DESCRIÇÃO

DA

PATENTE DE INVENÇÃO

N.º 100 037

REQUERENTE: NEW YORK UNIVERSITY, norte-americana, com sede em New York, New York 10016, Estados Unidos da América

EPÍGRAFE: "Processo de expressão-clonação para detectar proteínas alvo para tirosina-quinases de receptores, sonda polipeptídica e proteínas GRB humanas"

INVENTORES: Joseph Schlessinger, Edward Y. Skolnik e Benjamin L. Margolis

Reivindicação do direito de prioridade ao abrigo do artigo 4.º da Convenção de Paris de 20 de Março de 1883.

Estados Unidos da América em 18 de Janeiro de 1991 sob o nº 07/643 237

INPI MOD 113 RF 15732

73 578
SCHLESSINGER 2 PORT

PATENTE Nº 100 037

"Processo de expressão-clonação para detectar proteínas alvo para tirosina-quinases de receptores, sonda polipeptídica e proteínas GRB humanas"

R E S U M O

O presente invento refere-se a um novo processo de expressão-clonação para detectar e identificar proteínas capazes de se ligarem a domínios, fosforilados na tirosina, de tirosina-quinases de receptores, usando uma nova sonda. Esta sonda compreende uma sequência de aminoácidos derivada da porção fosforilada na tirosina da molécula de receptor, ou de um seu derivado funcional, e tendo, pelo menos, um resíduo de tirosina fosforilado e sendo desprovida da porção tirosina-quinase do receptor, e é marcada detectavelmente.

O invento refere-se também ao processo de preparação da sonda, a um processo de mapeamento, num cromossoma, de um gene que codifica uma proteína capaz de se ligar a domínios, fosforilados na tirosina, de tirosina-quinases de receptores e a um processo para purificar essa proteína com a sonda.

O invento refere-se ainda a duas proteínas obtidas pelo processo de clonação referido anteriormente, GRB-1 e GRB-2, ao ADN que codifica estas proteínas e a processos para detectar estas proteínas.

MEMÓRIA DESCRITIVA

ANTECEDENTES DO INVENTO

Campo do Invento

O invento, no campo da biologia molecular e celular, refere-se a um novo processo, baseado na clonação por expressão directa, para identificar proteínas celulares capazes de se ligarem ao domínio do terminal carboxi de tirosina-quinases de receptor e que servem como substrato para esta enzima em vias de transdução de sinal maior. O invento refere-se também a novas proteínas identificadas usando este processo.

Descrição da Arte Anterior

Vários factores de crescimento polipeptídicos e hormonas medeiam os seus efeitos celulares interactuando com receptores da superfície celular possuindo actividade enzimática da tirosina-quinase (para uma revisão ver Williams, L.T. et al., Science 243:1564-1570 (1989); Ullrich, A. et al., Cell 61:203-212 (1990); Carpenter, G. et al., J. Biol. Chem. 265: 7709-7712 (1990)). A interacção destes ligandos com os seus receptores induz uma série de eventos , que incluem a dimerização e estimulação da actividade da proteína tirosina-quinase. Para o receptor do factor de crescimento epidérmico (EGFR), bem como para outros receptores com actividade da tirosina-quinase, tais como o receptor do factor de crescimento derivado de plaquetas (PDGFR), a activação da quinase e a autofosforilação do receptor resultam na associação física do receptor com vários substratos citoplásmicos (Ullrich, et al.,supra).

Foram agora definitivamente identificados, em células vivas, dois substratos para a quinase de EGFR: (a) a fosfolipase C-gama específica de fosfatidilinositol (PLC-gama) e (b) a proteína activadora de GTPase (GAP), uma proteína que pode estar no anel

("loop") efector da proteína ras (Margolis, B. et al., Cell 57: 1101-1107 (1989b); Meisenhelder, J. et al., Cell 57: 1109-1122 (1989); Molloy, C.J. et al., Nature 342: 711-714 (1989); Wahl, M.I. et al., J. Biol. Chem. 265: 3944-3948 (1990); Ellis, C. et al., Nature 343: 377-381 (1990); Kaplan, D.R. et al., Cell 61: 121-133 (1990)).

Similarmente, mostrou-se que o PDGFR activado fosforila a tirosina e que se associa com PLC-gama, GAP, e tirosina-quinases celulares tais como $pp60^{src}$ (Gould, K.L. et al., Molec. Cell Biol. 8: 3345-3356 (1988); Meisenhelder, J. et al., Cell 57: 1109-1122 (1989); Molloy, C.J. et al., Nature 342: 711-714 (1989); Kaplan, D.R. et al., Cell 61: 121-133 (1990); Kazlauskas, A. et al., Science 247:1578-1581 (1990); Krypta, R.M. et al., Cell 62: 481-492 (1990); Margolis, B. et al., Science 248:607-610 (1990)). Embora os sítios exactos responsáveis pela associação de EGFR com PLC-gama ou com GAP não tenham sido completamente clarificados, iniciaram-se recentemente os trabalhos para identificar as regiões do substrato e do receptor que contribuem para a associação.

Os domínios SH2 ("src homology 2") (homologia src 2) parecem ser as regiões responsáveis pela associação de vários substratos de tirosina-quinase com receptores de factores de crescimento activados. Os domínios SH2 são sequências conservadas de cerca de 100 aminoácidos encontradas em tirosina-quinases citoplásmicas de não receptores tais como pp60src, PLC-gama, GAP e v-crk (Mayer, B.J. et al., Nature 332:272-275 (1988); Pawson, T. Oncogene 3:491-495 (1988)). Embora tenham domínios catalíticos distintos, todas estas moléculas partilham domínios SH2 e SH3 ("src homology 3") conservados e a capacidade de se associarem com receptores com actividade da tirosina-quinase (Anderson, D. et al., Science 250:979-982 (1990)).

A activação da tirosina-quinase e a autofosforilação do receptor são requisitos prévios para a associação entre receptores de factores de crescimento e proteínas contendo o domínio SH2 (Margolis, B. et al., Mol. Cell. Biol. 10: 435-441

(1990); Kumjiam et al., Proc. Natl. Acad. Sci. USA 86: 8232-8239 (1989); Kazlauskas, A. et al., Science 247:1578-1581 (1990)). Em particular, o fragmento carboxi-terminal (C-terminal) do EGFR, que contém todos os sítios de autofosforilação conhecidos, liga-se especificamente aos domínios SH2 de GAP e de PLC-gama (ver abaixo). Assim, existe um sítio major de associação entre o domínio SH2 destas proteínas substrato e a cauda C-terminal fosforilada em tirosina do EGFR.

Com o reconhecimento de que a ligação ao receptor de tirosina-quinase activada é conservada entre várias proteínas substrato, têm sido feitos esforços para identificar substratos adicionais que partilhem estas propriedades. As proteínas alvo que se ligam a receptores activados têm sido identificadas por análise de proteínas que co-imunoprecipitam com receptores de factores de crescimento ou que se ligam a receptores fixados em matrizes imobilizadas (Morrison, D.K. et al., Cell 58:649-657 (1989); Kazlauskas, A. et al., EMBO J. 9:3279-3286 (1990)). Embora a identidade de algumas destas proteínas seja conhecida, muitas outras detectadas por estas abordagens não foram completamente caracterizadas. Além disso, é possível que raras moléculas alvo que interactuam com receptores activados não tenham sido detectadas devido à sensibilidade limitada destas técnicas; a estequiometria real de ligação pode ser baixa, e a solução detergente necessária para solubilizar proteínas pode quebrar a ligação.

As abordagens convencionais para isolar e clonar estas proteínas têm sido difíceis, necessitando do uso de grandes quantidades de tecido ou de linhas celulares para purificar quantidades suficientes de proteína para análise por microsequenciação e subsequente clonação de ADNc convencional. Deste modo, reconhece-se na arte que há necessidade de novas abordagens para a clonação e posterior isolamento e identificação destas proteínas.

SUMÁRIO DO INVENTO

Uma das necessidades mais prementes no nosso esforço para compreender e tomar o controlo sobre a regulação do crescimento celular e da oncogénese é a capacidade de identificar proteínas alvo para tirosina-quinases.

