glu Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr Asp Asn Ala			
275	280	285	

10

20

30

40

Lys Thr Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile Asn Cys Thr
 290 295 300
 Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gln Arg Gly Pro
 305 310 315 320
 Gly Arg Ala Phe Val Thr Ile Gly Lys Ile Gly Asn Met Arg Gln Ala
 325 330 335
 His Cys Asn Ile Ser Arg Ala Lys Trp Asn Asn Thr Leu Lys Gln Ile
 340 345 350
 Asp Ser Lys Leu Arg Glu Gln Phe Gly Asn Asn Lys Thr Ile Ile Phe
 355 360 365
 Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile Val Thr His Ser Phe Asn
 370 375 380
 Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser
 385 390 395 400
 Thr Trp Phe Asn Ser Thr Trp Ser Thr Lys Gly Ser Asn Asn Thr Glu
 405 410 415
 Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn
 420 425 430
 Met Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile Ser Gly
 435 440 445
 Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp
 450 455 460
 Gly Gly Asn Ser Asn Asn Glu Ser Glu Ile Phe Arg Pro Gly Gly Gly
 465 470 475 480
 Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val
 485 490 495
 Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val
 500 505 510
 Val Gln Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Leu Phe Leu Gly

10

20

30

40

515 520 525
 Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Met Thr Leu
 530 535 540
 Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn
 545 550 555 560
 Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr
 565 570 575
 Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Ile Leu Ala Val Glu Arg
 580 585 590
 Tyr Leu Lys Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys
 595 600 605
 Leu Ile Cys Thr Thr Ala Val Pro Trp Asn Ala Ser Trp Ser Asn Lys
 610 615 620
 Ser Leu Glu Gln Ile Trp Asn Asn Met Thr Trp Met Glu Trp Asp Arg
 625 630 635 640
 Glu Ile Asn Asn Tyr Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser
 645 650 655
 Gln Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys
 660 665 670
 Trp Ala Ser Leu Trp Asn Trp Phe Asn Ile Thr Asn Trp Leu Trp Tyr
 675 680 685
 Ile Lys Leu Phe Ile Met Ile Val Gly Gly Leu Val Gly Leu Arg Ile
 690 695 700
 Val Phe Ala Val Leu Ser Val Val Asn Arg Val Arg Gln Gly Tyr Ser
 705 710 715 720
 Pro Leu Ser Phe Gln Thr His Leu Pro Ile Pro Arg Gly Pro Asp Arg
 725 730 735
 Pro Glu Gly Ile Glu Glu Glu Gly Gly Glu Arg Asp Arg Asp Arg Ser
 740 745 750

10

20

30

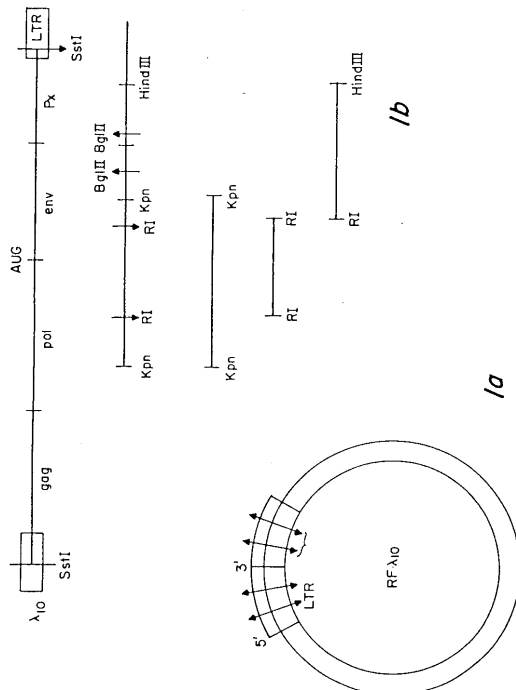
40

Ile Arg Leu Val Asn Gly Ser Leu Ala Leu Ile Trp Asp Asp Leu Arg
 755 760 765
 Ser Leu Cys Leu Phe Ser Tyr His Arg Leu Arg Asp Leu Leu Leu Ile
 770 775 780
 Val Thr Arg Ile Val Glu Leu Leu Gly Arg Arg Gly Trp Glu Ala Leu
 785 790 795 800
 Lys Tyr Trp Trp Asn Leu Leu Gln Tyr Trp Ser Gln Glu Leu Lys Asn
 805 810 815
 Ser Ala Val Ser Leu Leu Asn Ala Thr Ala Ile Ala Val Ala Glu Gly
 820 825 830
 Thr Asp Arg Val Ile Glu Val Val Gln Gly Ala Tyr Arg Ala Ile Arg
 835 840 845
 His Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg Ile Leu Leu
 850 855 860

