

Handwritten signature or initials.

92.844

MEMORIA DESCRITIVA

Resumo

O invento divulga a sequencia de aminoácidos de Hidrolase de Acil-Peptideos. O invento está relacionado com uma molécula de DNA codificadora da Hidrolase de Acil-Peptideos, um vector que inclui a molécula e a utilização de tal vector para transformar um hospedeiro.

=====

THE GENERAL HOSPITAL CORPORATION

"METODO PARA A PRODUÇÃO E UTILIZAÇÃO DE HIDROLASE DE ACIL-PEPTIDEOS"

Handwritten signature or initials.

O invento está ainda relacionado com a expressão pelo hospedeiro para produzir Hidrolase de Acil-Peptídeos e com o uso de APH para catalisar a hidrólise de um peptídeo ou proteínas N^α-acetilado ou a reação entre um dador de N^α-acetil-aminoácidos e uma proteína aceitadora com um grupo alfa-NH₂ livre. O invento está ainda relacionado com a utilização das sequências codificadoras da Hidólise de Acil-Peptídeos na detecção de cancro.

Referencia Cruzada a pedidos de patente relacionados:

Este pedido é uma continuação em parte do Pedido de Patente U.S.No.de serie 07/296,996 (entregue em 13 de Janeiro de 1989), o qual é uma continuação em parte do Pedido de Patente U.S.No. de Serie 07/087,936 (entregue em 21 de Agosto de 1987), ambas aqui incluídas por referencia.

Campo do Invento

O presente invento é dirigido à produção de Hidrolase de Acil-peptídeos por tecnologia de DNA recombinante. Ele também se dirige à utilização de enzima para catalisar a hidrólise de um peptídeo ou proteína acilado e à reação entre um dador de N^ω-acetil-aminoácido derivatizados e a uma proteína aceitadora com um grupo alfa-NH₂ livre. O invento diz ainda respeito a uma sequência genética que codifica a hidrolase de acil-peptídeos de rato. O invento também se dirige ao diagnóstico do carcinoma de células pequenas através da utilização da hidrolase de acil-peptídeos e de sequências de genes que codificam a hidrolase de acil-peptídeos.

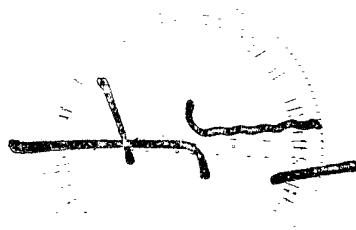
Breve descrição de trabalhos anteriores:

Desde a descoberta de um grupo acetilol no extremo amina da proteína da capsíde do vírus do mosaico do tabaco, foram encontrados uma série de proteína N^α-acetiladas em animais, plantas e seus vírus e também em bacterias e fungos.

A N^α-acetilação é portanto considerada como uma das modificações típicas de proteínas em organismos vivos. Ainda, nalgumas células eucarióticas, foi sugerido que mais de 80% das proteínas intracelulares sejam N^α-acetiladas (Brown, J.L. J.Biol.Chem. 254:1447-1449 (1979)).

O significado biológico da N^α-acetilação das proteínas é ainda uma questão em aberto (ver Tsunasawa et al, Method Enzymol. 106: 165-170 (1984)). Foi proposta que esta modificação pós-tradução protege as proteínas intracelulares da proteólise. No entanto, não se verifica para todas as proteínas. No caso da actina do "slime mold", a degradação proteolítica torna-se mais lenta quando a proteína é N^α-acetilada. Pelo contrário, a hemoglobolína do gato é degradada à mesma velocidade independentemente da N^α-acetilação (Tsunasawa et al, 1984)

Resultados recentes de sequenciação de DNA mostraram que nos genes estruturais das proteínas secretadas que são N^α-acetiladas o códon do resíduo amino terminal acetilado é directamente precedido pelo códon de



iniciação sem a inserção de códons adicionais para os aminoácidos (Tsunasawa et al, 1984). Fez-se pouco esforço para compreender a relação entre N^α-acetilação e o transporte de proteínas secretadas de membranas biológicas. Para compreender completamente a função de N^α-acetilação será importante identificar os aminoácidos N^α-acetilados nas proteínas e peptídeos numa escala microanalítica. Com esta finalidade a remoção do grupo N^α-acetilo do N^α-acetil-aminoácido deve ser eficazmente conseguido.

A Hidrolase de Acil-peptídeos (APH) tem sido usada com êxito para a hidrólise de aminoácidos N^α-acilados. Tal enzima que foi purificada a partir de fígado animal, pode libertar o N^α-acetil-aminoácido de peptídeos relativamente pequenos derivados de proteínas N^α-acetiladas (Tsunasawa et al, 1984). A especificidade do substrato é larga para o resíduo amino-terminal. APH cliva o aminoácido N^α-terminal acetilado ou formilado de um peptídeo bloqueado (Jonas et al, BBRC 126:933 (1985)). Esta enzima catalisa a hidrólise de um número diverso de peptídeos e apresenta diferentes pHs ótimos para certos substratos ao fazê-lo. Esta enzima pode também desempenhar papel "pivot" no procedimento de cadeias polipeptídicas durante a biossíntese.

A PH foi purificada a partir de fígado de rato (Tsunasawa et al, J. Biochem. 77, 89-102(1975) / a partir do fígado bovino (Gade et al, Biochim. Biophys Acta 662:86-83 (1981)); a partir do fígado porcino (Tsunasawa et al, J. Biochem. 93:1217-1220 (1983)); a partir de cérebro de rato (Marks et al, J. Neurochem. 41: 201-208 (1983)); e a partir de eritrócitos humanos (Jones et al, Biochem. and Biophys. Res-Comm, 126:933-940 (1985)).

Uma hidrolase de acil-peptídeos de fígado de rato (APH) que catalisa a hidrólise de resíduo acetilado a partir de peptídeos N^α-acetilados foi recentemente purificada até à homogeneidade e várias experiências de inibição indicaram que é provavelmente uma protease serina, utilizando um sistema de substituição de cargas envolvendo serina, histidina e provavelmente um grupo carboxilo (Kobayashi, K. e Smith, J.A. J.Biol.Chem.262: 11435-11445 (1987)). No entanto, não está ainda claro se a hidrolase de acil-peptídeos é uma protease serina única.

Para facilitar uma compreensão mais completa de regulações de hidrolase de acil-peptídeos de rato in vivo, é portanto, desejável clonar e sequenciar o gene da hidrolase de acil-peptídeos.

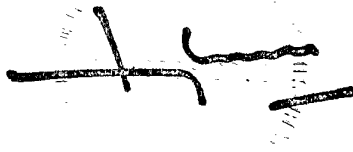
Sumario do Invento

A hidrolase de acil peptídeos catalisa a hidrólise de um resíduo de aminoácido N^α-acetilado. de um peptídeos N^α-acetilado. Duas sondas oligonucleotídicas degeneradas e sobrepossíveis baseadas na sequência de um peptídeo triptico, derivado da hidrolase de acil-peptídeos de rato purificada, foram sintetizadas e usadas no despiste de uma biblioteca de cDNA derivado de fígado de rato em λgt11. Um cDNA de 2,5 kb foi clonado e sequenciado. Este clone continha 2364pb da sequência da hidrolase de acil-peptídeos de rato mas não possuía um codão de iniciação. Usando uma sonda de 220pb derivada do extremo 5'

deste cDNA de quase tamanho completo para fazer novo despitte da biblioteca, isolaram-se clones de tamanho completo, os quais continham um codão ATG na mesma grelha de leitura nos nucleotídeos 6-8 e codificadores da sequência NH₂-terminal, Met-Glu-ArgpGln...

A sequência de DNA codificava uma proteína de 732 resíduos de aminoácidos, 80% dos quais foram confirmados por resultados da sequência proteica de 19 peptídeos CNBr ou tripticos. A enzima isolada está bloqueada no extremo NH₂ (Kobayashi, K, e Smith J.A.(1987). J.Biol.Chem. 262 : 11435-11445). e com base na sequência proteica NH₂-terminal deduzida a partir de sequência de DNA e na sequência do peptídeo CNBr do extremo NH₂, é provável que o resíduo NH₂-terminal seja um resíduo de metionina acetilado, uma vez que três resíduos estão frequentemente justapostos aos resíduos glutâmico (Person, B.et al (1985) Eur.J.Biochem. 152, 523-527). A análise da transferência de RNA revelou um único mensageiro de 2,7 kb em vários tecidos de rato examinados. Sabem que se saiba que esta enzima é inibida por fluorofosfato de diisopropilo e por acetilalanina-clorometilcetona (Kobayashi, K. e Smith, J.A.(1987), J.Biol.Chem.262:11435-11445), não se encontrou qualquer semelhança forte na sequência proteica com outras proteases serina. Este resultado sugere que a hidrolase de acil-peptídeos poderá ser uma protease serina diferente.

Este invento é dirigido a uma proteína. Hidrolase de Acil-peptídeos (APH), a qual se caracteriza pela sequência de aminoácidos de Figura 1. Ele também é dirigido à produção de APH por tecnologia de DNA recombinante e à utilização de APH na hidrólise ou aminoacidificação de



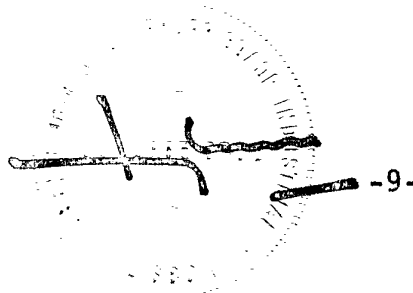
ou proteínas.

O invento diz também respeito à clonagem e análise da sequência de uma hidrolase de acil-peptídeos derivada de fígado de rato descrita por Kobayashi K. et al, (J. Biol.Chem. 264:8892-8899) (Maio, 1989) cuja referência é aqui incluída por referência.

Uma molécula de DNA recombinante codificadora de APH do presente invento pode ser usada para transformar qualquer um de uma série de hospedeiros, criando novas fontes de fornecimentos ilimitados de APH.

O invento ainda compreende as seqüências genéticas codificadoras de uma enzima tendo a seqüência de aminoácidos designada na Figura 1, veículos contendo a sequência genética, hospedeiros transformados com eles, produção da enzima por expressão no hospedeiro transformado e utilização da enzima na hidrólise ou na amino-âcilação de peptídeos ou proteínas. É uma finalidade deste invento proporcionar novas fontes de APH substancialmente pura de que se disporia numa fonte ilimitada.

Em adição este invento engloba a utilização da enzima para catalisar a hidrólise de uma proteína N α -acilada ou a reação entre um dador de N α -acetil-aminoácidos e uma proteína aceitadora com um grupo alfa-NH₂ livre.



Portanto, são finalidades adicionais deste invento proporcionar num meio de hidrolise de proteínas N^ε-aciladas e de amino-acilação de qualquer polipeptideo ou proteina a partir de um dador de N^ε-acetilaminocidos e de um aceitador com um grupo alfa-NH₂, livre usando a APH.

Detalhadamente, o invento diz respeito a hidrolase de Acil-peptideos na forma pura. O invento tambem diz respeito a Hidrolase de Acil-peptideos sem glicosilação nativa.

O invento ainda diz respeito a uma molecula de acido nucleico recombinante, RNA, DNA genômico ou cDNA, a qual contem uma sequencia genetica codificadora da Hidrolase de Acil-peptideos. A molecula de acido nucleico pode ser um vector ou plasmideo.

O invento também diz respeito a um hospedeiro como seja uma bacterias, uma levedura ou uma celula de mamifero, etc., transformada com qualquer uma das moleculas de acido nucleico recombinante atrás descritas.

O invento tambem diz respeito a um método para a produção da Hidrolase de Acil-peptideos que se caracteriza por:



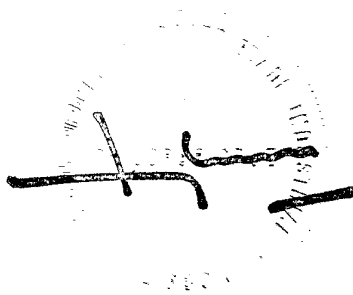
- (a) obtenção de qualquer uma das moléculas de ácido nucleico atrás descritas, em que a molécula é DNA;
- (b) inserção da molécula de DNA num vector;
- (c) transformação de um sistema hospedeiro com o vector;
- (d) expressão da sequência de DNA da hidrolase de Acil-peptídeos da molécula de DNA recombinante no hospedeiro; e
- (e) recuperação da hidrolase de Acil-peptídeos produzida pela expressão.

O invento também inclui a Hidrolase de Acil-peptídeos produzida pelo método atrás descrito.

O invento também inclui a Hidrolase de Acil-peptídeos atrás descrita na forma imobilizada.

O invento também inclui um método de hidrólise do acil-aminoácido N-terminal de um polipeptídeo acilado, o qual compreende o contacto do polipeptídeo com a Hidrolase de Acil-peptídeos atrás descrita.

O invento também inclui um método de catálise de reacção entre um dador de N^α-acetil-aminoácidos derivatizado e um aceitador com um grupo alfa-NH₂ livre que compreende o contacto do dador como aceitador na presença da Hidrolase de Acil-peptídeos atrás descrita.



O invento também engloba um método de detecção do carcinoma de células pequenas que se caracteriza por:

a. incubação de uma amostra de ácido nucleico de um paciente suspeito de ter carcinoma de células pequenas, na presença de uma molécula de ácido nucleico tendo uma sequência seleccionada do grupo constituído por:

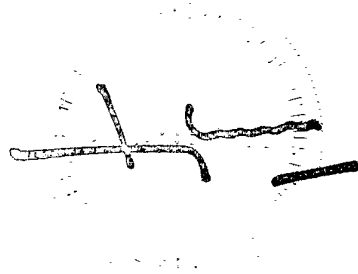
a: uma sequência que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos; e

b. uma sequência que é complementar de uma sequência que codifica toda ou parte da enzima hidrolase de acil-peptídeos; a incubação sendo em condições suficientes para permitir a hibridação de ácidos nucleicos entre a amostra de ácido nucleico e a molécula de ácido nucleico e para assim formar uma molécula híbrida; e

b. detecção como seja por uma análise de polimorfismo de comprimentos de fragmentos de restrição, de carcinoma das células pequenas determinando se a molécula híbrida difere na sequência de uma molécula referência compreendendo uma amostra de ácido nucleico de um indivíduo normal hibridada com uma molécula de ácido nucleico que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos.

O invento ainda inclui uma molécula de ácido nucleico de cadeia dupla compreendendo:

A. uma primeira cadeia tendo uma sequência seleccionada de um grupo constituído por:



a. uma sequência que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos; e

b. uma sequência que é complementar de uma sequência que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos; a primeira cadeia usa e sendo hibridade com:

B. uma segunda cadeia, a segunda cadeia tendo uma sequência que é substancialmente complementar em sequência da sequência da primeira cadeia, a sequência complementar da segunda cadeia derivada de um indivíduo suspeito de ter carcinoma das células pequenas.

Descrição das Figuras

A Figura 1 ilustra a sequência de aminoácido de APH. A sequência proteica deduzida da sequência de cDNA (Figura 3) está indicada pelo código de uma única letra para os aminoácidos. As linhas entre parêntesis indicam os extremos dos peptídeos CB, CB-R e CB/R.

As setas que apontam para a direita indicam que o correspondente resíduo de aminoácido foi identificado como o resíduo de aminoácido PTH durante a degradação automática de Edman (Tabela 1). Um "blank" indica que um aminoácido PTH não foi identificado neste ciclo degradativo. Um asterisco indica que um Pth-Trp juntamente com um derivado Pth não identificado e que elui uma fase tardia foi identificado em vez de Pth-Lis durante este

ciclo degradativo. Os resíduos cisteína foram identificados como derivados Pth de γ - ^{14}C 7S-carboximetilcisteína. A serina activa está apresentada nas posições 620-627 da Figura 1 (caixa preenchida com diagonais). A descrição dos peptídeos apresentados aqui está definido na Tabela 3.

A Figura 2 ilustra a clonagem e sequenciação de cDNA codificador da hidrolase de acil-peptídeos de rato.

(A) Sondas oligonucleotídicas usadas para o despiste inicial da biblioteca de cDNA de fígado de rato em λ gt11. A sequência de aminoácidos de um peptídeo triptico purificado por RPLC (CB18-R11-13C; Tabela 3) foi usada como base para a síntese de dois oligonucleotídeos degenerados sobreponíveis YS17.2 e YS20.1.