Os presentes inventores desenvolveram um novo sistema de expressão/clonação para a clonação rápida de proteínas alvo para receptores celulares do tipo da tirosina-quinase. O processo de clonação baseia-se na capacidade de uma certa classe de substratos se ligarem especificamente ao término carboxi (término C) fosforilado em tirosina do receptor do factor de crescimento epidérmico (EGFR) (ver Exemplo VI, abaixo; Margolis, B. et al., EMBO J. 9:4375-4380 (1990)).

A abordagem concebida pelos presentes inventores tem vantagens importantes em relação aos processos de clonação convencionais, incluindo a de evitar o trabalho laborioso e dispendioso de purificar potenciais proteínas alvo para análise por microsequenciação. Além disso, a abordagem do presente invento proporciona um processo para identificar moléculas alvo raras cuja associação com receptores de activação não podia ser detectada usando técnicas convencionais. Acresce ainda que este processo permite a identificação de proteínas estrutural ou funcionalmente afins, as quais, embora só fracamente homólogas ao nível do ADN, são similares nas suas propriedades de ligação a receptores activados com actividade da tirosina-quinase. Esta última capacidade é importante visto que os processos de pesquisa convencionais usados para identificar genes afins são tipicamente baseados em hibridação de ácido nucleico de baixa severidade ("strigency"). De facto, essa pesquisa baseada na hibridação não teria tido sucesso na clonação e identificação das novas proteínas do presente invento, GRB-1 e GRB-2, devido à sua falta de semelhança ao nível do ADN.

A abordagem dos presentes inventores tem as suas raízes nos processos anteriormente utilizados para clonar proteínas de

ligação de ADN e de ligação de jun (Singh, H. et al., Cell 52: 415-423 (1988); MacGregor, P.F.et al., Oncogene 5:451-458 (1990)). Os processos do presente invento tiram partido da concepção dos inventores e da observação de que o término C da proteína EGFR no qual os resíduos tirosina estão fosforilados podem ligar substratos (ver abaixo). Ao criarem uma sonda marcada, na qual as tirosinas estão fosforiladas com ^{32}P, os presentes inventores pesquisaram bibliotecas de expressão de ADNc de lambda gt11 de vários tecidos humanos e identificaram numerosas proteínas que se associavam ao término C fosforilado do EGFR. Os presentes inventores designaram as proteínas observadas desta maneira por "GRB" (Growth factor Receptor Bound, "ligante de receptor de factor de crescimento"). A metodologia de clonação do presente invento foi designada por "CORT" (Cloning Of Receptor Targets, "Clonação de Alvos Receptores").

O processo do presente invento é proposto como uma nova abordagem possuindo tanto generalidade como rapidez de identificação e clonação de moléculas alvo para tirosina-quinases.

O presente invento dirige-se, assim, a um processo para detectar a expressão de uma proteína, que é um alvo de uma tirosina-quinase de receptor, por uma célula albergando um vector de expressão, sendo a proteína capaz de se ligar a uma porção polipeptídica fosforilada em tirosina da tirosina-quinase de receptor, compreendendo o processo:

(a) fazer contactar a célula, um seu extracto, um seu lisado ou um seu sobrenadante com um portador em fase sólida, provocando a ligação da proteína ao portador;
(b) incubar a proteína ligada ao portador com o polipéptido fosforilado em tirosina, permitindo que o polipéptido se ligue à proteína ligada ao portador;
(c) remover materiais não ligados ao portador; e
(d) detectar a presença ou medir a quantidade de polipéptido fosforilado em tirosina ligado ao portador,

detectando-se assim a expressão da proteína.

Numa concretização preferida, o receptor é o receptor do factor de crescimento epidérmico, o receptor do factor de crescimento derivado de plaquetas ou o receptor do factor de crescimento de fibroblasto.

Este processo é preferivelmente realizado usando uma célula procariótica, mais preferivelmente uma célula bacteriana tal como E. coli. A célula também pode ser eucariótica, tal como uma uma célula de mamífero ou levedura.

Preferivelmente, o polipéptido fosforilado é marcado detectavelmente, tal como com o radiomarcador ^{32}P.

O portador em fase sólida é preferivelmente uma membrana de nitrocelulose, para a qual são transferidas proteínas libertadas das células bacterianas lisadas quando está a ser pesquisada uma biblioteca de lambda gt11.

O presente invento proporciona também um processo para mapear num cromossoma humano um gene que codifica uma proteína que é capaz de se ligar a uma porção polipeptídica fosforilada em tirosina de uma molécula de tirosina-quinase de receptor, compreendendo o processo:

(a) infectar células bacterianas com uma biblioteca de expressão de genes humana;
(b) detectar um clone que expresse a proteína usando um processo de acordo com a reivindicação 1;
(c) sequenciar o ADN do clone; e
(d) mapear a sequência num cromossoma humano.

O presente invento refere-se também a uma sonda polipeptídica útil na detecção da expressão de uma proteína capaz de se ligar a uma porção polipeptídica fosforilada em tirosina de uma tirosina-quinase de receptor. A sonda compreende uma sequência de aminoácidos derivada da porção fosforilada em tirosina da molécula de receptor, ou um seu derivado funcional, é desprovida do domínio de tirosina-quinase e a sequência deve

conter pelo menos um resíduo fosfotirosina, preferivelmente 4 ou 5 fosfotirosinas. A sonda deve estar marcada detectavelmente, preferivelmente com ^{32}P.

Uma sonda preferida tem entre cerca de 25 e 250 resíduos de aminoácidos.

A sonda do presente invento é útil para detectar proteínas alvo para tirosina-quinases de receptor incluindo receptor do factor de crescimento epidérmico, receptor do factor de crescimento derivado de plaquetas e receptor do factor de crescimento de fibroblasto.

O presente invento também inclui um processo de preparação da sonda anterior, compreendendo:

(a) proporcionar o receptor ou um derivado do tipo receptor construído por engenharia genética, sob forma substancialmente pura, tendo o receptor ou o derivado do tipo receptor um domínio de tirosina-quinase e um domínio fosforilado em tirosina, incluindo o domínio fosforilado em tirosina pelo menos um resíduo tirosina capaz de ser fosforilado pela tirosina-quinase;

(b) incubar o receptor ou o derivado do tipo receptor com [gama-^{32}P]adenosina-trifosfato sob condições que permitam a fosforilação do resíduo tirosina, causando a fosforilação do resíduo tirosina

produzindo-se assim a sonda. Numa concretização preferida, o processo inclui o passo de:

(c) tratar adicionalmente a molécula de receptor fosforilada com um agente capaz de clivar a molécula entre o domínio da tirosina-quinase e o domínio fosforilado em tirosina.

Um agente de clivagem preferido é brometo de cianogénio.

Noutra concretização, o processo anterior envolve um derivado do tipo receptor construído por engenharia genética, o qual é um polipéptido codificado por uma molécula de ADN compreendendo uma sequência de ADN que codifica tirosina-quinase,

ligada a uma sequência de ADN que codifica um sítio de clivagem enzimática selectivo, ligada a uma sequência de ADN que codifica o domínio fosforilado em tirosina e sendo o agente uma enzima capaz de clivagem neste sítio de clivagem. As enzimas preferidas são Factor Xa e trombina.

É também proporcionado um processo para purificar, a partir de uma mistura complexa, uma proteína que é capaz de se ligar a uma porção polipeptídica fosforilada em tirosina de uma molécula tirosina-quinase de receptor, compreendendo o processo:

(a) fazer contactar a mistura complexa com um portador em fase sólida ao qual está ligada uma sonda, permitindo que a proteína se ligue à sonda;
(b) remover os materiais não ligados ao portador; e
(c) eluir a proteína ligada do portador,

purificando assim a proteína.

O presente invento dirige-se também a uma proteína, GRB-1, possuindo a sequência de aminoácidos apresentada na Figura 4. O invento inclui uma proteína, GRB-2, que inclui a sequência de aminoácidos apresentada na Figura 17.

O invento dirige-se também a uma molécula de ADN codificando a proteína GRB-1 e a uma molécula de ADN codificando a proteína GRB-2. Estão incluídas moléculas de ADN que codificam derivados funcionais destas proteínas. Quando a molécula de ADN ocorre naturalmente, ela é substancialmente desprovida das sequências nucleotídicas com as quais está nativamente associada. As moléculas de ADN deste invento podem ser veículos de expressão, tal como plasmídeos.