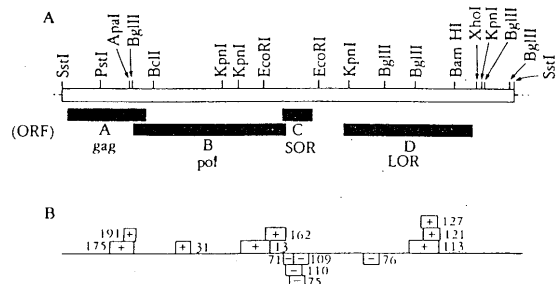
10

20

【 1 】



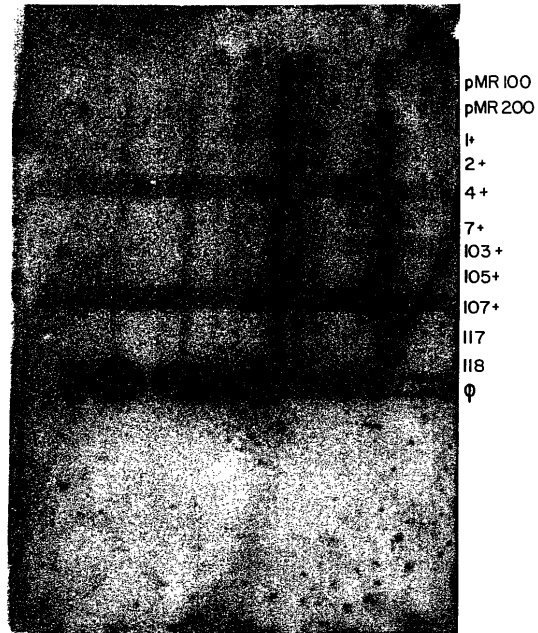
【 2 】



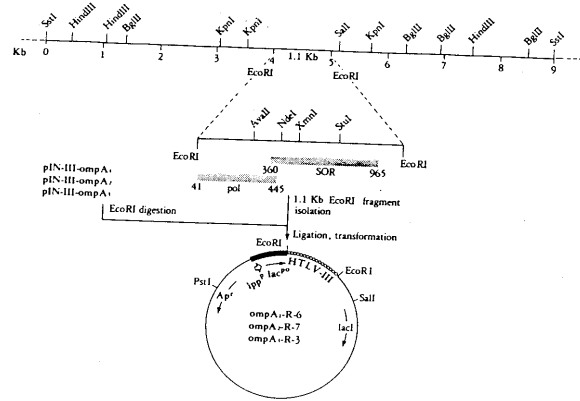
CLONE	AMINO ACID RESIDUE	NUCLEOTIDE POSITION
BH10 BH5		420
BH10 BH5		345
BH10 BH5		270
BH10 BH5		195
BH10 BH5		120
BH10 BH5		45
BH10 BH5		1
BH10 BH5		39
BH10 BH5		75
BH10 BH5		150
BH10 BH5		221
BH10 BH5		296
BH10 BH5		371
BH10 BH5		446
BH10 BH5		521
BH10 BH5		596
BH10 BH5		671
BH10 BH5		746
BH10 BH5		821
BH10 BH5		896
BH10 BH5		971
BH10 BH5		1046
BH10 BH5		1121
BH10 BH5		1196
BH10 BH5		1271
BH10 BH5		1346
BH10 BH5		1421
BH10 BH5		1496
BH10 BH5		1571
BH10 BH5		1646
BH10 BH5		1721
BH10 BH5		1796
BH10 BH5		1871
BH10 BH5		1946
BH10 BH5		2021
BH10 BH5		2096
BH10 BH5		2171
BH10 BH5		2246
BH10 BH5		2321
BH10 BH5		2396
BH10 BH5		2471
BH10 BH5		2546
BH10 BH5		2621
BH10 BH5		2696