(B) Mapa de restrição e estratégia de sequenciação do DNA dos clones: Usando os oligonucleotídeos degenerados na Figura 2A, APH5.2 foi obtido a partir de uma biblioteca de cDNA de fígado de rato em λ gt11, como descrito abaixo. As setas indicam a direcção e grau de determinação da sequência de DNA para cada fragmento. A análise da sequência de DNA para este clone revelou o sítio de hibridação esperado perto do seu extremo 5' (região aberta em "bold line") uma sequência poli(A), no seu extremo 3' e uma sequência não relacionada no seu extremo 3' (caixa tracejada).

Após novo despiste da biblioteca de cDNA de fígado de rato em λ gt11 com o fragmento XmnI-KpnI derivado de APH5.2 (caixa aberta maior). A PH 36.1 não

possuia um codão de iniciação ATG e também continha um fragmento de 120 pares de bases codificador da albumina de soro de rato (caixa de/com diagonais). Após novo despeite da biblioteca de/com um fragmento BamII-PstII de 220pb derivado de APH 36.1 (caixa pequena aberta), APH 2.7 foi clonado, o qual foi subsequentemente subclonado no plasmídeo Bluescript (Stratagene) e sequenciado em parte.

Abreviaturas: B, BamI; P, PstI;
X, XmnI; e K, KpnI.

A Figura 3 mostra a sequencia de nucleotídeos e a sequencia de aminoácidos deduzida da hidrolase de acil-peptídeos de fígado de rato. O cDNA completo codificador da hidrolase de acil-peptídeos de fígado de rato foi derivado combinando os resultados da sequencia de DNA de APH 26.1 e APH 2.7 (Figura 2B).

A sequencia proteica deduzida está indicada pelo código de uma letra para os aminoácidos.

A Figura 4 mostra a sequencia de nucleotídeos do gene da hidrolase de acil-peptídeos de rato e a sua região delimitante. O sítio de iniciação da transcrição do gene está indicado pela seta vertical. O nucleotídeo nesta posição está assinalado com o número. A sequencia de DNA intrônica está apresentada em letras minúsculas e a sequencia de DNA exônico está apresentado em letras minúsculas.

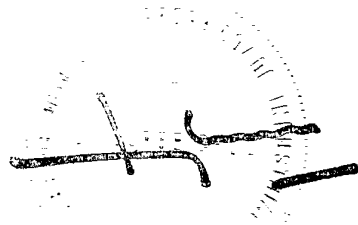
O começo e o fim de cada intrão estão marcados por linhas verticais. O sítio de iniciação da tradução está situado nos nucleotídeos 625-627.

O sinal de poliadenilação está situado nos nucleotídeos 9708-9713. A sequência tipo caixa "TATA" (nucleotídeos 248 -30) e a sequência tipo caixa CAAT" (nucleotídeos 95 a -99) estão dentro de retângulos. As repetições GC estão sublinhadas. As repetições de 200pb em tandem estão sublinhadas a tracejado.

A figura 5 mostra uma organização estrutural do gene da hidrolase de acil-peptídeos de rato. A Figura 5A mostra fagos recombinantes sobreponíveis contendo o gene da hidrolase de acil-peptídeos. Os clones genômicos sobreponíveis, APHE5 e APHE6, tendo em conjunto todo o gene da hidrolase de acil-peptídeos, estão indicados por linhas horizontais e cheio.

A Figura 5B mostra o mapa de restrição do gene da hidrolase de acil-peptídeos e as suas regiões delimitantes. Os sítios EcoRI(E), BamHI(B), HindIII(H) e PstI(P) estão indicados por bases verticais. Esta apresentada a orientação da transcrição de 5'(esquerda) para 3'(direita) deste gene.

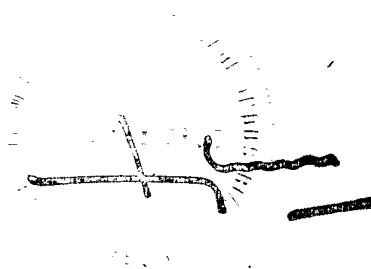
A Figura C mostra a organização exão-intrão do gene de hidrolase de acil-peptídeos de rato. A localização dos 23 exões dentro do gene da hidrolase de acil-peptídeos de rato está indicada por retângulos a cheio. As localizações do códon de iniciação da tradução, ATG, e o



sinal de poliadenilação, AATAAA, estão marcados por linhas verticais.

A Figura 6 mostra uma comparação das sequencias de aminoácidos da hidrolase de acil-peptídeos e das proteínas DNF 1552.

Nas Figuras, os aminoácidos foram designados por uma unica letra do alfabeto:
A= Alanina, B= Acido Aspártico ou Asparagina,
C = Cisteina , D= Acido Aspártico, E = Acido Glutâmico,
F = Fenilalanina, G = Glicina, H = istidina; I = Isoleucina, K = lisina, L = Leucina, M = Metionina, N = Asparagina. P = prolina, Q = Glutamina, R = Arginina, S = Serina, T = Treonina, V = Valina, W = Triptofano, Y = Tirosina, Z = Glutamina ou Acido Glutâmico.



Discussão Detalhada do Invento

Definições

Para ajudar na compreensão da especificação e reivindicações, incluindo o âmbito a ser dado a tais termos, são dadas as seguintes definições.

Transcrição

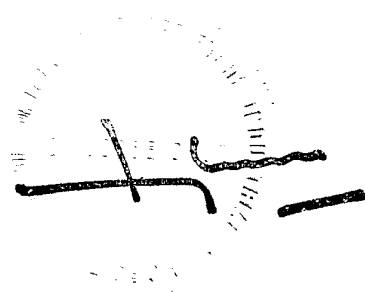
O processo de produção de mRNA a partir de um gene estrutural.

Tradução

O processo de produção de um polipeptídeo a partir de mRNA.

Expressão

O processo sofrido por um gene estrutural para produzir um polipeptídeo. É uma combinação de transcrição e tradução.

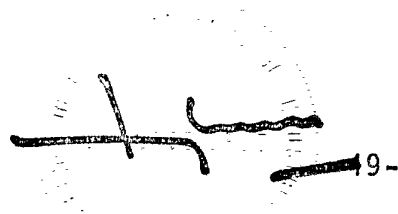


Plasmídeo

Uma molécula de DNA circular de cadeia dupla que não faz parte do cromossoma principal de um organismo contendo genes que conferem resistência a anticorpos específicos. Quando o plasmídeo é colocado dentro de um organismo unicelular, as características deste podem ser alteradas ou transformadas caso resultado do DNA do plasmídeo. Por exemplo, um plasmídeo portador do gene da resistência à tetraciclina (Tet^R) transforma uma célula previamente sensível à tetraciclina numa que lhe é resistente. Uma célula transformada por um plasmídeo é designado um "transformante".

Veículo de clonagem

Um plasmídeo, DNA fágico ou outras sequências de DNA que são capazes de se replicar numa célula hospedeira. O veículo de clonagem é caracterizado por um ou um pequeno número de sítios de endonuclease de restrição nos quais tais sequências de DNA podem ser cortadas de modo determinável sem perda de uma função biológica, essencial do DNA, o qual pode conter uma marca adequada para usar na identificação de células transformadas. São marcas, por exemplo a resistência à tetraciclina ou à ampicilina. Um veículo de clonagem é por vezes designado num vector.



Moléculas de DNA recombinante ou DNA híbrido

Uma molécula consistindo em segmentos de DNA de diferentes genomas que foram ligados extremo a extremo fora das células vivas e que tem a capacidade de infectar algumas células hospedeiras e nelas se manter.

Operador

Uma sequência de DNA capaz de interagir com o repressor específico, controlando assim a transcrição de gene (s) adjacente(s).

Promotor

Uma sequência de DNA em que a RNA polimerase se liga e inicia a transcrição de um gene(s) adjacente(s).

Hidrolase de Acil-peptídeos (APH).

Este termo pretende, incluir uma (s) hidrolase(s) de acil-peptídeos de qualquer espécie, a qual tem actividade de libertação do aminoácido acilado N α -terminal de qualquer proteína de peptídeo num sistema in vivo ou in vitro. O termo hidrolase de acil-peptídeos é também usado neste invento para incluir qualquer analogo, homologo, mutante ou derivado de uma hidrolase de acil-peptídeos natural a qual cliva o aminoácido N α -acetilado de

fracção N^o-acetilado da fracção N^o-terminal de um peptideo ou proteina. O termo pretende tambem incluir fragmentos tendo menor numero de aminoácidos do que o natural, como sejam fragmentos parciais de hidrolases de acil-peptideos naturais que retem a actividade de clivagem do aminoácido acilado do extremo de N-terminal de uma proteina a/ou peptideo. O termo é tambem usado para incluir qualquer produto que compreenda a sequencia de uma hidrolase de acil-peptideos natural ou um seu homologo, juntamente com um ou mais aminoácidos delimitantes que apresente actividade de hidrolase de acil-peptideos. O termo hidrolase de acil-peptideos tambem inclui sinónimos tais como factor de libertação de acil-aminoácidos, enzima libertadora de acil-aminoácidos hidrolase de acil-aminopeptideos e acetilaminoacil-p-nitro anilidase,

Forma substancialmente pura

Tal como aqui é usado, o termo "substancialmente pura" ou "substancialmente purificada" pretende descrever a proteina que está substancialmente livre de qualquer composto normalmente associado com o factor no seu estado natural. O termo pretende ainda descrever que é homogéneo por um ou mais características de pureza ou homogeneidade normalmente usadas. Por exemplo, num factor substancialmente puro apresentará característica constantes e reprodutíveis com desvios padrão experimentais para parâmetros tais como os que se seguem: peso molecular, tecnicas cromatograficas e outros parametros do mesmo tipo. O termo, no entanto, não pretende excluir misturas artificiais ou sinteticas do factor com outros compostos. O termo não pretende incluir /excluir a presença de impurezas existentes em quantidades muito pequenas que não interferem com

a actividade biologica do factor e que podem estar presentes por exemplo devido a purificação incompleta.

O peso moldecular da APH de figado de rato, conforme calculado por filtração em gel, é 290 000 - 320 000. Parece possuir quatro subunidades identicas, com uma serina activa por subunidade. O extremo N^α da APH é acilado. APH parece ser uma protease-serina, com um sistema de substituição de cargas envolvendo serina, histidina e grupos carbonilo. A serina activa está mostrada nas posições 620-627 da Figura 1 (caixa preenchida com diagonais). A sequencia de aminoácidos deste sitio é MGGSHGGF.

O ambiente do sitio activo defere de outras proteases da familia tripsina, devido à presença de histidina e à ausência de ácido aspártico. Se bem que APH apresente uma larga especificidade para substractos, ela cliva peptídeos contendo Ac-Ala-, Ac-Ser- e Ac-Met-(os residuos N-terminais mais vulgarmente acilados) mais eficazmente do que outros dipeptídeos acilados. APH tem uma actividade baixa ou nenhuma para peptídeos contendo Ac-Trp-, Ac-Asp-, Ac-Glu, Ac-Arg-, Ac-Fen e Ac-Pro.

A Hidrolise de Acil-peptídeos (APH) deve-se distinguir da N^α-acetil transferase, a qual catalisa a reacção em que uma proteina aceita o grupo acetilo de um acetil-CoH (Tsanasawa et al. Methods in Embriology 106:165-170(1984)). A Hidrolase de Acilpeptídeos deverá ser distinguida da Amino-Acilase (Szajami, Acta Biochim. et Biophys. Acad. Sci. Hung. 15:223-228(1980))

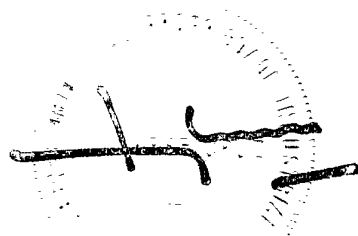
[também conhecida como hidrolase de alfa-N-acilaminoácidos (Gade et al, Biochim. Biophys. Acta 662:86-93(1981)7.

Se bem que APH tenha sido isolada e purificada a partir de varias fontes, atéagora não foi feita sequenciação de APH. O presente invento descreve aquela sequencia (Figura 1).

A sequencia de DNA codificado-ra de APH pode ser derivada de uma unidade de variedade de fontes. Por exemplo, mRNA codificador de APH pode ser isoaldp a partir dos tecidos de qualquer especie que produz APH, usando o metodo de transferencia Northern (Alwine et al, Method. Enzymol. 68:220-242 (1979) e sondas oligonucleotídicas marcadas.

O mRNA pode então ser conver_tido em cDNA por técnicas conhecidas. As sondas podem ser sintetizadas com base na sequencia de aminoácidos conhecidas de peptídeos APH.

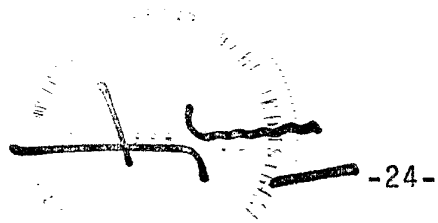
Como alternativa, sondas de ADN degenerativas podem ser usadas para fazer o despiste de uma biblioteca de DNA de uma especie que produz APH, isolando assim um clone que contem a sequencia de DNA codificadora de APH. A biblioteca de DNA foi criada cortando o DNA genômico com uma ou mais endonucleases de restrição, seguido de incorporação em vectores de sua utilização para transformar células hospedeiras as quais são então semeadas e despistadas.



A sonda de DNA pode ser marcada com um grupo detectável. Tal grupo detectável pode ser qualquer material tendo uma propriedade física ou química detectável. Tais materiais foram bem desenvolvidos no campo dos imunoenaios e em geral qualquer marca util em tais métodos pode ser aplicada ao presente invento.

São particularmente uteis os grupos enzimaticamente activos, como sejam enzimas (ver Clin. Chem. 22:1243 (1976), substractos de enzimas (ver E.p. Pat Britânica 1548 741) coenzimas (ver Pat.US Nos. 4230 797 e 4 238 565) e inibidores de enzimas (ver Pat.U.S. No. 4 134 792); compostos fluorescentes (ver Clin. Chem. 25 353 (1979)) cromóforos; substancias luminescentes tais como quimioluminescentes e bioluminescentes (ver Clin. Chem. 25: 512(1979)), ligando especificamente ligaveis; pares de interacção peroximal; e radioisótopos, tais como ^3H , ^{35}S , ^{32}P , ^{125}I e ^{14}C .

Tais como arcas e pares de marcação são detectados em base nas suas proprias propriedades físicas (e.g. compostos fluorescentes, cromóforos e radioisótopos) ou na sua reactividade ou propriedades de ligação (e.g. enzimas, substractos, coenzimas e inibidores). Por exemplo, uma sonda marcada com cofactor pode ser detetada pela adição da enzima para a qual a marca é um cofactor e um substracto da enzima. Por exemplo, pode-se usar uma enzima que actue num substracto para gerar um produto com uma propriedade física mensurável. Exemplos desta ultima incluem mas não está limitada a β -lactosidade fosfatase alcalina e peroxidase.



Uma sequência de DNA codificadora de APH pode ser recombinada com DNA vector de acordo com técnicas convencionais, incluindo a ligação de extremos coesivos ou coesivos para ligação, digestão com enzimas de restrição para se obter os extremos apropriados, preenchimento dos extremos coesivos conforme adequado, tratamento com fosfatase alcalina para evitar ligações indesejáveis e ligação com ligases apropriadas.

Para expressar APH são necessárias sinais de transcrição e tradução reconhecidos por um elemento do hospedeiro adequado. Os hospedeiros podem ser células de mamífero capazes de serem cultivadas in vitro, particularmente leucócitos, mais particularmente células de mieloma ou outros linfócitos transformados ou oncogênicos, e.g., células transformadas por EBV.

Como alternativa, podem ser empregues células que não as de mamífero, tais como bacterias, fungos, e.g. levedura, fungos filamentosos ou similares.

São hospedeiros possíveis para a produção de APH células de mamífero, cultivadas in vitro em culturas de tecidos ou in vivo em animais. As células de mamífero podem proporcionar modificações pós-tradução a moléculas APH incluindo enrolamento ou glicosilação correctas nos sítios certos. As células de mamíferos que podem ser úteis como hospedeiros incluem células de origem fibroblástica tais como VERO ou CHO-K1 ou células de origem linfóide, tais como o híbrido SPa/O AG14

ou o mieloma P3x63 Sgh e seus derivados.

Geralmente a construção APH será parte de um vector tendo um sistema de replicação reconhecido pela células hospedeira.