Proporciona-se também um hospedeiro transformado com cada uma das moléculas de ADN anteriores.

O presente invento também inclui um processo para preparar a proteína GRB-1 ou a GRB-2 substancialmente isentas de outras proteínas com as quais cada uma está nativamente associada, ou um

seu derivado funcional, compreendendo:

(a) cultivar uma célula hospedeira capaz de expressar a proteína sob condições de cultura;
(b) expressar a proteína ou o derivado funcional; e
(c) recuperar a proteína ou o derivado funcional da cultura.

BREVE DESCRIÇÃO DOS DESENHOS

A Figura 1 é um padrão de mancha no filtro mostrando que o término carboxi do EGFR interactua com GAP-SH2 imobilizada em filtros de nitrocelulose. Proteína de fusão trpE/GAP-SH2 expressa bacterianamente ou trpE, como controlo foi colocado a várias concentrações em filtros de nitrocelulose. Os filtros foram hibridados durante a noite com domínio C-terminal do EGFR, marcado com [^{32}P]. Auto-radiografou-se durante 2 horas.

A Figura 2 é um diagrama esquemático representando o processo de clonação dos alvos receptores (CORT). O domínio C-terminal do EGFR é fosforilado com fósforo radiomarcado. A biblioteca de lambda gt11 foi colocada em placas, a uma densidade de 4 x 10^4 placas por placa de 150 ml. As placas foram cobertas com filtros de nitrocelulose impregnados com IPTG durante 12 horas, e em seguida as placas foram transferidas para nitrocelulose e incubadas com a sonda marcada. As colónias positivas foram então seleccionadas para análise posterior.

A Figura 3 mostra autoradiogramas de fago expressando a proteína GRB-1. A) Primeira imagem demonstrando um sinal positivo (seta) em 40 000 fagos colocados em placa. B) Purificação das placas de fago expressando GRB-1. Todas as placas se ligaram ao domínio C-terminal do EGFR, marcado com [^{32}P].

A Figura 4 mostra a sequência de ADN e a sequência de aminoácidos prevista de GRB-1. A proteína tem 724 resíduos de aminoácidos.

A Figura 5 compara as sequências dos domínios SH2 de GRB-1 com outras proteínas com motivos semelhantes. A) Domínios SH2 de GRB-1, c-src, v-abl, PLC-gama bovina, GAP e V-crk. N e C referem-se, respectivamente, ao terminal N e ao terminal C dos domínios SH2. As substituições de aminoácidos conservativas são as definidas por Schwartz e Dayhoff: (A,G,P,S,T); (L,I,V,M); (D,E,N,Q); (K,R,H); (F,Y,W); e C. As letras a cheio identificam as posições em que o mesmo aminoácido ou uma substituição do aminoácido conservativa está presente em 5 ou mais posições. As caixas identificam motivos conservados. B) Uma comparação semelhante do domínio SH3 de GRB-1.

A Figura 6 é um diagrama esquemático comparando a organização estrutural dos domínios SH2 e SH3. O esquema inclui proteínas conhecidas contendo domínios SH2 e SH3, tais como c-src, V-crk, PLC-gama, GAP1 e GRB-1.

A Figura 7 é uma mancha Northern de ARNm de macaco com sonda para GRB-1. 5 μg de poli(A)+ ARNm, obtido de vários tecidos de macaco, foram submetidos a electroforese em gel de agarose--formaldeído 1,2%/2,2M. A mancha foi hibridada com uma sonda de ADN traduzida por corte com [^{32}P] correspondendo à inserção do clone ki4.

A Figura 8 é um padrão de gel mostrando que anticorpos para GRB-1 imunoprecipitam uma proteína de 85 kDa a partir de células marcadas biossinteticamente. As células foram marcadas metabolicamente com [^{35}S]metionina, e depois prepararam-se lisados e imunoprecipitaram-se com soro imune (I) ou pré-imune (P). A proteína imunoprecipitada foi separada num SDS/PAGE a 8%. A autoradiografia foi executada durante a noite. As linhas celulares usadas incluem linha de células de glioblastoma humano, U1242, linha de células de carcinoma da bexiga de ratazana, células NBT-II e NIH-3T3.

A Figura 9 representa várias proteínas do tipo selvagem e mutantes usadas nos estudos. (A) Construções do receptor EGF com os seus sítios de autofosforilação conhecidos ou previstos. A do

tipo selvagem (W.T.), a quinase negativa (K721A) e a delecção do terminal carboxi (CD126) foram imunoprecipitadas a partir de células NIH-3T3 transfectadas descritas anteriormente expressando -300 000 receptores de EGF. EGFR-C representa um mutante de delecção contendo o domínio citoplásmico do receptor de EGF produzido por células SF9 infectadas com baculovírus. (B) Estrutura de proteínas PLC-gama e trpE/GAP SH2 indicando a localização dos domínios SH2 e SH3 e os sítios de fosforilação da tirosina em PLC-gama.

A Figura 10 é um padrão de gel mostrando a associação de PLC-gama com mutantes de EGFR. O EGFR do tipo selvagem (HER14), de delecção do terminal carboxi (CD126) ou quinase negativa (K721A) foi imunoprecipitado com mAb108 anti-EGFR. Os receptores foram autofosforilados com [gama-^{32}P]-ATP. Concomitantemente adicionou-se EGFR-C a contas de proteína A-Sepharose sózinha ou a receptores K721A imunoprecipitados com ou sem ATP. Depois de mais lavagens para remover ATP, adicionou-se lisado de 15 x 10^6 células 3T-P1 sobreexpressando PLC-gama e misturou-se durante 90 min a 4°C. Após lavagem para remover PLC-gama não ligada, as proteínas foram separadas num gel de SDS a 6% e transferidas para nitrocelulose para imunotransferência ("immunobloting"). Um oitavo da amostra foi utilizado para transferência anti-PTyr, o restante para transferência anti-gama-PLC (tempo de exposição 14 h).

A Figura 11 é um padrão de gel mostrando que a fosforilação de PLC-gama reduz a sua ligação ao receptor de EGF. O EGFR de comprimento total foi imunoprecipitado com mAb108 e deixado autofosforilar. Adicionou-se lisado de células 3T-P1 sobreexpressando PLC-gama e misturou-se durante 90 min a 4°C. Depois da ligação, adicionou-se ATP a metade das amostras deixando-se que as moléculas de PLC-gama fossem fosforiladas pelo receptor de EGF. Adicionou-se então tampão de amostra SDS-PAGE a metade dos complexos EGFR-PLC-gama (SEM LAVAGEM, painel da esquerda) e carregou-se directamente no gel a 6%. A outra metade foi lavada três vezes com HNTG e depois carregada no gel (LAVAGEM, painel da direita). Depois de se passarem amostras em duplicado em SDS-PAGE, as proteínas foram transferidas para

nitrocelulose e sondadas com anti-PLC-gama e [^{125}I]proteína A. As bandas foram subsequentemente retiradas da nitrocelulose e quantificadas num contador gama. Depois de três lavagens com HNTG, 50±5% (Média±SEM, n = 4) da PLC-gama não fosforilada permaneceu ligada ao EGFR enquanto que apenas 22±4% da PLC-gama fosforilada permaneceu ligada (tempo de exposição: 12 h).

A Figura 12 é um padrão de gel mostrando a ligação de EGFR-C a proteínas trpE. (A) O EGFR-C (0,5 μg) foi imunoprecipitado com anticorpo C e lavado. Adicionou-se então $MnCl_2$ sózinho ou $MnCl_2$ e ATP para facilitar a autofosforilação. trpE ou trpE/GAP SH2 (aproximadamente 2 μg). Os imunoprecipitados foram separados num gel SDS a 10%, e transferidos para nitrocelulose e efectuou-se a imunotransferência com anti-trpE. Para comparação, cerca de 0,1 μg de lisado de trpE ou trpE/GAP SH2 foram carregados directamente no gel (painel direito de A). (B) trpE ou trpE/GAP SH2 foi imunoprecipitado com anticorpos anti-trpE e lavado. Adicionou-se então EGFR-C fosforilado e não fosforilado (0,5 μg) e deixou-se ligar como acima. Depois da lavagem, as amostras foram separadas num gel a 10%, transferidas para nitrocelulose e sondadas com anticorpo C. As duas amostras da direita representam 0,5 μg de quinase fosforilada e não fosforilada carregada directamente no gel (tempo de exposição: 2 h).