[illegible][illegible]

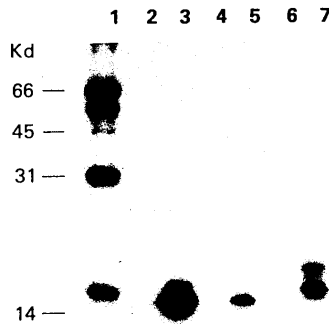
【 15 】



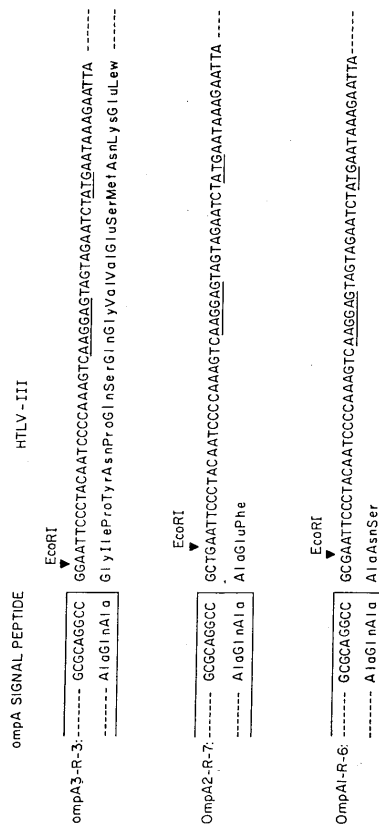
【 16 】



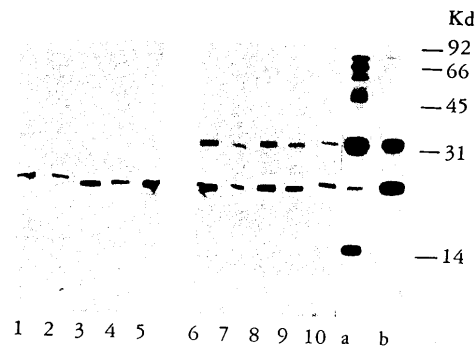
【 17 】



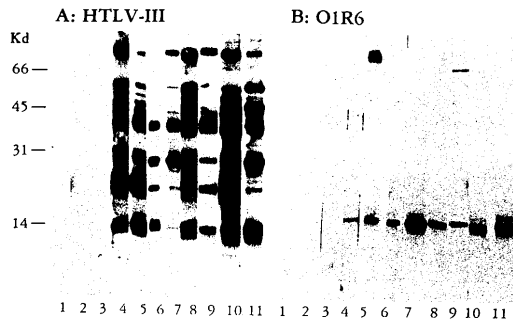
【 18 】



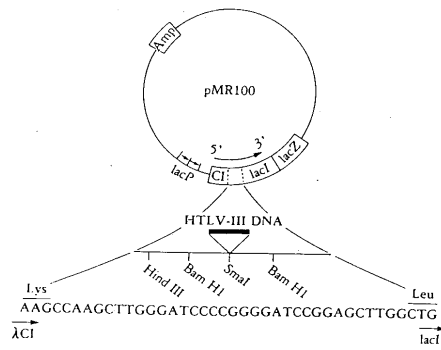
【 19 】



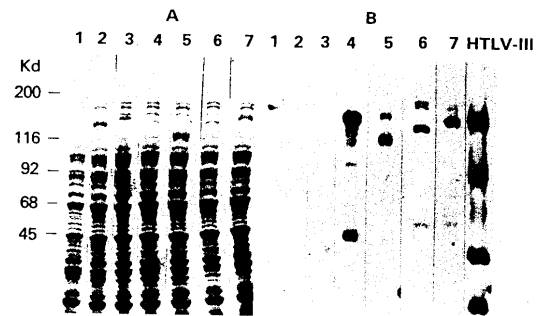
【 20 】



【 図 2 1 】



【 図 2 2 】



フロントページの続き

(51) Int.Cl.⁷

F I

G 0 1 N 33/569

G 0 1 N 33/569

H

G 0 1 N 33/60

G 0 1 N 33/60

Z

// (C 1 2 P 21/02

C 1 2 P 21/02

C

C 1 2 R 1:19)