Numa realização preferida, uma célula procariótica é transformada por um plasmídeo portador do gene codificador de APH. Os hospedeiros bacterianos de particular interesse incluem E.coli K12 estirpe 294 (ATCC 31446), E.coli X1776 (ATCC31537), E.coli W3110; (F. Lambda⁻, proptotrófica (ATCC 27325) e outras enterobactereáceas, tais como Salmonella Typhimurium ou Serratia marcescens, e várias espécies de Pseudomonas. Em tais condições, a APH não será glicosilada. O hospedeiro procariótico deve ser compatível com o replicação e sequências de controle no plasmídeo de expressão. Um hospedeiro procariótico com um plasmídeo contendo o cDNA codificador de APH foi depositado em 21 de Agosto, 1987 na American Type Culture Collection, Rockville, M.D. USA, e dado o número de acesso ATCC 67504.

Em geral, tais vectores contendo replicação e sequências de controle que derivam de espécies compatíveis com uma célula hospedeira são usados em ligação com o hospedeiro. O vector normalmente é portador de um sítio replicação assim como de genes específicos que são capazes de proporcionar selecção fenotípica em células transformadas. A expressão do DNA codificador de APH pode ser também colocado sob o controle de outras sequências reguladoras que podem ser homologas do organismo no seu estado não transformado.

Por exemplo, DNA cromossômico de E.coli dependente de lactose compreende um operão lactose ou lac que medeia a utilização de lactose ao elaborar a enzima B-galactosidase. Os elementos de controle lac podem ser obtidos a partir de bacteriófago lambda plac5, o qual é infeccioso para E.coli. O sistema promotor-operador lac pode ser induzido por IPTG.

Outros sistemas promotor (operador ou porções deles podem igualmente ser empregues. Por exemplo podem ser usados colicina E1, galactose, fosfatase alcalina, triptofano, tax e similares.

Para um hospedeiro mamífero existem disponíveis para expressão vários sistemas vector possíveis. Uma classe de vectores utiliza elementos de DNA que proporcionam plasmídeos extracromossômicos de replicação autónoma, derivados de vírus animais tais como vírus do papiloma bovino, vírus polioma, adenovírus ou vírus SV40. Uma segunda classe de vectores baseia-se na integração das sequências de genes pretendidas no cromossoma hospedeiro.

As células que têm estávelmente integrado o DNA introduzido nos seus cromossomas pode ser seleccionado introduzindo também uma ou mais marcas que permitam a selecção de células hospedeiras que contem o vector de expressão. A marca pode proporcionar prototrofia a um hospedeiro auxotrófico, resistência a biocidas e.g. antibióticos ou metais pesados, tais como cobre ou similares,

O gene da marca seleccionavel pode ser directamente ligado à sequencias de DNA a serem expressas ou introduzidas na mesma celula por cotransformação. Podem ser igualmente necessarios elementos adicionais para a sintese optima de mRNA. Estes elementos podem incluir sinais de "splicing", assim como promotores de transcrição., potenciadores e sinais de terminação. Os vectores de expressão e cDNA que incorporam tais elementos incluem os descritos por Okayama, H. Mol. Cel. Biol. 3: 280(1983) e outros.

Podem ser empregues uma gran de variedade de sequencias reguladoras da transcrição e tradução, dependendo da natureza 17 do hospedeiro. Os sinais de transcrição e tradução podem derivar de fontes de virais, tais como adenovirus, virus do papiloma bovino, virus simio ou similares, sempre que os sinais reguladores estejam associados a um gene particular que tenha um nivel elevado de expressão. Como alternativa podem ser empregues promotores de produtos de expressão de mamiferos, tais como actina, miosina, etc.

Podem tambem ser seleccionados sinais de iniciação da transcrição, que permitam a repressão ou activação de modo a poder-se modular a expressão dos genes. São de interesse sinais reguladores que sejam sensiveis à temperatura de tal forma que variando a temperatura, a expressão pode ser reprimida ou iniciada ou que estejam sujeitos a regulação quimica. e.g. metabolito.

Uma vez preparado o vector ou sequencia de DNA contendo as construções para expressão as construções de DNA podem ser introduzidas num hospedeiro

adequado. Podem ser empregues varias tecnicas tais como fusão de protoplastos, precipitação com fosfato de calcio, electroporação ou outras tecnicas convencionais. Após a fusão, as celulas foram cultivadas em meio e despistadas quanto às actividades adequadas. A expressão do(s) gene(s) resulta na produção da APH.

As celulas hospedeiras para a produção de APH podem ser celulasimortalizadas, principalmente celulas de mieloma ou de linfoma. Estas celulas podem ser cultivadas num meio nutritivo adequado em frascos de cultura ou injectadas num hospedeiro sinergístico, e.g. ratinho ou rato, ou hospedeiro imunodeficiente ou sitio imunodeficiente de um hospedeiro, e.g., ratinho labro ou bolsa de hamster.

A APH do invento pode ser isolada e purificada de acordo com as condições convencionais como seja extracção, precipitação, cromatografia, cromatografia de afinidade, electroforese ou similares.

Utilizações

APH, uma vez produzida e purificada, pode ser usada, por exemplo no contexto de produção farmacêutica para hidrolisar um peptídeo N^α-acilado ou para amino-acilar o extremo N^α-de um peptídeo.

Este último é efectuado numa solução aquosa e torna resistentes proteínas susceptíveis à degradação de Edman. Aquela pode ser efectuada num ambiente quase anidro e é útil na redução da degradação de proteínas e a serem usadas terapêuticamente. Ver a discussão a seguir a A. Klibrinov, "Unconventional Catalytic Properties of Conventional Enzymes" em Basic Biology of New Developments in Biotechnology, pag. 497-518. (A Hollander ed, 1973), sobre a utilização de enzimas em sistemas bifásicos para síntese orgânica.

O ambiente quase anidro alterará a especificidade de substrato de APH, de tal modo que ocorra a amino-acilação dos peptídeos. A especificidade do substrato de uma enzima em solventes orgânicos pode ser radicalmente diferentes, e muitas vezes oposta, à na água (ver Zaks et al, J. Am. Chem. Sci. 108:2767-2768 (1984)).

Demonstrou-se que os peptídeos podem ser sintetizados por tripsina e alfa-quimotripsina em solventes miscíveis em/ou imiscíveis com água (ver Puhniere et al, Proteins, Structure, Function and Genetics) 1:134-138 (1986)). As klibases pancreáticas porcina, de

levedura e de fungos actuam fortemente como catalisadores numa serie de solventes quase anidros.

A actividade das lipases nos meios organicos depende do pH da soluçãõ aquosa a partir da qual a lipase é recuperada. A actividade maxima de lipase bo solventes organico coincide com o pH optimo da actividade enzimatica em água (ver Zaks et al, Proc.Natl.Acad.Sci. USA 82:3192-3196(1985)). Foi tambem demonstrado que uma carboxipeptidase serina, como seja, a carboxipeptidase Y derivada de levedura pode sintetizar um peptideo a partir da reacção de um éster ou amida de aminoácido ou outro substracto com um aminoácido ou outro componentes amina (Patente U.S. No.4339 534).

Enzimas tais como APH podem funcionar bem como catalisadores em solventes organicos incluem (1) uma escolha correcta do solventes (sendo preferidos os hidrofóbicos se não removerem a camada essencial de agua da molecula de enzima); (2) a utilização de uma enzima recuperada a partir de uma soluçãõ aquosa do pH optimo para a actividade enzimatica; e (3) eliminacão das limitacões de difusão por agitacão vigorosa e dispersão fina do pó de enzima no solvente organico (ver Zaks et al, 1986).

Os reagentes na reacção de condensacão catalisado por APH sao polipeptideos aceitadores e.g. proteínas com um grupo terminal N^o livre em um substracto tal como um derivado de alcool benzilico de um aminoácido acilado. A concentraçãõ de substracto necessita de ser suficiente para que se dê a reacção de amino-acilacão. O sol-

mente escolhido é um que seja hidrofóbico e que não remova a camada essencial de molécula de água que rodeiam o enzima. A APH, antes de ser colocada no solvente, é recuperada da solução aquosa do pH ótimo para a actividade enzimática. A dispersão do pó fino de APH no solvente e agitação forte são usados para ultrapassar limitações de difusão (Zaks et al, J.A..Chem. Soc 108:2767-2768 (1986)). Em adição o ambiente organico facilitará a extracção da APH devido à insolubilidade da enzima em meios organicos (Zaks et al, Proc. Natl Acad. Sci. USA 82:3192-3196(1985)).

APH pode ser suspensa no seu pó fino hidratado ou pode ser imobilizado num veciculo. A estabilidade das enzimas em relação a agentes inactivadores, tais como alcoois mono-hidricos é muitas vezes aumentada por imobilização. Foi demonstrado que a tripsina e alfa-quimiotripsina, quando imobilizadas num complexo insolúvel de alumina-fosfocolamina, apresenta notavel resistencia a solventes organicos, incluindo alcoois mono-hidricos misciveis em água (Pugniere et al, 1986).

APH podem ser imobilizada por metodos conhecidos em esferas ou outros veiculos, os quais podem então ser usados em bloco ou em colunas.

Tendo descrito de um modo geral este invento o mesmo será melhor compreendido com referencia a exemplos especificos, os quais aqui são incluidos apenas com fins ilustrativos e não pretendem ser limitantes a menos que seja especificado.

EXEMPLOS

Exemplo 1

Extracção e purificação da Hidrolise de Acil-peptídeos (APH)

Materiais

DEAE.Sepharose CL-6B, colunas de EPLC (Mono-Q HR5/5 e MONO S HR5/5). Sephacryl S-300 super fina, Octil-Sepharose e Polybuffer 74 foram da Pharmacia. A coluna de Sphergel CAA-HIC (0,46 × 10 cm) foi da Beckman. A hidroxilapatite (BioGel HT) foi da Bio-Rad. O glicerol foi da BRL. Reactigel 6× foi da Pierce. Os aminoácidos (Ac-L-Ala) foram da Sigma. Todos os outros produtos químicos foram grau de pureza reagente ou melhor.

Purificação da enzima

APH foi purificada a partir de 300 g de fígado de rato (macho, estirpe CD) como descrito por Tsunasawa et al, J.Biochem. (Tokyo) 77:89-102(1975), excepto no que respeita à substituição DEAE-celulose e Sepharose 6B por DEAE-Sepharose CL-68 e Sephacryl S-300, respectivamente. Os tamanhos das colunas e os gradientes foram também alterados. Para a cromatografia em hidroxilapatite o gradiente de partida foi de tampão fosfato 5 mM em vez de fosfato 20 mM e usou-se 10% de glicerol no gradiente.

Durante a cromatografia em DEAE-Sepharose CL-68 observou-se um aumento na actividade total. Para confirmar a homogeneidade da proteína da coluna de Sephacryl S-300, efectou-se uma cromatografia adicional: (i) cromatografia de permuta iónica com o

sistema de FPLC Pharmacia em Mono Q e Mono S com varios tam-
pões a pH entre 5 e 8; (ii) cromatografia de interacção
hidrofóbica em Octil-Sepharose e Spherosgel CAA-HIC ;
(iii) cromatofocagem em Mono P com Polybuffer 74; e
(iv) cromatografia de afinidade usando Ac-L-Ala-Sepharose,
preparada a partir de Reacti-Gel 6x (Pierce) e acetil-
L-alanina. Em nenhum dos casos foi observado outra separa-
ção ou aumento de actividade. A purificação está resumida
na Tabela 1.

TABELA 1

Purificação da Hidrolase de Acil-peptídeos de
fígado de rato

PASSO		Actividade (unidades)	Proteína (mg)	Actividade específica (unit/mg)	rendi-mento (%)	Purifi- cação
1	Homogenato	194	44200	0,00439	100	1,00
2	12000 × g	194	39400	0,00492	100	1,12
3	Sobrenadante Sulfato de amó- nio (20-50%)	150	25400	0,00591	77	1,35
4	Tratamento α _p /calor	139	2520	0,0552	65	11,5
5	DEAE-Sepharose	208	29,3	7,10	108	1630
6	Hidroxilapatite	148	5,90	25,1	76	5780
7	Sephaeryl S-300	118	4,04	29,2	61	6090

Exemplo 2

Sequenciação de aminoácidos de fragmentos de APH tripticos e de brometo de cianogênio.

Materials

APH foi purificada como no Exemplo 1. A pureza foi confirmada por electroforese em gel de SDS-poliacrilamida pelo metodo de Laemmli; Nature 227:680-685 (1970).

- As medições de UV foram obtidas usando um espectrofotômetro de UV Hewlett-Packard 8450A. A quantidade de proteina foi determinada pelo método de Bradford, M.M. (Anal. Biochem. 72:248-254(1976)) usando albumina de soro bovino como padrão e expressa em nmol de subunidade de hidrolase de acil-peptideos de figado de rato, assumindo que 1 nmol de enzima se refere a 1 nmol de subunidade de $M_r=80000$ da enzima (Kobayashi, K.e. Smith, J.A. J. Biol. Chem. 262:11435-11445 (1987)).

As amostras radiocativas foram contadas num contador de cintilação Beckmann LS3801.

Brometo de cianogenio, guanidina-HCl, 2-mercaptoetanol, acido trifluoroacético (TFA) foram adquiridos à Pierce. O acetonitrilo (grau de pureza para HPLC com absorção de UV a 188 nm) foi de J.T.Baker.

A tripsina tratada com N-tosil-Fen CH_2Cl foi adquirida a Worthington. O reagente Bradford para o ensaio de proteínas e os reagentes para electroforese foram adquiridos a Bio-Rad, excepto os marcadores de peso molecular e Tris, os quais foram adquiridos a Sigma.

Zwittergent 3-14 foi da Calbiochem e o ácido $[\text{ }^{-14}\text{-C}7]$ iodoacético (9,8 mCi/mmol) foi da New England Nuclear. Todos os outros reagentes foram do maior grau de pureza disponível comercialmente.

Análise dos aminoácidos

A hidrolase de acil-peptídeos foi dialisada extensivamente contra ácido acético 0,1M, liofilizada e hidrolisada a 110°C durante 24 h e 48h em 6M HCl contendo 0,1% de fenol. A 110°C da composição em aminoácidos foi determinada usando um Analisador de Aminoácidos Beckman (ver Moore, S, Em. Chemistry and Biology of Peptides, (Meinhofer, J.Ed), pp.629-652, Ann Arbor Science Ann Arbor MI(1972) (Tabela 2).

TABELA 2Composição em aminoácidos da Hidrolase de Acil-Peptid_i
os de fígado de rato

A composição teórica foi determinada a partir de sequen
cia primária deduzida da sequencia nucleotídica na Fig.3
A composição observada foi calculada por análise de aminoá
cidos de Hidrolase de Acil-Peptideos de fígado de rato
purificada (N=3). A composição observada foi calculada
assumindo uma subunidade $M_r = 80,000$.

<u>Aminoácido</u>	<u>Teórico</u>	<u>Observado</u>
Asx	57	55
Thr	29	29
Ser	67	64
Glx	80	84
Gly	54	52
Ala	45	45
Val	61	61
Met	19	15
Ile	24	23
Leu	75	77
Tyr	24	26
Phe	29	36
His	19	19
Lys	30	31
Arg	34	34
Pro	50	65
Cys	19	ND ^a
Trp	16	ND
TOTAL	732	

^a Abreviaturas: Asx = Asn + Asp; Glx = Gln + Glu;
ND, não determinado.

Redução de pontes dissulfureto e alquilação com ácido iodó acético

APH de rato purificada (3 mmoles) foi dissolvido em Tris-HCl (pH 8,5) contendo 7 M guanidina HCl/2mM EDTA e reduzida com 8-10mM 2-mercaptoetanol em atmosfera de argon à temperatura ambiente durante 12 h ou a 37°C durante 3 horas. A mistura foi adicionado (0,19 ml) de ácido ^{-14}C iodoacético (2,6 umoles) em 30 μl de 0,5M Tris-HCl (pH 8,5) /7M guanidina HCl/2mM EDTA) e a reacção decorreu durante 1h a 37°C no escuro.

Adicionou-se então 2-mercaptoetanol para uma concentração final de 0,2M. Removeu-se o sal à proteína por precipitação com quatro volumes de acetona/metanol (3:1 v/v) ou por diálise contra 0,1M ácido acético e liofilizou-se numa concentração/evaporador Savant.