A Figura 13 é um padrão de gel mostrando a ligação de trpE/GAP SH2 a EGFR do tipo selvagem e mutante. (A) O receptor do tipo selvagem (HER14) ou o receptor de delecção do terminal carboxi CD126 foram imunoprecipitados com mAb 108. Adicionou-se então $MnCl_2$ sózinho ou $MnCl_2$ e ATP à metade autofosforilada das amostras contendo receptor. Um conjunto de CD126 foi também fosforilado cruzadamente com 0,5 μg de EGFR-C. Adicionou-se então trpE/GAP SH2 durante 90 min a 4°C e, após mais três lavagens, carregou-se em SDS-PAGE. Após transferência para nitrocelulose, as manchas foram sondadas com anti-trpE (painel da esquerda), anti-EGFR RK2 (painel central) ou anti-PTyr (painel da direita). O RK2 e o anti-PTyr são ambos 1/8 da amostra total e foram separados em SDS-PAGE a 7%. O resto da amostra foi carregado num gel a 10% para transferência anti-trpE (tempo de exposição:

14 h). (B) Imunoprecipitaram-se lisados de células NIH3T3 2.2 não contendo EGFR (3T3) ou de células com receptores negativos para quinase (K21A) com mAb108. A todos os imunoprecipitados se adicionaram 0,5 μg de EGFR-C e em seguida $MnCl_2$ sózinho ou $MnCl_2$ e ATP. Adicionou-se trpE/GAP SH2 e as amostras foram preparadas e imunotransferidas como em (A) (tempo de exposição: 19 h).

A Figura 14 é um padrão de gel mostrando a ligação de PLC-gama e trpE/GAP SH2 ao fragmento C-terminal, clivado com CNBr, de EGFR. Incubou-se EGFR-C (10 μg) num Centricon 30 em HEPES 20 mM, pH 7,5 com 100 μg de BSA como uma proteína portadora. Os EGFR-C fosforilado e não fosforilado foram então divididos, cada um, em dois, sendo uma metade armazenada em tampão enquanto que a outra metade foi clivada com CNBr. As quatro amostras com ou sem ATP e com ou sem CNBr foram então, cada uma, desenvolvidas em 500 μl de tampão de lise Triton X-100 a 1%, divididas em duas e imunoprecipitadas com anticorpo anti-C. Depois da lavagem dos imunoprecipitados, adicionaram-se os lisados contendo PLC-gama e trpE/GAP SH2. A imunotransferência então realizada nas amostras como acima com anti-trpE ou anti-PLC-gama. Para o painel da direita, carregou-se directamente no gel, sem imunoprecipitação, uma fracção do EGFR-C clivado e não clivado (0,1 μg) e imunotransferida com RK2 (tempo de exposição: 14 h). A banda escura observada em todas as linhas da mancha anti-trpE corre a cerca de 40 kDa (observada também na Figura 13) e representa a ligação de [^{125}I]proteína A à cadeia pesada do anticorpo imunoprecipitante.

A Figura 15 é um padrão de gel mostrando a ligação do fragmento de EGFR C-terminal, fosforilado em tirosina, a trpE/GAP SH2 mas não a trpE. O EGFR-C (5 μg) foi autofosforilado pela adição de [gama-^{32}P]ATP1. O EGFR-C fosforilado foi concentrado num Centricon 30 e em seguida clivado com CNBr em ácido fórmico a 70%. Metade da amostra (350 000 c.p.m.) foi deixada ligar-se a trpE ou a trpE/GAP SH2 tal como na Figura 12B, lavada e passada num gel SDS- a 10%. (A) Ligação de EGFR-C, clivado com CNBr, fosforilado a trpE. (B) Ligação de EGFR-C, clivado com CNBr, fosforilado a trpE/GAP SH2. (C) 3 000 c.p.m. de EGFR-C clivado

com CNBr. (D) para comparação 3 000 c.p.m. de EGFR-C clivado (tempo de exposição: 20 h). EGFR 984/1186 indica a sequência do fragmento autofosforilado em tirosina gerado por CNBr.

A Figura 16 mostra a sequência nucleotídica parcial e a sequência de aminoácidos prevista de GRB-2.

DESCRIÇÃO DETALHADA DAS CONCRETIZAÇÕES PREFERIDAS

Os presentes inventores desenvolveram uma nova sonda e processo para usar esta sonda para clonação por expressão rápida de proteínas que codificam ADN que têm a característica de ligar uma molécula de tirosina-quinase de receptor, em particular o receptor do factor de crescimento epidérmico (EGFR) à cauda C-terminal fosforilada na tirosina.

Com o termo "expressão" entende-se a transcrição e tradução de uma sequência de ADN codificando uma proteína , para originar uma molécula de proteína. A clonação por expressão é um processo em que o ADN a ser clonado codifica uma proteína. O ADN desejado, tipicamente na forma de uma biblioteca de ADNc, é detectado por meio da sua expressão e detecção directa da proteína que codifica. Os sistemas de clonação por expressão são bem conhecidos na arte (ver: Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, 2ª Edição, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)), o qual é aqui incorporado por referência). Tipicamente a proteína é expressa a partir de uma biblioteca tal como uma biblioteca lambda gt11, via infecção de bactérias com o vector de expressão lambda gt11. A proteína expressa, libertada na área da placa bacteriofágica, é transferida para um filtro de nitrocelulose e tipicamente detectada usando um anticorpo para corar o filtro. Estes processos de clonação por expressão convencionais não podem ser usados para novas proteínas ou para proteínas que não estão estão disponíveis em quantidades suficientes para preparar um anticorpo ou para proteínas que não tenham um ligando conhecido para detecção baseada na afinidade.

O processo de clonação por expressão do presente invento pode ser facilmente aplicado a muitos outros sistemas receptores. por exemplo, certas moléculas alvo ligam-se à porção fosforilada na tirosina de PDGFR e ao factor-1 de estimulação de colónias (CSF-1) (Coughlin, S.R. et al., Science 243: 1191-1194 (1989); Kazlauskas, A. et al., Cell 58: 1121-1133 (1989); Shurtleff, S.A. et al., EMBO J. 9: 2415-2421 (1990); Reedjik, M. et al., Mol. Cell. Biol. 10: 5601-5608 (1990)). Nestes dois receptores, a fosforilação da tirosina ocorre num domínio de insercão da quinase, e não no domínio C-terminal como no caso do EGFR. Portanto, para clonação por expressão podem ser usadas similarmente sondas específicas utilizando o domínio de inserção da quinase ou um seu derivado funcional (definido abaixo) tanto do PDGFR como do receptor de CSF-1. Podem também construir-se sondas similares para o receptor do factor de crescimento de fibroblast (FGF) (o qual é fosforilado em tirosina no domínio C-terminal) ou para o receptor HER 2/neu, sendo ambos também capazes de interagir com proteínas contendo SH2 tais como PLC-gama. Noutros receptores, tais como o receptor de insulina, a fosforilação da tirosina ocorre no próprio domínio da quinase.

Assim, embora deva ter-se em consideração que sítios diferentes são fosforilados em tirosina em diferentes proteínas, p. exp., o domínio C-terminal no EGFR, o domínio da quinase no receptor de insulina e um domínio de quinase inserido em PDGFR, o presente invento reconhece as características comuns de todas estas estruturas, a presença de um ou mais resíduos fosfotirosina e a capacidade de certas proteínas celulares se ligarem, com base na afinidade, a um polipéptido contendo uma ou mais fosfotirosinas. Embora abaixo se faça referência na generalidade a uma sonda que é um domínio C-terminal, relativamente ao EGFR, esta linguagem não pretende ser limitativa e pretende incluir todos os outros domínios alternativos fosforilados em tirosina discutidos anteriormente.