C 1 2 R 1:19

(72)発明者 ロバート・シー・ガロ

アメリカ合衆国メリーランド州20817, ベセスダ, ソーンデン・テラス 8513

(72)発明者 フロジエ・ウォング - スタール

アメリカ合衆国カリフォルニア州92130, サンディエゴ, モントレー・サイプレス・ウェイ
12737

審査官 田村 明照

(56)参考文献 特開昭61-173782(JP, A)

特開昭55-104887(JP, A)

特表昭62-500282(JP, A)

特表昭62-500592(JP, A)

特表昭61-501431(JP, A)

专利名称(译)	HTLV-III DNA的克隆和表达		
公开(公告)号	JP3569766B2	公开(公告)日	2004-09-29
申请号	JP2002137039	申请日	2002-05-13
[标]申请(专利权)人(译)	圣码公司 森托科尔公司		
申请(专利权)人(译)	Sentoko公司		
当前申请(专利权)人(译)	Centocor公司公司		
[标]发明人	ナンシーティーチャン ロバートシーガロ フロジューウングスタール		
发明人	ナンシー・ティー・チャン ロバート・シー・ガロ フロジュー・ウング・スタール		
IPC分类号	G01N33/53 A61K38/00 C07K14/15 C07K14/155 C07K14/16 C07K14/705 C07K16/10 C07K19/00 C12N1/21 C12N9/38 C12N15/02 C12N15/09 C12N15/48 C12N15/49 C12P21/02 C12P21/08 C12Q1 /68 C12Q1/70 C12R1/19 G01N33/543 G01N33/566 G01N33/569 G01N33/577 G01N33/58 G01N33/60		
CPC分类号	C12Q1/703 A61K38/00 C07K14/005 C07K16/1054 C07K16/1063 C07K2319/40 C07K2319/61 C07K2319/80 C12N2740/14022 C12N2740/16122 C12N2740/16222 C12Q1/6813		
FI分类号	C12N15/00.ZNA.A C07K14/16 C12P21/02.C G01N33/53.M G01N33/566 G01N33/569.H G01N33/60.Z C12R1/19 C12N15/00.A C12N15/00.AZN.A C12Q1/68.A C12Q1/6813.C C12Q1/6813.Z		
F-TERM分类号	2G045/AA28 2G045/BA13 2G045/BB03 2G045/BB10 2G045/BB16 2G045/BB20 2G045/CA25 2G045 /CA26 2G045/CB01 2G045/CB07 2G045/CB14 2G045/DA12 2G045/DA13 2G045/DA14 2G045/DA36 2G045/FB02 2G045/FB03 2G045/FB04 2G045/FB08 4B024/AA01 4B024/AA14 4B024/BA14 4B024 /CA03 4B024/CA04 4B024/CA07 4B024/CA09 4B024/CA12 4B024/DA06 4B024/EA04 4B024/FA04 4B024/FA10 4B024/GA11 4B024/HA08 4B024/HA12 4B063/QA01 4B063/QQ03 4B063/QQ10 4B063 /QQ53 4B063/QR32 4B063/QR55 4B063/QS33 4B063/QS34 4B064/AG32 4B064/CA02 4B064/CA19 4B064/CC24 4B064/DA01 4H045/AA10 4H045/AA20 4H045/BA10 4H045/BA41 4H045/CA04 4H045 /DA75 4H045/EA29 4H045/EA52 4H045/FA74		
优先权	06/659339 1984-10-10 US 06/693866 1985-01-23 US		
其他公开文献	JP2003038190A		
外部链接	Espacenet		

摘要(译)

要解决的问题：基于克隆HTLV-III DNA，提供一种有助于艾滋病的诊断，治疗和预防的方法。解决方案：本发明涉及由含有HTLV-III cDNA的重组载体转化的细胞表达的免疫反应性HTLV-III多肽，产生该多肽的方法，编码HTLV-III基因的分离的cDNA，由a组成的DNA探针。DNA序列与HTLV-III病毒特异性的HTLV-III基因组的一部分基本上同源，并且通过将HTLV-III多肽与一种或多种其他多肽结合而获得的杂合蛋白质。

【 2 】