Clivagem com brometo de cianogénio

A proteína carboximetilada foi dissolvida em 70% de ácido fórmico (0,1-0,2 mol) ao qual se adicionou 10-15 μl de solução de CNBr (100 mg/ml) em ácido fórmico a 70%). A mistura foi incubada à temperatura ambiente durante 24 h e seca sob vácuo após diluição com água. Os fragmentos peptídicos obtidos por clivagem com CNBr foram precipitados por HPLC de fase reversa (RPLC) ou por liofilização num concentrador/evaporador Savant ou ainda fragmentos por digestão triptica.

Digestão com Tripsina

A mistura bruta dos peptídeos CNBR (3mmoles) foi dissolvida em 0,2 ml de bicarbonato de amônio 0,2M contendo 0,2% de Zwittergent 3-14 e digerida com Tripsina (50 pmoles) tratada com N-tosil-Fen CH₂Cl durante 20 h a 37^oC. O produto da digestão foi seco sob vacuo e dissolvido em 6M guanidina HCl em 0,1% TFA para purificação por RPLC.

Purificação dos fragmentos peptídicos por HPLC de fase reversa

Os peptídeos foram purificados por RPLC num sistema de HPLC beckman 344, usando uma coluna C₄ (Vydac, 0,46 × 25 cm, particulas de 10 microns com 300Å de poro) para fragmentos CNBr ou uma coluna Phenyl (Vydac, 0,46 × 25 cm, particulas de 5 micron com 300Å de poro) para os fragmentos tripticos. A mistura de peptídeos bruta foi aplicada numa coluna equilibrada com 0,1% de TFA e eluida com um gradiente linear de 0-80% de acetonitrilo em 0,1% TFA (para peptídeos CB/R) em 160 min a um fluxo de 1 ml/min.) Uma mistura de peptídeos tripticos derivada de uma mistura bruta de peptídeos CNBr foi aplicada como descrito atrás e eluida com um gradiente linear de 2-60% de acetonitrilo em 0,1% TFA em 180 minutos a um fluxo de 1ml/min.

As eluições foram escolhidas e controladas por absorvância a 214 nm e a 280 nm. Cada pico foi colhido manualmente e, se necessario, ainda purificado por RPLC isocrática usando a mesma HCl-0,1% TFA.

A concentração optima de acetonitrilo para a separação dos peptideos de cada fracção foi calculada a partir do padrão de eluição da primeira HPLC (ver equação de Wong et al, Proc. Natl. Acad. Sci. USA 82:7711-7715 (1985)).

Sequenciação dos peptideos

As analises de sequencias peptidicas foram efectuadas usando um Sequenciador de Proteinas Applied Biosystems 470A e um Analisador Pth Applied Biosystems 120A (ver Hewick et al, J. Biol. Chem. 256: 7900-7997 (Tabela 3)).

TABELA 3

ANALISE DE SEQUENCIAS PROTEICAS

Os peptídeos CNBR e tripticos foram derivados e sequenciados como descrito atrás. Os peptídeos são identificados de acordo com a clivagem e purificação: CB um fragmento de clivagem com CNBR purificado por RPLC; CB-R, um peptídeo triptico purificado por RPLC derivado de um peptídeo CNBR purificado por RPLC; e CB/R, um peptídeo triptico purificado por RPLC derivado de uma mistura bruta de peptídeos CNBR. A recuperação de Pth-aminoácidos detectados em cada degradativo está indicada em pmmols da qual se calcula o rendimento repetitivo. Os números correspondem a varias fracções colhidas derivadas do gradiente de RPLC e as letras minúsculas correspondem às várias fracções colhidas manualmente derivadas de RPLC isocrática, necessarias para posterior purificação.

Peptido	Pth-Aminoácido/ciclo degradativo (pmmol)																								Ry ^a	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
<u>CB17-R13</u>	Glu	Arg	106	258	299	285	254	139	234	139	119	90	133	140	133	87	82	15								92%
<u>CB16-R18, 19-C</u>	Gly	Leu	21	304	78	152	119	43	93	163	80	76	107	63	107	51	40	41	54	52	22	19	25	24	20	89%
<u>CB17-R3</u>	Thr	Val	185	117	62	179	37	89	39	8																89%
<u>CB/R97</u>	Gln	Tyr	108	220	133	121	84	57	76	116	186	67	90	63	56	60	46	51	53	20	81	28	11	25	93%	
<u>CB/R96</u>	Lys	Ala	98	90	126	92	108	108	78	42	52	46	27	25	35	25	13									86%
<u>CB17-R33</u>	Ser	Phe	143	9	85	76	54	45	37	35																

-41-

TABELA 3 (CONT)

Uma vez que se encontrou um resíduo Lis na sequência proteica deduzida a partir da sequência de cDNA, Trp provavelmente representa uma forma modificada de Lis, ou no entanto a sua estrutura química não foi ainda determinada.

Exemplo 3

Preparação de sondas

Construção de sondas

Sintetizaram-se duas sondas de oligonucleotídeos degeneradas e sobreponíveis, YS20.1 e YS17.2 (Figura 2A), derivados da sequência de aminoácidos do peptídeo CB18-R11-13c (Tabela 3) e usaram-se para o despiste de uma biblioteca de cDNA de fígado de rato em λ gt11. As sondas oligonucleotídicas e os iniciadores foram sintetizados com um sintetizador de DNA Applied Biosystem 380A usando fosforamídicos de B-cianoetil (Sinha, N.D. et al, Nucl. Acid. Res. 12:4539-4557(1984) e purificados por eletroforese em gel de poli-acrilamida de acordo com o Manual da Applied Biosystems por precipitação com etanol a partir de uma solução de oligonucleotídeo contendo 10mM $MgCl_2$.

YS17.2 e YS20.1, respectivamente conjuntos representam oligonucleotídeos 128 vezes degenerados. As sondas YS17.2 e YS20.1, tinham 17 e 20 nucleotídeos de comprimento, respectivamente. As duas sondas sobrepoem-se 12 nucleotídeos, de tal forma que a utilização sequenciada das sondas para fazer o despiste da biblioteca de DNA faz um despiste eficaz de um pedaço de 25 nucleotídeos de DNA codificador de APH.

Preparação de RNA

Fígado de rato da estirpe CD foi rapidamente congelado em azoto liquido e descongelado em isotiocianato de guanidina, e o RNA foi purificado por centrifugação em CsCl (Chirgwin et al., Biochem. 18 :5294-5299 (1979). O rendimento foi de 60 μ g de RNA. Selecionou-se RNA Poli(A) + por passagem de RNA total através de uma coluna de oligo(dT) celulose (Aviv, H. et al., Proc. Natl. Acad. Sci (USA) 69:14 8-1412(1972). Obteve-se quarenta e cinco μ g de RNA poli(A) + que se verificou não estar degradado por análise de transferências de RNA de 1 μ g de RNA e hibridando com uma sonda de cDNA da actina (Spiegelman, B.M. et al., J. Biol. Chem. 258:10083-10089 (1983).

Preparação da biblioteca de cDNA

Sintetizou-se DNA complementar a partir de 10 μ g de RNA (Poli(A) + pelo método de Gubler, U. et al (Gene, 25:263-269(1983) e clonou-se em λ gt11 (Young, RA et al., Proc. Natl. Acad. Sci.(USA) 80:1194-1198 (1983)), como foi descrito por Klickstein, L.B. (Em: Current Protocols in Molecular Biology (Ausubel, F.M. et al., Eds), pags 5,8,4, Wiley-Interscience and Greene Publishing Associates, New York, NY (1987)

A produção de recombinante foi de 4 milhões a partir de 100 ng de cDNA e 10 μ g de DNA e 10 μ g de DNA de λ gt11. A biblioteca foi amplificada em E.coli estirpe Y10088 (lacU169, supE, supF, HsdR-, HsdM⁺, metB, trpR, tonA21, proC:Tn5 (pMC9) e guardada a 4⁰C.

Isolamento de clones de cDNA

A biblioteca foi semeada a 25 000 placas de placa de 150 mm (para o despiste de 10^6 ou menos placas) ou a 10^6 placas por placa de 225 mm x 225 mm (para o despiste de mais de 10^6 placas) e obtiveram-se filtros de transferências em duplicado para cada uma (Maniatis, T et al, Molecular Cloning: A Laboratory Manual Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982)).

Para o despiste com oligonucleotídeos, os oligonucleotídeos foram marcados no extremo 5' para uma actividade específica de $2-8 \times 10^8$ cpm/ μ g com $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ e polinucleotídeo-cinase de T4 (Zoller, M. et al, DNA 3:479-488 (1985)). Os filtros foram hibridados com oligonucleotídeo com 6 x SSC, 0,1% SDS, 0,1% SDS, 0,05% pirofosfato de sodio, solução do Denhardt 1 x e 100 μ g/ml de DNA de esperma de salmão a 65°C durante a noite.

As T_d max e T_d min foram calculadas para cada mistura de oligonucleotídeos com a formula : $T_d = 4(G+C)+2(A+T)$, como préviamente descrito para sequencias curtas (Suggs, S.V. et al, Em: Developmental Biology Using Purified Genes (Brown, D. Ed). pag.683-693 Academic Press, New York, NY (1981)). As sequencias foram lavadas a temperaturas progressivamente mais altas em 6 x SSC 0,05% de pirofosfato de sodio, e 0,1% de SDS até estar reduzida a ligação inespecífica.

Para o despiste com sondas de cDNA (fragmento XmnI-KpnI de APH5.2 ou fragmento BamII-PstI de APH36-1; Figura 2B) os filtros foram hibridados durante a noite com sondas sujeitas a "nick-translation" em 50% de formamida, 5XSSC, solução de/ou Denhardt 's fosfato do sodio 10mM, 0,1% de SDS, 1mM EDTA, e 50 ug/ml de DNA de esperma de salmão desnaturado e sonificado a 42^oC. Os filtros foram lavados em 0,2 x SSC, 0,1% de SDS, fosfato de sodio 1mM, pH 7,0 e 1 mM EDTA a 55°C. Os filtros lavados foram expostos a filme Kodak XAR com um écran intensificado a -70^oC. Os fogos que deram sinais em duplicado foram purificados por placas em ciclos repetidos de despiste.

Análise de sequencia de DNA

Fragmentos de restrição de APH5.2 e APH 36.1 foram subclonados em M13mp18 ou M13mp19 e sequenciados pelo método de terminação de cadeias didesoxi-nucleotídicas (Sanger F. et al., J. Molec. Biol. 94 :441-448(1975). A sequencia de alguns clones foi obtida construido primeiro mutantes de deleção usando exonuclease III(Henikoff, S. Gene 28: 351-356 (1984). A inserção de cDNA do APH2.7 foi subclonado no plasmideo Bluescript (Stratagene) e sequenciada pelo método de terminação de cadeias didesoxi, modificado para sequenciação de cadeia dupla por Guo et al., (Nucleic Acid. Res. 12 :387-394 (1983) Os resultados de sequenciação de DNA foram analisados com os dados do University of Wisconsin Genetics Computer User Group (Devereaux, J. et al, Nucle. Acid. Res. 12: 387-395 (1984).

Clonagem e Sequenciação de cDNA codificador da Hidrolase de Acil-Peptideos de Fígado de Rato

Encontrou-se vinte e sete de 45 000 clones recombinantes que hibridaram com a sonda YS20.1. Doze destes clones foram novamente despintados com a sonda, YS17.2 para dar um único clone, APH 5.2, contendo uma inserção de 1,3Kb (Figura 2B). A sequência de DNA de APH 5.2 codificava toda a sequência peptídica do peptídeo triptico CB18-R11-13C confirmando que APH5.2 era um clone autêntico. Uma vez que APH5.2 continha uma sequência poli(A)+ no seu extremo 5' (Figura 2B, rectângulo sombreada com linhas cruzadas) provavelmente um artefacto criado durante a construção da biblioteca de cDNA, usou-se um fragmento XmnI-KpnI de APH5.2 para fazer novo despiste de um milhão de clones da mesma biblioteca de cDNA e obteve-se o clone APH36.1, contendo uma inserção de 2,5kb (Figura 2B).

A sequência proteica deduzida a partir de sequências de DNA de APH36.1 continha todas as sequências proteicas na Tabela 3, excepto os três resíduos de extremo amina do peptídeo CB17-R13. No entanto, o seu extremo 5' continha um fragmento de 120pb codificador de albumina sérica de rato (Figura 2B, rectângulo com diagonais).

Para se obter os resultados de sequenciação 5' que faltam, usou-se um fragmento BamII-PstI de 220pb (Figura 2B) para fazer novo despiste da mesma biblioteca de cDNA. Obtiveram-se cinco clones positivos com diferentes tamanhos de caudas poli(A) a partir dos 5 milhões de recombinantes. Quatro clones de cDNA,

incluindo APH 2.7, começavam com a mesma sequência de nucleotídeos e continham um codão ATG na mesma grelha de leitura, nos nucleotídeos 6-8, enquanto que o extremo 5' do quinto clone de cDNA não possuía 18 pares de bases. Nos nucleotídeos 2344-2349 encontrou-se um sinal de poliadenilação, AATAAA. A Figura 3 ilustra a sequência completa de cDNA de APH, conforme derivado de APH5.2, APH36.1 e APH 2.7.

Estrutura primária de hidrolase de acil-peptídeos deduzida a partir de cDNA

A sequência de DNA completa foi determinada combinando as sequências de APH36.1 e APH2.7 (Fig.3). A sequência de DNA codifica uma proteína contendo 732 resíduos de aminoácidos assumindo que o ATG nos nucleotídeos 6-8 é o codão de iniciação de tradução. A sequência proteica deduzida contém todas as sequências peptídicas na Tabela 3 (Fig.1).

A proteína tem um peso molecular calculado de 81 347 e a composição em aminoácidos baseada na sequência proteica deduzida está de acordo com a composição observada (Tabela 2). Conforme deduzido a partir de sequência de DNA, foram identificados três resíduos de aminoácidos 118, 291 e 443 os quais correspondem as posições onde Pth-Trp juntamente com um derivado de Pth que elui na fase tardia (Tabela 3 e Figa.1).

Nos resíduos 134-136, 233-235 e 243-245 foram identificadas tres sequencias consensus de N-glicosilação (i.e., Asn-Xxx-Tre/Ser (Parodi, A.J. et al, Biochim. Biophys. Acta 559: 1-37 (1979)).

Analise da hidrofobicidade

O perfil de hidrofobicidade foi determinado usando o algoritmo de Kyte, M. et al (J. Molec. Biol. 157:105-132 (1982) com uma janela de 8.

A sequencia de proteica da hidrolase de acil-peptideos de rato foi recuperada com as bases de dados de proteinas National Biomedical Research Foundation e Swiss usando os programas Wordsearch e Bestfit da Universidade de Wisconsin Genetics Computer User Group (Deveraux J. et al, Nucl. Acid. Res. 12: 387-395 (1984) e o programa FASTP baseado no algoritmo de Ripman, D.J. et al., (Science 227 : 1435-1441 (1985)). Para identificar possiveis regiões de sitios activos na hidrolase de acil-peptideos de rato, a suasequencia foi comparada com as sequencias peptidicas, contendo os residuos serilo, histidilo ou aspartilo de sitios activos, derivados de proteases serina conhecidos.

A representação grafica de hidrofobicidade revela que a proteina contem uma região hidrofílica localizada entre os residuos 80 e 220, maq, não é claro que se esta região tem um papel especifico nas interações com outras proteinas ou na catalise. A pesquisa baseada em computador de base de dados do National Biomedical Research Foundation e da Swiss Protein não revelou

proteínas fortemente homologas. Em adição a comparação entre hidrolase de acil-peptídeos de rato e peptídeos curtos do centro-activo, contendo serina, histidina e ácido aspártico, derivados de proteases serina conhecidas não revelou semelhantes significativas.

Se bem que a hidrolase de acil-peptídeos se tenha demonstrado ser uma protease-serina por experiencias de inibição, usando fluorofosfato de diisopropilo, acetilalanina, clorometilcetona e outros inibidores de enzimas (Kobayashi, K. Smith, J.A. J.Biol Chem. 262:11435-11445(1987); Tsunasawa et al J.Biochem. (Tokyo) 77:89-102(1975) não se encontrou semelhança forte entre a hidrolase de acil-peptídeos de rato e peptídeos do sitio activo de outras proteases serina, sugerindo que esta enzima pode ser uma protease serina única.

Exemplo 4

Clonagem de todo o gene da hidrolase de acil-peptídeos de rato.