Os processos e abordagem do presente invento podem ser aplicados à clonação e identificação de todas as moléculas alvo que são capazes de interactuar de uma maneira específica com os

polipéptidos fosforilados na tirosina, tais como os receptores fosforilados activados descritos acima. Proteínas adicionais que se ligam a sequências fosforiladas na tirosina, tais como as fosfatases específicas para tirosina, p. exp. R-TPases (Sap, J. et al., S Proc. Natl. Acad. Sci. USA 87:6112-6116 (1990); Kaplan, R. et al., S Proc. Natl. Acad. Sci. USA 87:7000-7004 (1990)). Os processos são também aplicáveis na clonação e identificação de proteínas que se ligam a resíduos serina/treonina fosforilados, tal como com fosfatases específicas de serina/treonina.

O uso da sonda do presente invento permite a clonação rápida de ADN e identificação das proteínas codificadas por uma biblioteca de expressão de genes. O processo é particularmente útil com uma biblioteca de bacteriófagos lambda gt11. A pesquisa de uma biblioteca de expressão lambda gt11 de cérebro humano fetal permitiu aos presentes inventores clonar dois genes e caracterizar as proteínas que eles codificam. Um, denominado GRB-1, foi sequenciado na totalidade e verificou-se que codificava nova proteína humana com um peso molecular de cerca de 85 kDa, a qual continha dois domínios SH2 e um domínio Sh3 (Figura 4 e Figura 5). A GRB-2, que foi sequenciada parcialmente, continha também domínios únicos SH2 e SH3.

Os domínios SH2, tal como nas proteínas GAP e PLG-gama, são responsáveis pela assocaição destas proteínas com o término C fosforilado do EGFR (ver Exemplo VI, abaixo). Assim, uma das funções dos domínios SH2 é a de justapor a porção intracelular das moléculas de tirosina-quinase de receptor com os seus substratos para facilitar a fosforilação eficiente da tirosina.

A análise detalhada de um dos clones de ADNc do presente invento, GRB-1, identificado usando os processos do presente invento, revela uma sequência nova contendo dois domínios SH2 e um domínio SH3. Esta proteína é expressa em vários tecidos e linhas celulares. O seu peso molecular previsto, 85 kDa, é consistente com a sua migração em electroforese de dodecilsulfato de sódio-gel de poliacrilamida (SDS-PAGE).

O termo tirosina-quinase de receptor pretende designar uma proteína transmembranar possuindo um domínio de receptor extracelular e um ou mais domínios intracelulares, incluindo um domínio possuindo a actividade enzimática da tirosina-quinase. Os domínios intracelulares adicionais podem ter homologia de sequência com SH2. Estas moléculas são bem conhecidas na arte (Williams, L.T.et al., Science 243:1564-1570 (1989); Ullrich, A. et al., Cell 61:203-212 (1990); Carpenter, G. et al., J. Biol. Chem: 265:7709-7712 (1990)).

As proteínas que interagem com, e que podem ser fosforiladas pelas, tirosina-quinase de receptor são designadas por proteínas "alvo" para estes receptores , para distinguir dos "ligandos" destes receptores, os quais se ligam ao domínio extracelular do receptor.

De acordo com o presente invento, o processo de clonação por expressão é realizado directamente numa biblioteca de expressão de genes, tal como de lambda gt11. Numa concretização preferida, o ADN é ADNc humano. Mais preferivelmente, o ADN é ADN de cérebro humano fetal. A utilização desta fonte como material de partida para a clonação de genes humanos é de grande vantagem relativamente aos meios alternativos nos quais se retira uma grande quantidade de tecido e se produzem muitos anticorpos ou a proteína é purificada e parcialmente sequenciada e as sondas oligonucleotídicas são preparadas a partir desta sequência e usadas para analisar um ADN genómico ou uma biblioteca de ADNc. A vantagem de ultrapassar estes passos é da maior importância no caso dos genes humanos visto que o tecido não está, geralmente, disponível em grandes quantidades, com a excepção da placenta.

A biblioteca de expressão é pesquisada num só passo. Preferivelmente, as placas lambda são "transferidas" para um portador sólido, preferivelmente, nitrocelulose, permitindo a transferência das proteínas codificadas pelo ADN da biblioteca que são expressas nas bactérias infectadas e transferidas para o portador. Este portador é então incubado com a sonda do presente invento, como aqui descrito. A sonda é deixada ligar-se a

proteínas que têm a capacidade de se ligar ao polipéptido fosforilado na tirosina. Com base no marcador usado na sonda , preferivelmente o radioisótopo ^{32}P, utiliza-se um sistema de detecção apropriado para identificar as placas contendo a proteína de interesse. Os fagos nestas placas são então seleccionados e as inserções de ADN neles podem ser novamente clonadas, excisadas e colocadas em outros vectores, usados para expressão da proteína em larga escala, e semelhantes, como é bem conhecido na arte.

Um perito da arte vulgar terá em conta que as concentrações, tempos e temperaturas podem variar com a natureza precisa do sistema usado e saberão fazer variar os parâmetros sem experimentação indevida. Além disso, os processos gerais nesta área são apresentados em Sambrook et al. (supra).

Os materiais de que o portador de fase sólida pode ser feito incluem, não restritivamente, nitrocelulose, celulose, papel, poliestirenos substituídos, acrilonitrilos, policarbonato, polipenteno ou óxido de silicone.

A sonda do presente invento é uma molécula polipeptídica fosforilada na tirosina derivada do domínio C-terminal de uma tirosina-quinase de receptor, preferivelmente o EGFR. O polipéptido pode ter um comprimento de entre cerca de 25 e cerca de 250 aminoácidos. A sonda pode ser uma sequência nativa fosforilada ou um seu derivado funcional (definido abaixo). Obtém-se uma fosforilação muitíssimo eficiente usando o domínio da tirosina-quinase presente na molécula de receptor para autofosforilar a região C-terminal entre 1 e 5 resíduos de tirosina. Usam-se processos e condições conhecidas (descritas em detalhe no Exemplo I) para fosforilar os resíduos de tirosina. Um substrato preferido é [gama-^{32}P-adenosina-trifosfato]. A fonte de molécula de receptor usada como material fonte para fazer a sonda pode incluir moléculas purificadas quimicamente a partir de tecidos ou células ou moléculas produzidas por métodos de ADN recombinante.

Quando se usam técnicas recombinantes, pode ser produzido um

receptor nativo ou, alternativamente, pode ser produzido um derivado do tipo receptor. Um derivado do tipo receptor inclui o domínio da tirosina-quinase ligado ao domínio C-terminal. Noutra concretização, os dois domínios podem ser produzidos como moléculas separadas e misturadas entre si para conseguir a fosforilação na tirosina do polipéptido derivado do término C.

A sonda compreendendo uma porção C-terminal, fosforilada na tirosina, da tirosina-quinase de receptor, como aqui descrita, pode ser produzida por meios recombinantes na forma de uma proteína de fusão.

Como aqui usada, a expressão "proteína de fusão" pode referir-se a uma proteína fundida compreendendo uma proteína bacteriana e um polipéptido de interesse, tal como uma proteína possuindo um domínio SH2. Alternativamente, uma proteína de fusão também pode ser um derivado do tipo tirosina-quinase de receptor construído artificialmente, no qual uma sequência de ADN codificando o domínio da tirosina-quinase foi ligada a um sítio de clivagem enzimática selectiva, que, por sua vez, está ligado a um domínio C-terminal da tirosina-quinase de receptor possuindo um ou mais resíduos tirosina que podem ser fosforilados pela quinase. Uma tal construção genética que codifique este tipo de "proteína de fusão" pode ser inserida num veículo de expressão e expressada num hospedeiro bacteriano ou eucariótico. Uma vez expressa, esta proteína de fusão pode ser deixada autofosforilar, agindo a quinase de modo a fosforilar os resíduos tirosina do domínio C-terminal. A seguir a esta fosforilação, a utilização da enzima apropriada clivará no sítio de clivagem selectivo, separando assim a quinase N-terminbal do polipéptido fosforilado C-terminal, o qual pode agora servir como sonda.

Itakura et al., (Science 198:1056-1063 (1977)) e Riggs (Patente U.S. 4 366 246 (1982)) ensina métodos para expressão bacteriana de uma proteína estranha, na forma de uma proteína de fusão, que evitam a degradação intracelular da proteína estranha. Além disso, as proteínas de fusão sobreexpressas são frequentemente depositadas como corpos de inclusão nas células

bacterianas e são, portanto, mais fáceis de isolar e purificar (Marston, Biochem. J. 240:1-12 (1986)). Estas referências sugeriram métodos de clivagem alternativos baseados no conceito de sítios de clivagem selectivos: tal como a metionina pode ser inserida na junção desejada para servir como sítio alvo para o brometo de cianogénio, também podiam ser introduzidos sítios específicos para aminoácidos que podem ser atacados por enzimas proteolíticas com especificidade para aminoácido. Nagai et al., (Nature 309:810-812 (1984) revelam a produção de um produto genético desejado numa proteína de fusão, pela introdução de uma sequência de 4 aminoácidos que serve como sítio de clivagem para o factor Xa de coagulação do sangue. A clivagem pelo factor Xa também eliminou a metionina N-terminal extra, codificada pelo codão de iniciação que está presente na maioria das proteínas eucarióticas expressas em E. coli por métodos convencionais. Outras referências descrevem o uso de sítios de clivagem enzimática reconhecidos pelo factor Xa ou pela trombina, para aumentar a expressão em bactérias e para tornar mais fácil a purificação da proteína desejada (Germino et al., Proc. Natl. Acad. Sci. USA 81:692-4696 (1984); Scholtissek et al., Gene 62:55-64 (1988); Smith et al., Gene 67:31-40 (1988); Knott et al., Eur. J. Biochem. 174:405-410 (1988); e Dykes et al., Eur. J. Biochem. 174:411-416 (1988)).

O termo "sítio de clivagem selectivo" refere-se a um resíduo ou resíduos de aminoácido que podem ser selectivamente clivados quer com produtos químicos quer com enzimas e quando a clivagem pode ser conseguida de uma forma predizível. O sítio de clivagem enzimática selectivo é um aminoácido ou uma sequência peptídica que é reconhecida e hidrolisada por uma enzima proteolítica. Exemplos de tais sítios incluem sítios de clivagem de tripsina ou de quimiotripsina. Numa concretização preferida deste invento, o sítio de clivagem selectivo compreende a sequência Ile-Glu-Gly-Arg, que é reconhecida e clivada pelo factor Xa de coagulação do sangue. Noutra concretização, o sítio de clivagem selectivo tem a sequência Leu-Val-Pro-Arg, que é reconhecida e clivada pela trombina.

Na construção do derivado do tipo receptor pode incluir-se uma sequência de oligonucleótidos, 5' em relação à sequência que codifica o sítio de reconhecimente da enzima, e esta pode variar no comprimento. Por exemplo, numa concretização, 13 nucleótidos estão situados entre o codão para Ile (o início do sítio de reconhecimento do factor Xa) e a extremidade 3' da sequência que codifica o domínio da tirosina-quinase.

Assim, numa concretização do presente invento, a sequência Ile-Glu-Gly-Arg é introduzida entre o domínio da tirosina-quinase e o domínio C-terminal. Noutra concretização, é introduzida a sequência Leu-Val-Pro-Arg. As proteínas possuindo este sítio de clivagem são expressas em bactérias usando métodos standard. Em seguida, deixa-se ocorrer a autofosforilação do domínio C-terminal, preferivelmente com [gama^{32}P]adenosina-trifosfato, seguida de clivagem selectiva do domínio C-terminal fosforilado na tirosina, com o agente de clivagem apropriado, p.exp. factor Xa.

O presente invento também proporciona um processo para mapear um gene, preferivelmente um gene humano, que codifica uma proteína alvo para uma tirosina-quinase de receptor (tal como uma proteína GRB como aqui definida), num cromossoma humano particular. Este processo combina o novo processo de clonação por expressão aqui descrito com uma de várias técnicas conhecidas para mapear um gene num cromossoma particular. Assim, de acordo com o presente invento, usando os processos de clonação por expressão aqui divulagados é identificado um clone, tal como um clone lambda gt11, contendo uma inserção de ADN que codifica uma proteína GRB. A inserção pode ainda ser subclonada, se desejado, usando métodos bem conhecidos na arte, e construir-se uma sonda, quer marcando directamente o ácido nucleico do clone quer produzindo uma sonda de oligonucleótidos corrrespondendo a uma porção única da sequência do clone (ver:Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, 2ª Edição, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)) Esta sonda marcada pode então ser usada num ensaio de hibridação com "transferências" disponíveis comercialmente, tal como Chromossoms Blots da

Bios Corporation (New Haven, Connecticut) que contém ADN de um painel de híbridos de células somáticas de ser humano-hamster (Kouri, R. E. et al., Cytogenet. Cell Genet. 51:1025 (1989)). Confrontando quais os cromossomas humanos que permanecem na célula híbrida de ser humano-hamster e pela hibridação da sonda específica para o gene GRB de interesse, o gene é mapeado num cromossoma humano particular. Deste modo é estabelecida a ligação a genes humanos conhecidos (ou a doenças causadas por mutações neles) presentes neste cromossoma. Usando métodos bem conhecidos na arte para mapear acuradamente, p. exp., usando mutações de delecção humanas conhecidas, o gene GRB pode ser mapeado mais precisamente para outros genes humanos.

O polipéptido sonda C-terminal da tirosina-quinase de receptor fosforilado na tirosina, do presente invento, bem como as proteínas GRB do presente invento e proteínas GRB adicionais ainda desconhecidas que são obtidas usando os processos deste invento, são úteis em métodos de pesquisa de drogas e de outros agentes que sejam capazes de modular o controlo do crescimento celular que ocorre via transdução de sinal através de tirosina-quinases de receptor. Ligando um polipéptido sonda fosforilado na tirosina ou uma proteína GRB, ou seus fragmentos, a uma matriz portadora em fase sólida, cria-se uma sonda de afinidade que pode ser usada para isolar e purificar moléculas a partir de misturas complexas que são capazes de se ligar à sonda de afinidade. Além disso, uma tal sonda de afinidade é útil para detectar a presença, num fluido biológico, de uma molécula capaz de se ligar à sonda fosforilada em tirosina ou à proteína GRB. Similarmente, podem testar-se agentes químicos quanto à sua capacidade de interactuar com a sonda ou com GRB.

Os métodos de acoplamento de proteínas e péptidos à fase sólida, as substâncias em fase sólida úteis nestes métodos e os meios para eluição são bem conhecidos dos peritos na arte.

No caso dos receptores dos factores de crescimento que são tirosina-quinases de receptor, incluindo EDGFR, PDGFR e FGFR, a fosforilação na tirosina está ligada ao crescimento celular e à

transformação oncogénica. A ruptura da acção de uma GRB na célula pode evitar ou inibir o crescimento e pode servir como meio para neutralizar o desenvolvimento de um tumor. Além disso, uma mutação na porção C-terminal da tirosina-quinase de receptor ou na GRB, ou uma desregulação nas suas interacções mútuas, pode promover a susceptibilidade ao cancro.

O receptor de insulina (InsR) é também uma tirosina-quinase de receptor e a fosforilação na tirosina nas células que têm InsR está associada à função fisiológica normal. Pelo contrário, no caso do crescimento celular e do cancro, a ruptura das interacções normais entre a porção fosforilada na tirosina do receptor e a GRB neutralizaria os efeitos da insulina. Os níveis ou a actividade subnormal de uma proteína GRB podem actuar no sentido de remover os mecanismos normais contrarreguladores. Talvez a sobreexpressão ou a sobreactividade de uma proteína GRB possam inibir ou evitar completamente a acção da insulina sobre as células, conduzindo à diabetes (de uma variedade resistente à insulina). Assim, a susceptibilidade a diabetes pode ser associada à desregulação da proteína GRB-1 ou GRB-2.

Deste modo, os processos do presente invento para identificar genes de proteínas GRB normais ou mutantes, ou para detectar a presença ou a quantidade de GRB-1 ou de GRB-2 numa célula podem servir como processos para identificar susceptibilidade ao cancro, diabetes ou outras doenças associadas a alterações no metabolismo celular mediadas pelas vias da tirosina-quinase de receptor.

O presente invento proporcina processos para investigar a presença e o nível de proteína GRB-1 ou GRB-2 normal ou mutante , num paciente. A expressão alterada destas proteínas ou a presença de uma proteína GRB mutante, num indivíduo pode servir como um importante preditor da susceptibilidade à transformação oncogénica e ao desenvolvimento do cancro. Alternativamente, a expressão alterada da proteína GRB pode servir como um importante preditor da susceptibilidade à diabetes.