Materiais

Enzimas de restrição, ligase de T4, polinucleotídeo-cinase de T4, DNA-Polimerase I de E. coli e o seu fragmento Klenow, transcriptase reversa de AMV, exonuclease III, DNase I, RNase H, DNA-polimerase de T4, EcoRI-metilase; fosfatase alcalina de intestino de vitela e nuclease S1 foram de Boehringer Mannheim e New England Biolabs.

A RNase A foi da Sigma. O plasmídeo Bluescript, braços λ gt10 e extracto de empacotamento foram de Stratagene γ - 32 P ATP, α - 32 P dCTP e membrana GeneScreen Plus foram adquiridos à New England Nuclear α - 35 S γ -dATP α S foi da Amersham Corp. Oligo(dT)-celulose foi da Kollaboratives Research.

Os oligonucleotídeos sintéticos foram sintetizados com um sintetizador de DNA Applied Biosystems 380A usando a fase sólida baseada em sílica. (Matteucci, M.D. et al., J. Amc. Chem. Soc. 103:3185-3191 (1981) e o método com fosforamido de B-cianoetilo (Sinha, N.D. et al." Nucleic Acids Res. 12:4539-4544 (1984).

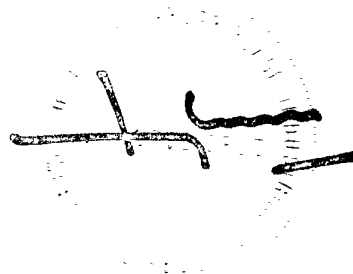
Preparação de DNA e RNA de fígado de rato

A fonte de DNA genômico de rato e de RNA citoplasmático de fígado é fígado de rato adulto Sprague-Dawley. O DNA de fígado foi preparado como descrito por Blin e Stafford (Blin, N, et al, Nucleic Acid Res. 3:2303-2308 (1976).

O RNA total de fígado de rato foi isolado pelo método de tiocianato de guanidina, como descrito por Chrigiwin et al (Chwigwin, J.M. et al, Biochemistry 18: 5294-5299 (1979). O RNA poliadenilado foi purificado por cromatografia em oligo(dT)-celulose (Avir, H. et al., Proc. Natl. Acad. Sci. USA 69:1408-1412(1972)

Análise de transferência de RNA

Purificou-se RNA como descrito atrás desnaturou-se a 65^oC e transferiu-se para membrana Zetabind (Thomas, P.S., Proc. Natl. Acad. Sci. (USA) 77:6201-5202 (1980). As transferências foram hibridadas com o fragmento de cDNA XmnI-KpnI de APH5.2 sujeito a "nick-translation" em 50% de formamida, 5XSSC, solução de Denhardt 5X, fosfato de sódio 10mM, 0,1% de SDS, 1mM EDTA e DNA de esperma de salmão desnaturado e sonificado (50 µg/ml). Os filtros foram lavados em 0,2XSSC, 0,1% de SDS, fosfato de sódio 10mM pH 7,0 e 1 mM EDTA a 55^oC.



A análise da transferência de RNA de RNA total, usando o fragmento de cDNA XmnI-KpnI derivado de APH5.2 como sonda (Fig.2) revelou que apenas um mRNA de 2,7 kb em quantidades sensivelmente equivalentes codifica a hidrolase de acil-peptídeos em vários tecidos de rato (i.e. baço, músculo, pulmão, fígado, rim e cérebro.).

Isolamento de clones genômicos

Usaram-se duas bibliotecas genômicas de rato para o despiste de gene APH. Uma biblioteca foi da Clonotech, a qual foi construída por digestão parcial com EcoRI de DNA de fígado de Sprague-Dawley e clonagem em Charon 4A do fago λ . Um fragmento de restrição EcoRI de 2,4 kb codificador de APH de rato derivado de APH 36.1 foi marcado por nick-translation (Sargent, T.D. et al, Proc. Natl. Acad. Sci. USA 76: 3256-3260 (1979) com λ -³²P7 dCTP para uma actividade específica de 10^8 cpm/ μ g e usando sonda para o despiste desta biblioteca genômica.

A outra biblioteca genômica que foi construída por digestão parcial com HaeIII e clonagem em Charon 4A do fago λ foi uma oferta generosa do Professor James Bonner de California Institute of Technology (Church, G.M. et al, Proc. Natl. Acad. Sci. USA 81:1991-1995 (1984). Um fragmento BamII-PstI de 200 pb de APH36.1 marcado por iniciação ao acesso (Feinberg, A. et al, Anal. Biochem. 132:6-13(1983) com λ -³²P7 dCTP para uma actividade específica de 10^9 cpm/ μ g, foi usado no despiste desta biblioteca, Aproximadamente 1×10^6 fagos de cada uma das biblio-

tecas foram despistadas por hibridação de placas (Chrurch, G.M. et al Proc. Natl. Acad. Sci. USA 81:1991-1995 (1984))

As placas positivas foram purificadas e isolado o DNA fagico (Maniatis, T. et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Cold Spring Harbor, New York (1982).

Analise de DNA por Mapeamento de restrição e hibridação de DNA

O mapa de restrição do gene clonado foi construido por digestão de DNA fagico com varias endonucleases de restrição (Maniatis, T. et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Cold Spring Harbor, New York (1982). Para a análise de transferencias de DNA, os fragmentos de restrição do DNA separados num gel de agarose, transferidos para membrana Gene Screen Plus e hibridados com a sonda marcada com ^{32}P de acordo com as recomendações do fabricante. As sondas são tres fragmentos de cDNA de APH de rato de APH36.1 marcados com ^{32}P : um fragmento 5' BanII-PstI de 200 pb, um fragmento kpn-EcoRI de 420 pb e um fragmento EcoRI de 2,4kb. Para a hibridação de transferencias de DNA genómico o fragmento de 2,4 kb do APH2.7 foi usado como sonda.

Sequenciação de DNA

Fragmentos de restrição dos clones genômicos de rato foram subclonados no plasmídeo Bluescript. Ambas as orientações das sequencias completa, assim como regiões a montante e a jusante, do gene APH de rato foram determinadas pelo metodo de terminação de cadeias didespxi de Sanger (Sanger, F. et al. Proc. Natl. Acad. Sci. USA 74: 5463-5467 (1977), modificado para sequenciação de ca² deia dupla (Guo, L. H et al, Nucleic, Acids. Res. 11:5521-5539, empregando estratégias de sequenciação do metodo de delecção com DNase I (Lin, H.-C. et al Anal. Biochem. 147:114-119(1985), o método de delecção com exonuclease III (Henikoff, S. Gene 28:351-359 (1984) e iniciadores oligonucleotidicos sinteticos. Os resultados de sequencia de nucleotídeos foram compilados e analisados pelo Genetics Computer Group Sequence Analysis Software Package, versão 5.0 (Devereux, J. et al, Nucleic Acids. Res. 12:387-395 (1984).

Preparação de uma biblioteca de cDNA contendo a região 5' não traduzida do mRNA de APH

Um oligonucleotideo sintetico de 17pb, 5' GTGACCTCCGGACCCAG 3', complememntar dos nucleotídeos 95-112 do APH2.7., foi usado como iniciador especifico para construir uma biblioteca de cDNA em λ gt10. O oligonucleotideo sintetico (1 μ g) foi emparelhado com 10 μ g de RNA poli(A)+ e a sintese da primeira e segunda cadeia de cDNA foi efectuada pelo método de Gubler e Hoffman (Gubler, U et al, Gene 25:263-269 (1983). Os extremos de cDNAs

foram tornados cerses com DNA-polimerase de T4 e os sitios EcoRI internos foram metilados. Os cDNAs de extremos cerses foram ligados a adaptadores EcoRI e após digestão com EcoRI o cDNAs foram fraccionados por tamanho numa coluna CL-4B (Wong, W.W. et al, Proc. Natl. Acad.Sci. USA 82: 7711-7715 (1985). Em seguida os cDNAs foram ligados a braços λ gt10 e o DNA dos fagos recombinantes foi empacotado de acordo com as recomendações do fabricante (Stratagene) Os fagos recombinantes foram despistados com o oligonucleotídeos sintéticos, 5'AAGTCCCGGAAGTGAGG 3' e 5' CTGACGCTCCA TAGTCG 3' cujas sequências foram derivadas da sequência (nucleotídeos 586-592, Fig.4) e de cDNA (nucleotídeos 1-17 de APH 2.7) respectivamente,.

O DNA fágico com a inserção maior foi purificado (Maniatis T, et al, Molecular Cloning; A Laboratory Manual, Cold Spring Harbor Laboratory Cold Spring Harbor New York, (1982) e a inserção foi subclonada no plasmídeo Bluescript.

Extensão do iniciador

Um oligonucleotídeo

5' TAGGAGTGAGAAAATCA 3' complementar da sequência de nucleotídeos no primeiro exão/nucleotídeos 44-60 Fig.4) foi marcado no extremo 5' com γ ^{32}P -ATP e polinucleotídeo-cinase de T4 e depois hibridado com 10 μg de RNA poli(A)+ de fígado de rato numa solução de 0,1M KCl, 5mM EDTA, tampão fosfato de sódio 5mM, pH 6,8. Para a hibridação, a temperatura da solução foi aumentada para 90°C durante 5 min. e voltou gradualmente para 42°C.

Os híbridos de RNA-DNA foram então sujeitos a reação com transcriptase reversa. Após digestão com RNase A, extração com fenol: Cloroformio e precipitação com etanol, o produto de extensão com iniciador foi então sujeito a eletroforese num gel de sequência de 8% de poliacrilamida.

Isolamento e sequenciação de DNA do gene de hidrolase de acil-peptídeos de rato

Inicialmente, uma biblioteca de genômica de rato, construída com Charon 4A a partir de uma digestão parcial com EcoRI do DNA de fígado de rato Sprague-Dawley, foi despistada usando como sonda, a inserção de cDNA de APH de fígado de rato de APH 36-1 marcada com ^{32}P . Foram despistadas um milhão de placas e vinte e oito hibridaram com a sonda. Todas as vinte e oito placas

foram isoladas e caracterizadas por mapeamento com enzimas de restrição e verificou-se serem idênticas. Transferências de DNA dos produtos de digestão com endonucleases de restrição derivados de cada DNA recombinante foram depositadas com as inserções de cDNA de APH36.1, o fragmento BanII-PstI de 200 pb (extremo 5' de APH 26.1), e o fragmento KpnI-EcoRI de 420 pb (extremo 3' de APH36.1) e cada um continha um fragmento EcoRI de 9,7Kb, correspondendo ao extremo 3' do cDNA. A um destes clones, APHE5 (Fig.5A) foi feito o mapa de restrição, subclonado em Bluescript e sequenciado. Uma segunda biblioteca genômica de rato, construída por digestão parcial com HaeIII do DNA de fígado de rato e clonada em Charon 4A (Gentilmente cedido pelo Professor J. Bonner, California Institute of Technology) foi despidada com o fragmento BanII-PstI de 200 pb de APH36.1. Oito de um milhão de placas hibridaram com a sonda e os seus DNAs fágicos isolados. Após mapeamento com enzimas de restrição e hibridação do DNA, um clone, APH H6 (Fig.5A) sobreponível com o clone APHE5 e prolongando-se mais na direção 5' foi analisado. Na Figura 5B está apresentado o mapa de restrição combinado de APHE5 e APHH6.

A análise de transferências de DNA do DNA genômico de rato revelou duas bandas correspondendo a fragmentos de restrição BamHI de cerca de 4 e 9,4 kb. O tamanho do fragmento maior está de acordo com o tamanho calculado através do mapa de restrição e o fragmento mais pequeno prolonga que para além do extremo 3' do mapa (Figura 5B). Duas bandas foram observadas correspondendo a fragmentos de restrição EcoRI com cerca de 7,5 e 9,7 kb. Os tamanhos de ambos os fragmentos estão de acordo com os tamanhos determinados a partir do mapa de restrição (Fig.5B). Isto indica que o gene APH está presente numa única forma do genoma de rato.

Para a análise das sequencias os fragmentos individuais EcoRI, PstI ou BamHI foram subclonados em Bluescript e os subclones individuais foram sujeitos a digestão unidireccional limitada com exonuclease III seguido de digestão com nuclease S1 ou sujeitos a cortes limitados acaso com DNase I seguido de digestão com enzimas de restrição para dar origem a uma serie de deleções. Preparou-se DNA de plasmídeo de cadeia dupla a partir de cada deleção e sequenciou-se pelo método de terminação de cadeias didexoxi. Para certas regiões, a sequencia foi determinada usando oligonucleotídeos sintéticos específicos como iniciadores de sequenciação.

Estes dois clones genomicos foram sequenciados em ambas as orientações. A sequencia de nucleotídeos completa está apresentada na Fig.4. Por questões de simplicidade na discussão da sequencia genómica, usou-se um sistema de numeração em que a posição +1 significa o sitio de iniciação da transcrição do gene APH. Como se mostra na Fig.5C e Tabela 4, as localizações precisas de cada uma das fronteiras exão-intrão 5' e 3' foram definidas alinhando a sequencia genómica com a sequencia de cDNA.

Análise da região 5' não traduzida do mRNA da hidrolase de acilpeptídeos de rato

Devido à sequencia de cDNA não possuir cerca de 300 pares de bases da região 5' não traduzida, calculado por comparação do tamanho do mRNAAPH com o tamanho do cDNA, construiu-se uma biblioteca de cDNA contendo a região 5' não traduzida do mRNA de APH como descrito atrás.

Esta biblioteca foi despistada com duas sondas oligonucleotídicas marcadas com ^{32}P com sequências complementares das primeiras dezassete bases do clone de cDNA APH2.7 ou com os nucleotídeos 586-602 do DNA genômico. (Fig.4). Uma inserção de cDNA de 466 pb foi isolada, subclonada em Bluescript e sequenciada pelo método de terminação de cadeias didesoxi. Esta sequência de extensão 5' continha a sequência de nucleotídeos correspondendo aos nucleotídeos 37 a 262, 492 a 636 e 711 a 805 do DNA genômico (Fig.4). Portanto, o códon de iniciação da tradução ATG, está situado nos nucleotídeos 625-627 (Fig.4) uma vez que está precedido na mesma grelha de leitura por um códon de terminação TAG, nos nucleotídeos 568-570 e portanto não existe outro códon ATG entre eles.

TABELA 4

JUNÇÕES INTRÃO-EXÃO NO GENE DA HIDROLASE DE
ACIL-PEPTIDEOS

exão no.	Tamanho do Exão	sítio de "splice" 5	sítio de "splice" 3	TAMANHO do Intrão (bp)
1	262	GCTCACAgtcggtc-----	ccccagCTGGTTG	229
2	145	GCGTCAGgtgagg-----	tgcgagGTGCTGC	74
3	133	CACACTGgtgtgta-----	cttcagAGTGGAC	452
4	127	GGGGGAgtaagt-----	ttctcagGCTGCTG	746
5	94	CTTGAGgtgagtc-----	tcctcagGCTGGG	92
6	76	GAGGATGgtgagg-----	catgtagACTGCTT	89
7	164	CATCAAGgtgctg-----	ttctcagGGGACC	141
8	138	TGGTCAGgtcagca-----	tttacagGCTTTT	94
9	92	ATCGCAGgtgagga-----	ttccagATCAGCT	76
10	41	AAGTGTGgtaagt-----	ggcctagAACTACT	91
11	122	CTGCCTGgtgagtt-----	cttcagTACGACT	462
12	61	CTGGGAGgtaagag-----	tttcagAGAGCTT	1003
13	98	TCGGCAGgtaaaag-----	gtttaagGACCTGT	601
14	52	ACAGCTGgtgagca-----	cctctagCGGGGTC	1104
15	89	AAGCCTGgtgagta-----	ttggcagAAAGTTG	358
16	139	CAATATGgtgagct-----	cctgcagCTGACCT	411
17	84	CCCCATGgtaggt-----	tctgcagGGGACC	148
18	81	CTTCTGGgtaatgc-----	ctttcagTGAACTA	454
19	89	TGTCCAGggtgcag-----	actttagTTGCAG	93
20	191	CTGFTGgtgagtg-----	tttatagGTGTATG	84
21	103	CCCTCAGgtactca-----	taccagGTAAAGA	87
22	107	CTGTCCGgtgagtg-----	cacatagGCTCCTG	85
23	262			

Determinação do sitio de iniciação da tradução da hidrolase de acil-peptídeos de rato

O extremo 5' de mRNA de APH foi mapeado por análise de extensão de iniciador. Um oligonucleotídeo marcado com ^{32}P , 5' TAGGAGTGAGAAAATCA 3' complementar dos nucleotídeos 44-60 (Fig.4/), foi hibridado com RNA poli(A)+ de fígado de rato. Este iniciador foi prolongado com transcriptase reversa na presença de desoxinucleotídeos e o comprimento do produto prolongado, determinado como atras foi de 60 nucleotídeos. Ainda, se RNA poli(A)+ de fígado de rato for substituído por RNA poli(A)+ de lebedura não se observa qualquer produto prolongado. Encontrou-se que a transcrição começa com um resíduo T no molde de DNA, o qual corresponde a um resíduo A na posição na Fig.4. Este resíduo A está situado 395 pb 5' do códon de iniciação da tradução ATG.