As sondas oligonucleotídicas que codificam várias porções da proteína GRB são usadas para testar células de um paciente quanto à presença de sequências de ADN ou de ARN que codificam a proteína GRB. Uma sonda preferida seria uma sonda dirigida à sequência de ácido nucleico que codifica pelo menos 4 resíduos de aminoácido, e preferivelmente pelo menos 5 resíduos de aminoácido da proteína GRB-1 ou GRB-2 do presente invento. Usando estas sondas podem ser efectuados ensaios qualitativos e quantitativos. Por exemplo, a análise de Northern (ver Exemplo III, abaixo) é usada para medir a expressão de um ARNm para proteína GRB numa preparação de célula ou de tecido.

Estes processos podem ser usados mesmo com quantidades muito pequenas de ADN obtido de um indivíduo, utilizando as técnicas de amplificação selectiva. As metodologias de ADN recombinante capazes de amplificar fragmentos de ácido nucleico purificados desde há muíto que foram reconhecidas. Tipicamente, estas metodologias envolvem a introdução do fragmento de ácido nucleico num vector de ADN ou de ARN, a amplificação clonal do vector e a recuperação do fragmento de ácido nucleico amplificado. Exemplos destas metodologias são proporcionados por Cohen et al. (Patente U.S. 4 237 224), Sambrook et al., (supra), etc.

Recentemente foi descrito um método enzimático, in vitro, que é capaz de aumentar a concentração dessas moléculas de ácido nucleico desejadas. Este método tem sido referido como "reacção em cadeia de polimerase" ou "PCR" (Mullis, K. et al., Cold Spring Harbor Symp. Quant. Biol. 51:263-273 (1986); Erlich H. et al., EP 50 424; EP 84 796, EP 258 017, EP 237 362; Mullis, K., EP 201 184; Mullis, K. et al., US 4 683 202; Erlich H., US 4 582 788; e Saiki, R. et al., US 4 683 194).

A reacção em cadeia de polimerase proporciona um método para aumentar selectivamente a concentração de uma sequência de ácido nucleico particular mesmo no caso dessa sequência não ter sido previamente purificada e está presente apenas como uma cópia única numa amostra particular. O método pode ser usado para amplificar ADN de cadeia simples ou de cadeia dupla. A essência

do método envolve o uso de duas sondas de oligonucleótidos para servirem como iniciadores para a replicação mediada pela polimerase, dependente do molde, de uma molécula de ácido nucleico desejada.

A natureza precisa das duas sondas de oligonucleótidos do método PCR é crítica para o sucesso do método. Como é conhecido, uma molécula de ADN ou de ARN possui direccionalidade, que é conferida pela ligação 5'-3' dos grupos fosfato da molécula. As sequências de ADN ou de ARN são ligadas entre si através da formação de uma ligação fosfodiéster entre o grupo fosfato 5' terminal de uma sequência e o grupo hidroxilo 3' terminal de uma segunda sequência. A amplificação dependente de polimerase de uma molécula de ácido nucleico prossegue pela adição de um 5' nucleótido trifosfato à extremidade 3' hidroxilo de uma molécula de ácido nucleico. Assim, a acção de uma polimerase estende a extremidade 3' de uma molécula de ácido nucleico. Estas propriedades inerentes são exploradas na selecção das sondas de oligonucleótidos da PCR. As sequências oligonucleotídicas das sondas do método PCR são seleccionadas de modo a conterem sequências idênticas ou complementares a sequências que flanqueiam a sequência de ácido nucleico particular cuja amplificação é desejada.

Mais especificamente, a sequência de oligonucleótidos da "primeira" sonda é seleccionada de modo a ser capaz de se hibridar com uma sequência de oligonucleótidos localizada a 3' da desejada sequência, enquanto que a sequência de oligonucleótidos da "segunda" sonda é seleccionada de modo a conter uma sequência de oligonucleótidos idêntica à presente a 5' da região desejada. Ambas as sondas possuem grupos 3' hidroxi e podem portanto funcionar como iniciadores para a síntese de ácido nucleico.

Na PCR, as condições da reacção são alternadas entre as que conduzem à hibridação e à polimerização do ácido nucleico e as que resultam na desnaturação de moléculas duplex. No primeiro passo da reacção, os ácidos nucleicos da amostra são aquecidos transientemente e em seguida arrefecidos de modo a desnaturar

quaisquer moléculas de cadeia dupla que possam estar presentes. As "primeira" e "segunda" sondas são então adicionadas à amostra a uma concentração que excede bastante a da molécula de ácido nucleico desejada. Quando a amostra é incubada sob condições conducentes à hibridação e polimerização, a "primeira" sonda hibridar-se-à com a molécula de ácido nucleico da amostra numa posição 3' em relação à sequência a amplificar. Se a molécula de ácido nucleico da amostra era, inicialmente, de cadeia dupla, a "segunda" sonda hibridar-se-á com a cadeia complementar da molécula de ácido nucleico numa posição 3' em relação à sequência que é o complemento da sequência cuja amplificação é desejada. Depois da adição de uma polimerase, as extremidades 3' da "primeira" e (se a molécula de ácido nucleico era de cadeia dupla) da "segunda" sondas serão prolongadas. A extensão da "primeira" sonda terá como resultado a síntese de um oligonucleótido possuindo a sequência exacta do ácido nucleico desejado. A extensão da "segunda" sonda terá como resultado a síntese de um oligonucleótido possuindo a sequência exacta do complemento do ácido nucleico desejado.

A reacção PCR é capaz de amplificar exponencialmente sequências de ácidos nucleicos específicas porque o produto de extensão da "primeira" sonda, necessária, contém uma sequência que é complementar a uma sequência da "segunda" sonda e que serve portanto como um molde para a produção de um produto de extensão da "segunda" sonda. Similarmente, o produto de extensão da "segunda" sonda, necessária, contém uma sequência que é complementar a uma sequência da "primeira" sonda e que serve portanto como um molde para a produção de um produto de extensão da "primeira" sonda. Assim, ao permitirem-se ciclos de polimerização e de desnaturação pode conseguir-se um aumento geométrico na concentração da molécula de ácido nucleico desejada. Revisões sobre PCR são proporcionadas por Mullis, K.B. (Cold Spring Harbor Symp. Quant. Biol. 51:263-273 (1986)); Saiki, R.K., et al., (Bio/Technology 3:1008-1012 (1985)); e Mullis, K.B. et al., (Meth. Enzymol. 155:335-350 (1987)).

Numa concretização, o invento dirige-se a proteínas GRB-1 e

GRB-2 de ocorrência natural. Noutra concretização, o invento dirige-se a proteínas GRB-1 e GRB-2 recombinantes. O invento proporciona a molécula proteica de ocorrência natural substancialmente isenta de outras proteínas com as quais está nativamente associada. "Substancialmente isenta de outras proteínas ou glicoproteínas" indica que a proteína foi purificada pelo menos até 90% (numa base ponderal) e mesmo até pelo menos 99%, se desejado, das outras proteínas e glicoproteínas com as quais está nativamente associada e que está portanto essencialmente isenta delas. Isto pode conseguir-se submetendo as células , tecido ou fluidos contendo a proteína GRB-1 ou GRB-2 a técnicas standard de purificação de proteína tal como colunas de imunoadsorvente possuindo anticorpos reactivos monoclonais contra a proteína.

A sequência de nucleótidos do gene de GRB-1 e a sequência de aminoácidos da proteína GRB-1 são apresentadas na Figura 4. A sequência de nucleótidos parcial de GRB-2 e a sequência de aminoácidos parcial são apresentadas na Figura 16.

Numa concretização preferida, a GRB-1 ou a GRB-2, ou uma proteína GRB desconhecida, podem ser isoladas e purificadas usando uma sonda de afinidade, a sonda do presente invento que é um domínio C-terminal fosforilado em tirosina de uma tirosina-quinase de receptor, ou um seu derivado funcional.

Alternativamente, a purificação pode ser conseguida por uma combinação de métodos atandard tais como precipitação com sulfato de amónio, cromatografia com peneiro molecular e cromatografia de permuta iónica.