Resumindo, estes exemplos mostram que o gene APH de rato está presente numa única forma. A sequência completa do gene APH de rato foi determinada, incluindo 3.58 kb de DNA delimitante 5', 2.75 de DNA exônico, 6.94 Kb de DNA intrônico e 1 kb de DNA delimitante 3' (Fig. 4). Os resultados de DNA exônico inequivocamente identificaram o sitio de iniciação da tradução correspondendo ao códon codificador da metionina no resíduo 1 da hidrolase de acil-peptídeos de rato, numa vez que existe nenhum códon ATG colocado entre o códon de paragem na mesma grelha de leitura (Figura 4, nucleotídeos 568-570) e o códon de iniciação da tradução no exão 2 (Figura 4).

Estes resultados também indicam que APH não é sintetizado como uma proteína precursora, uma vez que a sequência de APH proteica a seguir à Met NH₂-terminal pode ser identificada por degradação automática de Edman após clivagem com CNBr.

Como se mostra na Fig.5C, o gene APH de rato, abrangendo 9,69 kb está dividido em 23 exões. Os exões individuais variam de tamanho entre 41 e 262 pb (Tabela 4). O primeiro intrão interrompe a sequência 5' não traduzida; todos os outros intrões estavam dentro da região codificadora da proteína do gene. A Tabela 4 mostra as sequências das junções 5' e 3' de cada intrão e estas sequências são consistentes com as sequências consensus para as junções intrão-exão de outros genes eucarióticos (Sharp, P.A. Cell.23:643-646 (1981); Breathnach R. et al Ann. Rev. Biochem. 50, 349-383 (1981); Mount S.M. Nucleic Acids Res. 10 :459-472(1982). Todos os intrões começam com a sequência GT na fronteira 5' e terminam com a sequência AG para a fronteira 3' e em todos os casos as sequências delimitantes as fronteiras 5' e 3' são ricas em purinas e pirimidinas respectivamente.

A organização exão-intrão do gene APH, presumivelmente codificador de uma proteína/protease com um resíduo serina no sítio activo (Kobayashi, K. Smith J.A., J. Biol. Chem. 262 : 11435-11445 (1987), é muito mais complexo do que a de tripsina ou quimiotripsina, as quais contêm cinco a sete exões, respectivamente (Rogers, J. Nature 315: 458-459 (1985). Os resíduos dos sistemas de substituição de cargas destas enzimas, é conhecido que são codificados por exões separados, mas a distribuição dos resíduos correspondentes em APH espera mais estudos.

Com base numa pesquisa exaustiva da National Biomedical Research Foundation and Swiss Protein Databases, assim como uma comparação da organização exão-intrão de outras proteases serina (e.g. Tripsina quimio tripsina elastase, urocinase, calicrefina, adipsina, a hidrolase de acil-peptídeos não apresenta qualquer semelhança clara com estas proteases serina.

A análise do DNA delimitante 5' do gene APH revelou uma série de sequências conservadas. Uma destas é a sequência 5'TGATAAA 3', que poderá ser uma variante da sequência de uma caixa "TATA", situados nos nucleotídeos -24 a -30. Esta é uma localização típica da caixa TATA, a qual normalmente se encontra 26-34 nucleotídeos a montante de sítio de iniciação da transcrição (Cordon, J. et al., Science 209 :1406-1414 (1980). Uma outra sequência, 5'-TCAAT-3' (nucleotídeos -95 a -99) encontrou-se 95 nucleotídeos a montante do sítio de iniciação da transcrição e é semelhante à sequência da caixa "CCAAT" a qual geralmente se encontra a 70-80 bases do sítio de iniciação da transcrição. Em adição, uma sequência de 6pb, CGCGGG, está repetida três vezes. Uma repetição está situada nas posições -81 a -76 e presume-se estar dentro da região do promotor do gene APH.

As duas outras repetições, presentes na orientação contrária como CCGCCC estão situadas 5 nucleotídeos e 31 nucleotídeos a jusante do sítio de iniciação da transcrição. Esta descrito todas as regiões de ligação Sp1 contem uma ou mais cópias exactas desta sequência GCGCGG, as quais podem estar presentes em qualquer uma das orientações relativamente à transcrição (Dynam, W.S. et al., Nature 316:774-448 (1985).

Uma sequencia de 200 pb aparece em tandem começando a 917pb a montante do sitio de iniciação da transcrição. Comparado com outras sequencias no banco de genes, os dois terços 3' desta sequencia é semelhante à repetição Ala tipo â de ratinho (80% de semelhança) (Kominami; R et al, J. Mol. Biol. 165:209-228 (1983). Uma sequencia semelhante mas apenas como uma copia está presente na junção de SV40 e DNA de rato Fisher (Hasson, J.F. et al, J. Mol. Biol. 177:53-68(1984) e em varios outros genes (Min. H.Y. et al, Nucleic Acid Res. 14 :8879-8892(1986); Osumi, T. et al J. Biol. Chem. 262:8138-8143(1987); Cordon L.J. et al, Proc. Natl. Acad. Sci. USA 82: 7934-7938(1985); Phillips, M et al, J. Biol. Chem. 261 :10821-10827 (1986) Esta repetição poderá ter um efeito regulador.

Este gene poderá ser especificamente regulado por factores reguladores que actuam em cis e/ou em trans. Tal regulação pode estar associada à sintese ou degradação de proteínas.

As hidrolases de acil-peptideos têm sido isoladas a partir de varios tecidos de mamiferos e as suas propriedades moleculares e mecanismo de reacção foram parcialmente caracterizadas. Um aspecto de presente invento diz respeito à estrutura primária da hidrolase de acil-peptideos de figado de rato a qual foi deduzida a partir da sequencia de nucleotideos de dois clones de DNA isoladas a partir de uma biblioteca de figado de rato em λ gt11(Fig.3). Este cDNA codifica uma proteina de 732 residuos de aminoácidos e as analises da sequencia proteica derivada de 19 peptideos CNBr e tripticos confirmou a identidade dos 292 residuos. Esta enzima verificou-se consistir

em 4 subunidades de tamanho identico baseado em estimativas de Mr para a proteina nativa e para as suas subunidades por filtração em gel e SDS-PAGE, respectivamente (Kobayashi, K. e Smith J.A., J. Biol. Chem. 262:11435-11445 (1987).

Uma vez que todas as sequencias obtidas se encontrarão na sequencia proteica deduzida é provável que as quatro subunidades sejam identicas na sua estrutura primaria.

Uma comparação de sequencia proteica deduzida (Fig.1) e das sequencias de aminoácidos derivadas da degradação automática de Edman (Tabela 3) revela que a proteina contem tres grupos lisilo equivalentemente modificados (residuos 118, 291, 443) se bem, que a natureza quimica desta modificação não tenha sido ainda determinada.

Existem tres linhas indirectas de evidencia que sugerem que o residuo de metionina na posição 1 é de facto o extremo NH₂ da proteina. Primeiro, o peso molecular calculado está de acordo com a Mr da subunidade calculada por SDS-PAGE (81,347 versus 80 000 (Kobayashi K. Smith, J.A. J. Biol. Chem. 262:11435-11445(1987), respectivamente. Segundo a composição teorica em aminoácidos é semelhante à composição de aminoácidos observada para a hidrolase de acil-peptideos purificado (Tabela 2).

Terceiro, o codão de iniciação ATG, correspondendo à metionina está no contexto certo para um codão de iniciação como descrito por Kozak, M (Cell 44 283-292 (1986)

A sequencia NH_2 -terminal da estrutura primária deduzida da enzima é Met-Glu-Arg-Gln... No entanto, a análise previa da sequencia proteica indicou que o extremo NH_2 da proteína está bloqueado (Kobayashi, K. e Smith, J.A. J. Biol. Chem. 262:11435-11445 (1987)). Se se remover a metionina durante a tradução por uma aminopeptidase específica de metionina, será de esperar que um resíduo localizado mais perto do extremo C esteja bloqueado, neste caso a sequencia do peptideo CB17-R13 (Glu-Arg-Gln...) não podia ser obtida. Portanto, o resíduo metionina permanece na cadeia polipeptídica e sofre uma modificação NH_2 -terminal. Se bem que a natureza química deste grupo bloqueador não tenha ainda sido estabelecida, a ocorrência bem documentada de resíduos glutâmilo, assim como de resíduos aspartilo e asparaginilo adjacentes às metionina acetiladas sugere que a proteína provavelmente esteja N^α -acetilada. Neste caso a Ac-Met da hidrolase de acil-peptídeos não é aparentemente clivada in vitro ou in vivo por ela própria ou outras enzimas de processamento NH_2 -terminal durante o seu processamento ou deslocação intracelular.

Foi demonstrado que a hidrolase de acil-peptídeos não remove eficazmente um aminoácido acetilado das proteínas nativas ou desnaturadas in vitro (Kobayashi, K. e Smith J.A. J. Biol. Chem. 262 :11435-11445 (1987); Gade et al., Biochim. et Biophys Acta 662: 86-93(1981) se bem que tais resíduos eficazmente clivados dos peptideos N^α - acetilados (20 resíduos). Portanto, parece provável que os substratos in vivo para esta enzima possam ser pequenos peptideos N^α -acetilados resultantes da degradação de proteínas. No entanto, não pode ser excluído um papel desta enzima na remoção de aminoácidos N^α -acetilados, de outras cadeias polipeptídicas durante o processamento co-tradução (Rubenstein, P.A., et al., J. Biol. Chem. 258: 11354-11360 (1983)).

A análise de transferências de RNA indica que um único RNA de 2.7 kb codifica a hidrolase de acil-peptídeos em todos os tecidos de rato examinados. Ainda, a quantidade de mRNA detectado nestes tecidos parece ser quase equivalente, sugerindo que não existe regulação específica de tecidos dos níveis de mRNA da hidrolase de acil-peptídeos.

Exemplo 5

Deteccão e diagnóstico de carcinoma das células pequenas

Quatro tipos principais de neoplasmas do pulmão -- carcinoma das células pequenas (também referido como carcinoma das "células de aveia") carcinoma escamoso (também referido como carcinoma epidermóide) adenocarcinoma e carcinoma das células grandes -- constituem 95% dos neoplasmas substanciais e primários do pulmão. O carcinoma das células pequenas tem substancial importância clínica. Ele constitui cerca de 25% dos neoplasmas do pulmão. Enquanto outras formas de cancro do pulmão têm tempos de sobrevivência de 5 anos em 27-37% dos casos, menos de 1% de pacientes que sofrem de carcinoma das células pequenas sobrevivem 5 anos a partir da altura do diagnóstico (ver, por exemplo, Harrison's Principles of Internal Medicine, 11th Ed. Braunwald, E. et al., eds. (1987) pag.1115-1123, cuja referência aqui é incluída como referência.).

Presentemente, o carcinoma das células pequenas é geralmente detectado através de radiografia do torax de rotina; 5-15% de tais cancros são assintomáticos na altura da detecção. Diz-se que a doença está numa fase limitada quando está confirmada /confinada a um hemi-torax e nódulos linfáticos regionais; a doença é referida como estando numa fase extensiva quando se observa maior alastramento.

A detecção precoce do carcinoma das células pequenas até associada a um aumento substancial no prognóstico. Taxas de curas de 5 anos para a doença na fase limitada são potencialmente de 15-25%, no entanto, a taxa de cura potencial de 5 anos para a doença na fase extensiva é de apenas 1-5%.

O carcinoma das células pequenas é tratado com quimioterapia e radioterapia intensiva. O objectivo inicial do tratamento é conseguir uma regressão completa do tumor em 6-12 semanas de terapia. Se bem que o tumor muitas vezes volte, o grau de regressão está relacionado com a sobrevivência a médio e longo prazo. Devido ao seu potencial metastático, o carcinoma das células pequenas não é geralmente tratável com cirurgia. No entanto, se detectado na fase inicial, é possível a remoção cirúrgica e está associado a taxas de cura significativamente maiores.

Os estudos de cariótipos revelaram uma detecção/delação constante no cromossoma 3p (p14-p23) entre o carcinoma das células pequenas (Whang-Peng, J. et al., Science 215 : 181-182 (1981). Esta observação tem sido suportada por estudos com marcas de RFLP polimórficas (Naylor, S.L. et al. Nature 329:451-454 (1987) Kok, K. et al., Nature 330:578-581 (1987); Brauch, H. et al. N. Engl. J. Med. 317:1109-1113 (1987); Yakota, J. et al. Proc. Natl. Acad. Sci. USA 84: 9252-9256 (1987). A frequência da perda de alelos indica que virtualmente todos os carcinomas das células pequenas contêm uma deleção para uma porção do cromossoma 3 (Naylor, S.L. et al., Genómico 4: 355-361 (1989), cuja referência é aqui incluída por referência).

O braço pequeno do cromossoma 3 tem sido implicado em outros cancros do pulmão, em carcinomas de células renais e no síndrome de von Hippel-Lindau (Kok, K. et al., Nature 330:578-581 (1987); Brauch, H. et al. N. Engl. J. Med. 317:1109-1113 (1987); Zbar, B et al. Nature, 327: 721-724 (1987); Kovacs, G. et al., Proc. Natl. Acad. Sci. USA 85:1571-1575 (1988); Seizinger, B.R., et al., Nature 332:268-269 (1988) cujas referências são aqui incluídas como referência.

A perda de actividade da aminocilase 1 em tumores tipo carcinoma das células pequenas e a associação familiar algumas destas doenças ainda suporta mais a correlação entre estas doenças e uma deleção no braço pequeno do cromossoma 3 (Naylor, S.L. et al., Genomics 4: 355-361(1989).

Recentemente, uma sequencia de DNA (designada "DNF 15S2") foi clonada e mapeada no cromossoma 3 (Gerber, M.J. et al. Amer. J. Hum. Genet 43: 442-451 (1988), cuja referencia é aqui incluída por referencia; Naylor, S.L. et al Genomics 4: 355:361 (1989). Encontrou-se que o DNA clonado é capaz de identificar diferenças em RFLP entre DNA normal e DNA de carcinoma das células pequenas. Em particular, foram identificados polimorfismos usando a enzima de restrição TaqI .

A sequencia de aminoácidos da hidrolase de acil-peptídeos do presente invento é substancialmente semelhante à sequencia de aminoácidos codificada pela sequencia da sonda DNF 15S2. Dos 621 resíduos de proteína DNF 15S2, 67,6% eram idênticos dos encontrados na hidrolase de acil-peptídeos do presente invento (Figura 6). Assim, as sequencias nucleotídicas que codificam a hidrolase de acil-peptídeos do presente invento ou fragmentos desta enzima, podem ser usados como sondas para detectar e identificar carcinomas das células pequenas e outros cânceres associados a uma deleção no cromossoma 3.

Quando usado como sonda, tais sequencias são incubadas em condições que permitem que lhes hibridem com DNA ou RNA de um paciente a ser testado para determinar a presença de carcinoma das células pequenas.

São conhecidos métodos de hibridação adequados (ver, por exemplo, Hames B.D. e Higgins, S.J. Nucleic Acid. Hybridization, a practical approach., IRL Press Washington, D.C. (1985), cuja referencia é aqui incluída por referencia.

Após ter sido conseguida a hibridação métodos bem conhecidos para a detecção de polimorfismo (de preferencia análise do polimorfismo do comprimento dos fragmentos de restrição ("RFLP") podem ser empregues para determinar se uma amostra de acido nucleico contem um polimorfismo ou sequencia que esteja relacionada com a presença de carcinoma das celulas pequenas ou da outros. São conhecidos muitos métodos para efectuar a análise de polimorfismo que podem ser facilmente adaptados ao emprego das sequencias codificadoras da hidrolase de acil-peptideos do presente invento na detecção de cancro das celulas pequenas de outros cancros (ver, por exemplo, Wainscoat, J.S et al., Hum. Genet 75: 384-387 (1987); Rabin, D, et al, Him Genet 75 :120-122 (1987); Azuma C, et al, Amer. J. Obstet Gynecol. 160:734-736 (1986); Pakkala, S et al., Leuk Res. 12 :757-762 (1988); Todd, S. et al, Genomics 4 :53-59(1989). Chowdhury, M.K.U., Theor. Appl. Genet. 76:25-32(1988); Yam, P. et al, Amer. J. Hum. Genet. 41 :867-881(1987); Freeman, S.M. et al, Hum. Immunol. 20:1-12 (1987); Yoffe G, et al, Exper. hematol. 15: 725-728 (1987); Jones F.S. III et al, Gene 39:77-84 (1986); Bernheim, A. , et al, Proc. Natl. Acad. Sci. (USA) 80: 7571-7576 (1983); cujas referencias são aqui incluidas por referencia).

Os ensaios de detecção de polimorfismo têm sido usados para detectar e identificar cancros (Wada, M et al, Jpn, J. Canc. Res. 78:780-784 (1987); Naylor , S.L. et al Genemics 4 :355-361(1989); Gerber M.J. et al Amer. J. Hum. Genet 43: 442-451(1988); Kakehi, Y et al., Int. J. Cancer 43:391-394(1989); Gum, J.R. et al J. Biol. Chem. 264:6480-6487 (1989), cujas referencias são incorporadas aqui por referencia). Tais ensaios podem ser usados como um modelo geral para os ensaios do presente invento.

Se bem que o que foi dito se referia a realizações particulares perferidas, será compreendido que o presente invento não lhes está limitado.

Os familiarizados com a matéria verão que podem ser feitas varias mddificações às realizações divulgadas e que tais modificações pretendem estar dentro do âmbito do presente invento.

R E I V I N D I C A Ç Õ E S

1ª.- Método para a detecção de carcinoma das células pequenas, caracterizado por compreender:

A. incubação de uma amostra de ácido nucleico de um paciente suspeito de ter carcinoma das células pequenas, na presença de uma molécula de ácido nucleico tendo uma sequência seleccionada do grupo constituído por:

i. uma sequência que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos; e

ii. uma sequência que é complementar de uma sequência que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos; a referida incubação sendo feita em condições suficientes para permitir que se dê a hibridação de ácidos nucleicos entre as referidas amostras de ácido nucleico e a referida molécula de ácido nucleico para assim formar uma molécula de ácido nucleico nucleico para assim formar uma molécula híbrida; e

B. detecção de carcinoma das células pequenas determinando se a referida molécula híbrida difere de uma molécula referencia no que respeita à sequência, a referida molécula referencia compreende uma amostra de ácido nucleico de um indivíduo normal hibridado com uma molécula de ácido nucleico que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos.

2ª.- Metodo de acordo com a reivindicação 1, caracterizado por o referido passo B compreender uma análise dos polimorfismos do comprimento de fragmentos de restrição.

3ª.- Processo para a produção de uma molécula de ácido nucleico de cadeia dupla, caracterizado por se incorporar na referida molécula:

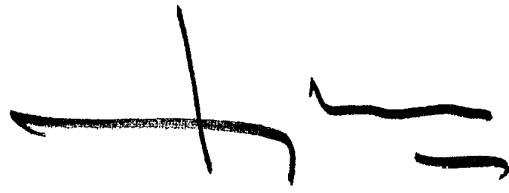
A. uma primeira cadeia tendo uma sequência seleccionada do grupo constituído por:

i. uma sequência que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos; e

ii. uma sequência que é complementar de uma sequência que codifica toda ou parte de uma enzima hidrolase de acil-peptídeos; sendo a referida primeira cadeia híbrida com:

B. uma segunda cadeia, tendo a referida segunda cadeia uma sequencia que é substancialmente complementar em sequencia de sequencia da referida primeira cadeia, sendo a referida sequencia complementar da referida segunda cadeia derivados de um individuo suspeito de ter carcinoma das células pequenas.

Lisboa, 11 de Janeiro de 1990



J. PEREIRA DA CRUZ
Agente Oficial da Propriedade Industrial
RUA VICTOR CORBON, 10-A, 1.^o
1200 LISBOA

2/23

320 340 360
AICSPRLSPDQCRIVYLQYPCPLAPHHQCSQLCLYDWYTKVTSVVVDIVPRQLGESFSGIY

77

380 400 420
CSLLPLGCWSADSQRVVFDSAQRSRQDLFAVDQTGSI TSLTAAGSAGSWKLLTIDKDLM

CB/R101
777777777777777777777777

440 460 480
VAQFSTPSPSLKVGFLPPPGKEQSVSWVSLEEAEP I PGIHWGVRVLHPPPDQENVQYA

CB18-R15
77777777 777777 777777 7

CB18-R11-13-b
77 77

500 520 540
DLDFEAILLQPSNPPDKTQVPMVMPHGGPHSSFVTAWMLFPAMLCKMGFAVLLVNYRGS

77 77777777

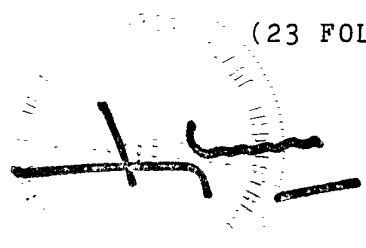
CB19-R12
777777 7 7 77

77

7

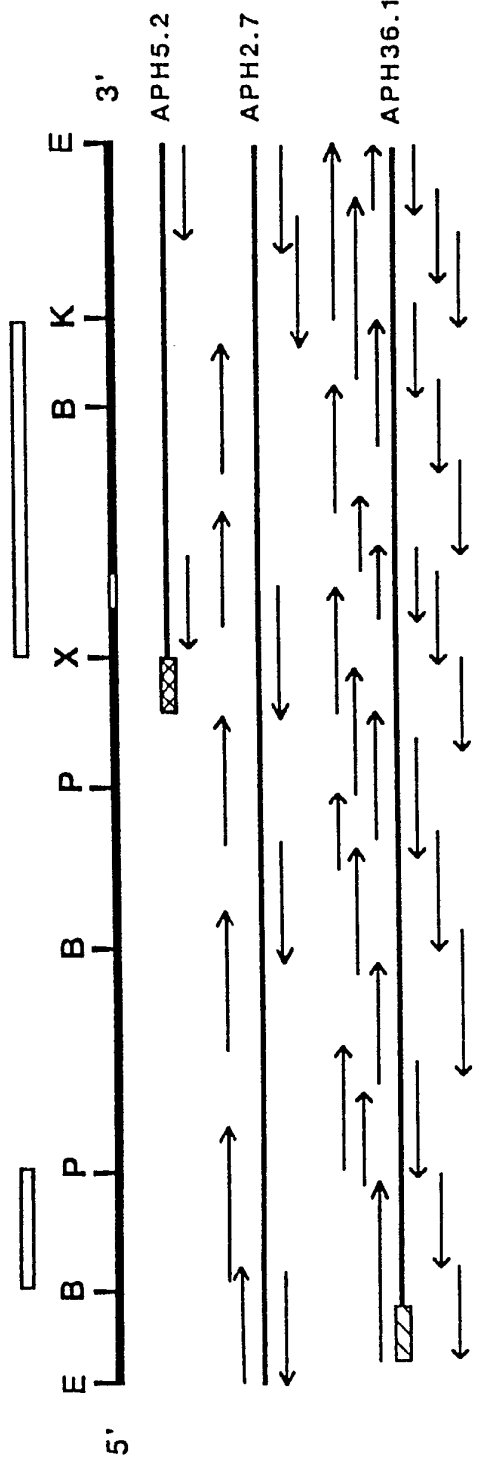
FIG. 1B

4/23



YS17.2 _____ YS20.1 _____
 Val-Leu-His-Pro-Pro-Asp-Gln-Glu-Asn-Val-Gln-Tyr-Ala-Asp-Leu-Asp-Phe-Glu-Ala-Ile-Leu-Leu-Gln-
 GTC-CTA-CAC-CCA-CCT-CCA-GAC-CAC-GAG-AAT-GTA-CAA-TAT-GCT-GAC-CTT-GAC-TTT-GAA-GCA-ATC-CTT-TTG-CAG-

FIG. 2A



0.4kb

FIG. 2B

5/23

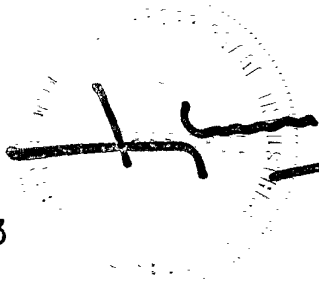
1 C G A C T A T G G A G C G T C A G G T G C T G C T G A G T G A G C C C C A G G A G G C G G C A G C G C T G T A C C G G G 60
M E R Q V L L S E P Q E A A A L Y R G
61 G C C T T A G C C C A G C C T T C G C T G A G C C C C C C T G C C T G G G T C C G G A G G T C A C C A C C A G C A G T 120
L S R Q P S L S A A C L G P E V T T Q Y
121 A T G G G G C C T C T A T C G G A C C G T G C A C A C T G A G T G G A C T C A G A G G G A C T T G G A G C G G A A T G G 180
G G L Y R T V H T E W T Q R D L E R M E
181 A G A A C A T C C G A T T C T G C C G C A G T A C C T A G T G T T T C A T G A C G G A G A C T C G G T G G T G T T T G 240
N I R F C R Q Y L V F H D G D S V F A
241 C C G G C C T G C A G G C A A C A G T G T G G A A A C C C G G G G A G C T G C T G A G C A G A G A G T C T C C T T 300
G P A G N S V E T R G E L L S R E S P S
301 C A G G C A C C A T G A A A G C T G T G C T T C G C A A G C C G G A G G C A C A G T C T C A G G A G A G G A A A A C 360
G T M K A V L R K A G G T V S G E E K Q
361 A G T T C T T G G A G G T C T G G G A G A A A C C G G A A G C T C A A G A G T T T C A A C C T G T C A G C G C T G G 420
F L E V W E K N R K L K S F N L S A L E
421 A G A A G C A T G G C C T G T C T A T G A G G A T G A C T G C T T T G G C T G C C T T C C T G G T C A C A C T C A G 480
K H G P V Y E D C F G C L S W S H S E
481 A G A C A C A C T T G T T G T A T G T G G G G A G A A G C G T C C C A A G G C T G A G T C C T T C T T T C A G A 540
T H L L Y V A E K K R P K A E S F F Q T
541 C C A A A G C C T T G G A C A T C A G T G C C A G C G A T G A T G A G A T G G C C A G G C C G A A G A G C C A G A C C 600
K A L D I S A S D D E M A R P K K P D Q

FIG. 3a

6/23

601 AGCCATCAAGGGGACCCAGTTTGTATTTTATGAAGACTGGGGAGAAACGATGGTTCCA 660
A I K G D Q F V F Y E D V G E T M V S K
661 AAAGCATCCCCTGCTCTGCGGTAGATATCGACAGTGGCAACATCTCTGTGCTTGAGG 720
S I P V L C V L D I D S G N I S V L E G
721 GAGTCCCTGAAAATGTGCACCTGGTCAGGCTTTTGGGCCCCCTGGAGACACGGGGGTGG 780
V P E N V S P G Q A F W A P G D T G V V
781 TGTGTTGGGCTGGGTCATGAACCCCTTCAGGCTCGGCACTCCGCTATTGCACCAATCGCA 840
F V G W H E P F R L G I R Y C T N R R
841 GATCAGCTCTGTAATGTGGACCTGAGCGGAGGGAAGTGAACACTACTCTCGGATGGTT 900
S A L Y Y V D L S G G K C E L L S D G S
901 CCCGTCATCTGTTCTCCCTCGGCTCAGCCAGCCAGCTGCGCATCGTCTACCTACAGT 960
L A I C S P R L S P D Q C R I V Y L Q Y
961 ATCCATGCTAGCCCTCATCAACCAATGCAGCCAGCTCTGCCCTGTACGACTGGTACACCA 1020
P C L A P H Q C S Q L C L Y D W Y T K
1021 AGGTCACCTCAGTGGTGGACATTGTACCCCGGCAGCTGGGAGAGAGCTTCTCTGGAA 1080
V T S V V V D I V P R Q L G E S F S G I
1081 TCTACTGCAGTCTTCTTCCCTTTGGGATGCTGGTTCAGCTGACAGCCAGAGAGTAGTCTTG 1140
Y C S L L P L G C W S A D S Q R V V F D
1141 ACTCTGCCAGCGCAGTCCGACGACCTGTTGCTGTGGACACCCAGACGGGCAGCATCA 1200
S A Q R S R Q D L F A V D T Q T G S I T

FIG. 3b



7/23

1201 CCTCCCTGACAGCTGGGGTTCAGCAGGAAGCTGGAAGCTGCTTACAATTTGACAAAGACC 1260
S L T A A G S A G S W K L L T I D K D L

1261 TTATGGTTGCCAGTTCTCCACACCCAGTTTGCCCCCAAGCCTGAAAGTTGGGTTCCCTGC 1320
M V A Q F S T P S L P P S L K V G F L P

1321 CTCCTCCTGGGAAGGAGCAGTCTGTGTCATGGGTGCCCTGGAGGAGGCTGAACCTATT 1380
P P G K E Q S V S W V S L E E A E P I P

1381 CTGGCATCCACTGGGGCGTCCGTGTGCTACACCCACCCCTCCAGACCAAGAGAATGTACAAT 1440
G I H W G V R V L H P P P D Q E N V Q Y

1441 ATGCTGACCTTGACTTTGAAGCAATCCTTTTGCAGCCCCAGCAATCCTCCAGATAAGACCC 1500
A D L D F E A I L L Q P S N P P D K T Q

1501 AGGTCCCAATGGTAGTCAATGCCCATGGGGACCCCATTCGTCTTGTCTACCTGCTGGA 1560
V P M V V M P H G G P H S S F V T A V H

1561 TGCTGTCCAGCCATGCTCTGCAAGATGGGCTTTGCAGTACTTCTGTGTAACATCGTG 1620
L F P A M L C K M G F A V L L V N Y R G

1621 GCTCCACTGGCTTTGGCCAAGACAGCATCCTTTCTCCAGGCAATGTGGCCACCAGG 1680
S T G F G Q D S I L S L P G N V G H Q D

1681 ATGTGAAGGATGTCAGTTTGCAGTGGAGCAGGTACTCCAGGAGGAACAACCTTGATGCAA 1740
V K D V Q F A V E Q V L Q E E H F D A R

1741 GACGGGTGGCCCTTATGGGTGGTTCCCATGGTGGGTTCCCTCTCCACCTGATTGGTC 1800
R V A L M G G S H G G F L S C H L I G Q

FIG. 3c

8/23

1801 AATACCCGGAGACCTACAGTGCCTTGTTATAGCCCGGAACCCAGTGATCAACATTGCCCTCCA 1860
Y P E T Y S A C I A R N P V I N I A S M
1861 TGATGGGCTCCACTGACATCCCTGATTGGTGTATGGTGGAGACTGGCTTCCCTTATAGCA 1920
M G S T D I P D W C M V E T G F P Y S N
1921 ACAGCTGCCCTCCAGATCTGAATGTGGGAGGAGATGCTGGATAAGTCACCCATCAAAT 1980
S C L P D L N V W E E M L D K S P I K Y
1981 ACATCCCTCAGGTAAGACACCAGTGTACTGATGCTGGGCCAGGAGGACCCGCGGTAC 2040
I P Q V K T P V L L M L G Q E D R R V P
2041 CCTTCAAGCAGGGTATGGAGTATTACCGTGCCCTCAAGGCCCGGAATGTGCCCTGTCCGGC 2100
F K Q G M E Y Y R A L K A R N V P V R L
2101 TCCTGTGTACCCCAAGAGCAACCACGCGCTGTCAGAGGTGGAGGCGGAGTCAGACAGCT 2160
L L Y P K S N H A L S E V E A E S D S F
2161 TCATGAACGCTGTGCTTTGGCTGCACACACTTGGGCAGCTGAAGCCCTATGTTCTTCT 2220
M N A V L W L H T H L G S *
2221 TGAGCTGGGCAGCCCGTCTTGCACCTTGGTCTTGGTAAGCTCCAGGATCTGACCGAACAG 2280
2281 CCAAACAGTCTCCTAGGCTGTGGACTCAGTGAATGGGCAGGACTAAGAACTATGTTAT 2340
2341 CATAATAAGGAAGACACAGAAAAAATAAAAAAAAAAAAAA 2381

FIG. 3d

1380 c t c c c c t g t a c t g a c a a g g t a t t t c t g t g t a t c a g a a t a a a g g a c t t g a t a t a a c a a c c t c c a g a c t g t c t t t g g
1300 g a t g a t t t c t g a a g t g g a a t g c t t a g c c a a c a a g t a a a t g c c a a t g c t t g t g a c a a t a t c c t c a a t g a t t a a a g a c g a
1220 g a g a g a g a g a g a g a g a g a g a g a g a g a a c a a c g t t c a g a a a t g t t g g t a c a c a a c t g g g g c t
1140 a a g t t a a a a t g a t g t c c a g g a a g g t a a a g g c c t g t c c c c a c c a a t g g g t t c c a c c a g t t t c a g g g t t c t g g t
1060 t c c c a g g a g t a g t g t g c c c c c t g g t g c t c a a g c a a g c t t t a c a t t t g t c t g c t c c a a c c t c a g a a t c c t c c
-980 c c c c c c c c c a g g c t a a g a c c t t a a t t c c a c a t a t a g t c a a g a g t a g t t c a a g c c g g g t g t c a t a g c c c a c a c t t t t
-900 t t t a a a a a g a t t a t t a t c a t a t a t a a g t a c a c t a t a c c a g t c t c a g a c a c a c a g a g a g g c a g t g t a t c t c a
-820 t t a c a g t a g g c t g t g a c c a c c a t g t g g t g t g g a t t g a g c t c a g g a c c t c t g g a a g a g c a g t c a g t c t t a a c c
-740 a c t g a c c a t c t c c a g c t c a t a g c c a c a c t t t t a a a t t t t a t t t a t t t a t g t a t a t g a c g a c a c t g t a g t g t
-660 c t t c a g a c a c a c c a g a a g g g t a t t g g a t c c c a t t a c a g a t g g t g t g a g c c a c c a t g t g g t t g a c t c a g g a c c t c t
-580 g g a a g a g c a g t c a g t c t c a a c c g t a g t c t a g t c a t c t c t c c a g a t c c c a a a g t c c a c a t t t t a a t c c c a g c t c t t
-500 g g g a g c a g a t a g g t g g g t c t c t g t g a g t a g a g c c a g g a c t a t g t a g a a a c t a a a c a a t a t a a g t c a a g a t c a
-420 c t a g a a a t a a g a a c a g a g a c t a a a a a g t g a t c a a t t t g c c t c g g a a g a a a g a a g t a c t c a t t c a a g g g a c c
-340 c a t a g c t g g t c t g a a a g a g c a g c c a g t c t c a a g g a t a t g g a g a a g g g a g a t g c c t a c t g t t g c c c a a c c t c t c t g c c
-260 c t c t g t g t c a c c t a a g t t c a a c c t a g t c t g t g g t g t g c a a g a g a g a c c g c c t g t c t a g g c t c t t c t t c c a c t g c

FIG. 4a (cont.)

80 acctgctcttccgtgaagtctgaccaccaccacatctgctcttggagtggctgactcagctcagatcagtcagc
00 atcaattgagtcactgaaaaggcggttaaggctgtagcaacagatagcttcggggcaagtctgggtggataaaccg
20 gggtagcaccagagattcAGCACCGCCCTTTGCCCGCCAGGGTCCGCCCGGGACCTGATTTTCTCACTCCTA
61 ACCGCCTGGCTTGCAGTAAATGGGGAGAGAGCTGAAAGAGATCCCTGGGAGTCCCGGCCCCGCAACGTTTCGACTTCAT
141 GGACCTTCCACCCACCCCAAATCCCCTTGTCCCTCATCCTTCAGGATCATTTGGCGTTCCTCTTAAGCACAGCCCTTC
221 GAATGAATACTCAGAGTTTTCTCTGACCCCGCAGCTCACAGtcggctgagtcctgctacttcttcaactcaatccc
301 agacctcccctatcccggaaacaccagctctgtttctggttacaattctgcccctcctcctcggaggcttcccctct
381 ctaaaaatcccaggaccgactcccagggttctgctcctccgacttgccaatcccgctcctcccaggctcctcgtagccc
461 cgcctcataccaccagaccgccccagCTGGTTGGCCCGAGCTTGGTGGCCCTGCAGGTCCCAGGTCGCCACT
541 GGGTGTAGCCGAAACTGGGACCCGCTTAGGCCCGCCCTCGGAGCCTCCTCCGGACTTAGCCGCTACAGAAAGATC
621 GACTATGGAGCGTCAGgtgaggctcggccctggtccctggtccctgtagccggctgagcctgacctgggtgctc
701 ccctgcagGTGCTGAGTGTGAGCCCCAGGAGCGCGCTGTACCAGGCGCTTAGCCGCCAGCCTTCGCTGAGCG
781 CCGCCTGCCTGGTCCGGAGGTCACCACGCAGTATGGGGCCTCTATCGGACCCTGCACACTGgtgtgtagaccgagggg
861 actgggtggcgcggccgtgcatcccaggctgggactgtggaaaggcagctgtgacacgggcttggggaggaaaccaa
941 ctgtagttcctgcgcaagggttcttgggggtgtgtgtagccccgtaaatgaaagtgtgggagtaaaatccagac

FIG. 4b

4621 ctgtttctccatttggaagggtgatcaagtgaccacctcttgagctcctgagtgagctgtgtactctgcaactctggtg
 4701 gataagtacgttatactaaacggttaaggcaactbacaatgccaagaccacaaagggaatagccaaataggaaagcatcg
 4781 agttcgcaataaagctcaacagctctgcctctccagctcctcagaaggttcccccccaacccccctcaacttctactcag
 4861 taaactcctcatgccaaggtaacctcaatgtctaggactatctacctcagatctcaggggcccctgggtgaaatccccca
 4941 cttcataatctcgaatccccaggccccctcctaagatggcagcttaacaccttctaccttgcagAGAGCTTCTCTGGAAT
 5021 CTACTGCAGTCTTCTTGGGATGCTGTGCTCAGCTGACAGCCAGAGACTAGTCTTTGACTCTGCCCAAGCCAGTCCGC
 |---intron 13
 5101 AGgtaaaagcctctgatgtgtggaaggtagcaataatctgcctggggcaaatgtagtaatgtggccctaggaggatg
 5181 ccgagcctgaaagctctaaactggctccaagaagcctttctgcctcttgctggacttagtttctccttaagaggaccacaaat
 5261 gtgggttgaggatttagctcagtggttagagcaacttgccataagcaaggcgaaggccctgggttcgggtcccccagctccga
 5341 aaaaaaagaaaaggaaaggaccacaaatgtttctgatgccccacagaagagcccccaagctttggggtgtgtctctgagat
 5421 gggccccctgaacacttggtaatggcagttgatggagacaggatgggcccccttgagggatgatgtgccgggatgtgccatgg
 5501 tcctttaggtatgataatctctgttctcataatcttctcgtctccttgcttgatatactagtaacttgtgaaggaaattgaat
 5581 ctagatccagggttatccccgttttacacagtagccatcaccttaggccaatgtccccctaaagacagtagaaaggacatcgg
 5661 gctgttgatgtgctgctacaagtggtctccccactgttttaagACCTGTTGCTGTGACACCCAGACGGCCAGCATCA
 |---intron 14
 5741 CCTCCCTGACAGCTGgtgagcaggaccctctgcagtagggtagtggtggctgggtggatggcagcttaggaaagcgcta

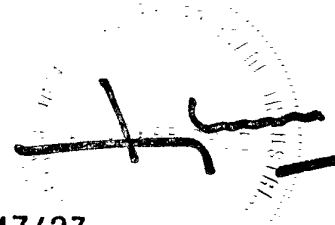
FIG. 4d

5821 gagcttggctggcattggcagcagtggtgtgctccacacgcccctggcccatagccccaccctgtcctaaaggtcacaga
5901 cctttcctgcaatctattgtagtcatccacctctgcccagaccctggagaaaccttgggtggaactctctagcattcctg
5981 cactctcccctatggtctgcctcagaagcagccttttgagtccctccaccgacctggactctgcaactctgcccatactctgct
6061 ttctctagggcgtcagcagggagctggaagacagaaagatgcttctcgggttccattgctgtgacgagacaccatg
6141 atcgtggctcctcttaataaggagacatttaattggggctggcttacagatccagaggttttagtccgttatctgtcatggc
6221 agggagcgggcagcgtcagcagattggctggcagcagaaagacagtgaagtagacctgacttgagcttctgaaacc
6301 tcaagcccacctcaatggcgcactttaaacagcggccaaagcaccctcttaatagtgccactcccctatgaggctactggg
6381 ggccatttttatcaaacactacaggtgataagtcctaatcttgaccctcattactagggttctccctagggtataca
6461 aagcctcgtgcacataaataagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagt
6541 agtagtagtagcagcagtaatacacaagcctagtaaaagcctaaatcacgtaagtaaaaggtaataataataatgg
6621 ataagcttagcccagctcctcttctccaccagcttcttgggctaggctgctgttcccttaccgttgcagctcttggga
6701 gtaaatagagtaccttggtagcttctgtgtccccaggtctggcaattcttgagtaataatcagtgctctcagatgaaca
6781 ggaggcttaccaggacatgtgcagttaggtagaaagtggtggccctgcccagcacttacaatcactgcttccctctagC
6861 GGGGTCAGCAGGAAGCTGGAAAGCTTACAATTGACAAAGACCTTATGTTGCCCCAGTTCTCCACACCAGTTGCCCC
|---intron 15
6941 CAAGCCTGgtgagtatgtgggtctcaggctagagccggatataacagactccatgccaactggaactgggctcgaagactg

FIG. 4d(cont.)

7021 accgtggctggccagcatgtagaaggcagagaggacttggctccagctagctctgtggttcttgggctcggcactcaa
7101 gagtagaacctaggccagcacagtgctctacagaagtgccagtcattattattagtagccagggaacaagaaaacc
7181 tagaagccaagaaatgtggtagggctagggtggccataatgtggatgtctattggcattatgtagctctctgtagaaatg
7261 ggctctgagagctggccgcctgtgactttccttccctggcagAAAGTTGGGTTCCCTGCCTCCTGGGAAGGAGC
7341 AGTCTGTGTCATGGGTGCCCTGGAGGAGGCTGAACCTATTCCTGGCATCCACTGGGGCGTCCGTGTGCTACACCCACCT
7421 CCAGACCAAGAGAAATGTACAATATGgtgagctggccgatggcgggtggaaggctctccagacccccacacacagcatic
7501 tgtggcgacccttccctagctcaagcttgaggcttcttaggatttccctctgaaatttcttctgactctatttggtcatgggga
7581 gtagatccaaagatctgagtaacctgtgtggttatacagcactgggtcctggcagggaatactctctgtggtgagaaagt
7661 taagatttaaaggaaagtaggaccttccagggctgctgactacacagaaatggtattcctccagacaataaggctctct
7741 gacaaggcagaaaggcacaagcttttaggggctccatgtgccaagggggagctgacctaggcctaccctggctgggct
7821 ctgttctactgtgacaggtttgtatcagacctgcagCTGACCTTGACTTTGAGCAATCCTTTTCAGGCCCAGCAATCCT
7901 CCAGATAAGACCCAGGTCCCACATGGTAGTCATGCCCCATGtaggtatttacctaagagccttgccaccctaccaccct
7981 cctagctggccttcatbgtgctgcttctgtctgcagggagagaaggcagtagtcaagaagaccagtcctgcttccaaa
8061 ggcttcttagccaagtcccatctgcagGGGACCCCATTCGTCTCCACTGCTGGATGCTGTTCCAGCCATGCT
8141 CTGCAAGATGGGCTTTGCAGTACTTCTGGgtaagtcagaacggaccctgcaattccaactggggtccaatgaagcagatggggt

FIG. 4e



Handwritten signature or initials.

8221 ggcctgggtgtatataccttttctatggtgtacttactcttttagaaacgggctgggcaatgggggacactccttttaatc
8301 ccagcacttagaaggcagagggtgaagttagtccaaggatcatcaggactacacagaaaaccctgttttgaaaa
8381 caaaaagaactactactagaaaattgccctggcttagcctctggttcgtaacctggtgatcagggtgataacttgaagc
8461 ttctgtgaaaggatcctatgggtctctbgttagcagcttcatctctctbgttagcctcccttgacagcagtgatgacag
8541 ctggcttagggcctatgtcacagccacatggaaatgggttggggactgagaccattgcccacaactatctbgtctcacttt
8621 cagTGAACtATCGTGGCTCCACTGGCTTGGCCAAAGACAGCATCCTTCTCCAGGCCAATGTGGCCACCAGGATGTG
1---intron 19
8701 AAGGATGTCCAGgttgaggcgggtgggtctgtggacagggagggtgatcaggcttctggctgcccagcagcgaact
8781 gtagtgctctctgccacactttagTTGCAGTGGAGCAGGTACTCCAGGAGGAACACTTTGATGCAAGACCGCGTGGCCCC
8861 TTATGGGTGGTTCCTCCATGGTGGGTTCCCTCCTGCCACCTGATTGGTCAATACCCGGAGACCTACAGTGCCTTGTATAGCC
8941 CGGAACCCAGTGATCAACAATTGCCCTCCATGATGGGCTCCACTGACATCCCTGATTGgtgagtgatcattagagtcctgccc
9021 ctctctccacccttccatggagatcctgtatcttcttgggtctttagGTGTATGGTGGAGACTGGCT
9101 TCCCTTATAGCAACAGCJGCTCCAGATCTGAATGTGTGGAGGAGATGCTGGATAAGTCACCCATCAAAATACATCCCT
1---intron 21
9181 CAGgtactcagggttctgtgtgtgcccaccctctctccagagtcaggctcaccacagctccttgcctcca
9261 atatacccagGTAAAGACACCAGTGTACTGATGCTGGGCCAGGACCCGAGTACCCTTCAAGCAGGGTATGGAGT
1---intron 22
9341 ATTACCGTGCCTCAAGGCCCGGAATGTGCCTGTCCGgtgagtgatcacagggtagtctcgggcagctggtgacagggtaaac

FIG. 4e (cont.)

20/23

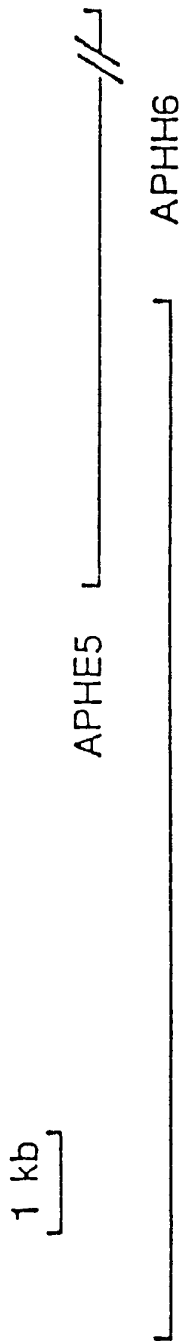


FIG. 5a

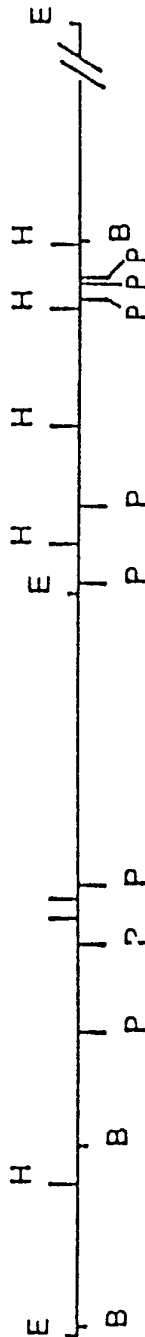


FIG. 5b

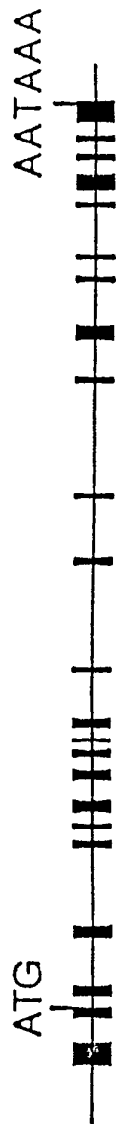


FIG. 5c