Deverá entender-se que as proteínas GRB-1 e GRB-2 do presente invento podem ser purificadas bioquimicamente a partir de uma variedade de fontes celulares e tissulares. Para a preparação de proteína GRB de ocorrência natural, preferem-se os tecidos tais como placenta ou cérebro de mamífero.

Alternativamente, em virtude do gene para GRB-1 e GRB-2 pode ser isolado ou sintetizado, o polipéptido pode ser sintetizado

substancialmente isento de outras proteínas ou glicoproteínas de origem de mamífero num organismo procariótico ou num organismo eucariótico não mamífero, se desejado. Como se entende no presente invento, uma molécula de GRB-1 ou GRB-2 produzida em células de mamífero, tais como, por exemplo, células COS transfectadas, NIH-3T3 ou CHO, é quer uma sequência proteica de ocorrência natural quer um seu derivado funcional. Quando é produzida uma proteína ou glicoproteína de ocorrência natural por meios recombinantes, ela é proporcionada substancialmente isenta de outras proteínas e glicoproteínas às quais está nativamente associada.

Alternativamente, são conhecidos métodos para a síntese de polipéptidos com uma desejada sequência em suportes de fase sólida e a sua subsequente separação do suporte. Em particular, a sonda do domínio C-terminal fosforilado em tirosina do presente invento, ou um seu derivado funcional, pode ser sintetizada usando um método de síntese peptídica no qual é proporcionada fosfotirosina no lugar de tirosina, resultando na síntese directa da forma fosforilada do polipéptido.

O presente invento proporciona "derivados funcionais" do polipéptido do domínio C-terminal fosforilado em tirosina e/ou as proteínas GRB-1 e GRB-2.

Por "derivado funcional" entende-se um "fragmento", uma "variante", um "análogo" ou um "derivado químico" da proteína GRB, termos que são definidos abaixo. Um derivado funcional retém pelo menos uma porção da função da proteína nativa o que permite a sua utilização de acordo com o presente invento.

Um "fragmento" de qualquer das proteínas ou polipéptidos do presente invento refere-se a qualquer subconjunto da molécula, isto é, a um péptido mais curto.

Uma "variante" da proteína refere-se a uma molécula substancialmente similar ao péptido inteiro ou ao seu fragmento. Os péptidos variantes podem ser convenientemente preparados por

síntese química directa do péptido variante usando métodos conhecidos na arte.

Alternativamente, as variantes da sequência de aminoácidos do péptido podem ser preparadas por mutações no ADN que codifica o péptido sintetizado. Estas variantes incluem, por exemplo, delecções, inserções ou substituições de resíduos na sequência de aminoácidos. Qualquer combinação de delecção, inserção e substituição também pode ser feita chegar à construção final,desde que a construção final possua a actividade desejada. Obviamente, as mutações que serão feitas no ADN que codifica o péptido variante não podem alterar a estrutura de leitura e preferivelmente não criarão regiões complementares que poderiam produzir a estrutura de ARNm secundária (ver Publicação da Patente Europeia nº EP 75 444).

Ao nível genético, estas variantes são usualmente preparadas por mutagénese dirigida ao sítio (como exemplificado por Adelman et al., DNA 2:183 (1983)) de nucleótidos no ADN que codifica a molécula de péptido, produzindo-se assim um ADN que codifica a variante, e em seguida expressando o ADN na cultura de células recombinantes (ver abaixo). As variantes exibem tipicamente a mesma actividade biológica qualitativa do péptido não variante.

Um "análogo" do polipéptido fosforilado em tirosina ou da proteína GRB refere-se a uma molécula não natural substancialmente similar à molécula inteira ou a um seu fragmento.

Um "derivado químico" do polipéptido fosforilado em tirosina ou da proteína GRB contém "porções" ("moieties") químicas adicionais que normalmente não fazem parte deste péptido. As modificações covalentes do péptido estão incluídas no âmbito deste invento. Estas modificações podem ser introduzidas na molécula fazendo reagir resíduos de aminoácidos alvejados do péptido com um agente de derivação orgânico que é capaz de reagir com cadeias laterais ou com resíduos terminais seleccionados.

Habitualmente fazem-se reagir resíduos cisteínilo com

alfa-haloacetatos (e correspondentes aminas), tais como ácido cloroacético ou cloroacetamida, para dar derivados carboximetilo ou carboxiamidometilo. Os resíduos cisteínilo dão também derivados por reacção com bromotrifluoroacetona, ácido alfa-bromo-beta-(5-imidozoil)propiónico, fosfato de cloroacetilo, N-alquilmaleimidas, dissulfeto de 3-nitro-2-piridilo, 2-piridil-dissulfeto de metilo, p-cloromercuribenzoato, 2-cloromercuri-4-nitrofenol ou cloro-7-nitrobenzo-2-oxa-1,3-diazol.

Os resíduos histidilo dão derivados por reacção com dietilprocarbonato a pH 5,5-7,0 porque este agente é relativamente específico para a cadeia lateral de histidilo. O brometo de para-bromofenacilo também é útil; a reacção é preferivelmente executada em cacodilato de sódio 0,1 M a pH 6,0.

Os resíduos lisinilo e amino terminal são feitos reagir com anidridos succínico ou de outros ácidos carboxílicos. A produção de derivados a partir destes agentes tem o efeito de inverter a carga dos resíduos lisinilo. Outros reagentes adequados para obter derivados a partir de resíduos contendo alfa-amino incluem imidoésteres tais como picolinimidato de metilo; fosfato de piridoxal; piridoxal; cloroboro-hidreto; ácido trinitrobenzenos-sulfónico; O-metiliso-ureia; 2,4-pentanodiona; e reacção, catalisada por transaminase, com glioxilato.

Os resíduos arginilo são modificados por reacção com um ou vários reagentes convencionais, de entre eles com fenilglioxal, 2,3-butanodiona, 1,2-ciclo-hexanodiona e ninidrina. A obtenção de derivados a partir de resíduos arginina requer que a reacção seja executada em condições alcalinas em virtude do pK_a elevado do grupo funcional guanidina. Além disso, estes reagentes podem reagir com os grupos da lisina bem como com o grupo epsilon-amino da arginina.

A modificação específica dos resíduos tirosilo per se tem sido estudada extensivamente, com particular interesse na introdução de marcadores espectrais nos resíduos tirosilo por reacção com compostos de diazónio aromáticos ou tetranitrometano.

Habitualmente, usam-se N-acetilimidizol e tetranitrometano para formar espécies O-acetil tirosil e derivados 3-nitro, respectivamente.

Os grupos carboxilo laterais (aspartilo ou glutamilo) são modificados selectivamente por reacção com carbodiimidas (R'-N-C-N-R') tais como 1-ciclo-hexil-3-(2-morfolinil-(4-etil)carbodiimida ou 1-etil-3-(4-azonia-4,4-dimetilpentil)carbodiimida. Além disso, os resíduos aspartilo e glutamilo são convertidos em resíduos asparaginilo e glutaminilo por reacção com iões amónio.

Os resíduos glutaminilo e asparaginilo são frequentemente desamidados dando os correspondentes resíduos glutamilo e aspartilo. Alternativamente, estes resíduos são desamidados sob condições moderadamente ácidas. Qualquer das formas destes resíduos cai no âmbito deste invento.

A obtenção de derivados com agentes bifuncionais é útil para ligar cruzadamente o péptido a uma matriz suporte insolúvel em água ou a outros portadores macromoleculares. Os agentes de ligação cruzada comummente usados incluem, p. exp., 1,1-bis(diazoacetil)-2-feniletano, glutaraldeído, ésteres de N-hidroxisuccinimida, por exemplo, ésteres com ácido 4-azido-salicílico, imidoésteres homobifuncionais, incluindo ésteres di-succinimidílicos tais como 3,3'-ditiobis(succinimidilpropionato) e maleimidas bifuncionais tais como bis-N-maleimido-1,8-octano. Os agentes de derivação tais como metil-3-[(p-azidofenil)ditio]propioimidato originam intermediários fotoactiváveis que são capazes de formar ligações cruzadas na presença de luz. Alternativamente, as matrizes reactivas, insolúveis em água, tais como hidratos de carbono activados com brometo de cianogénio e os substratos reactivos descritos nas Patentes U.S. nº 3 969 287; 3 691 016; 4 195 128; 4 247 642; 4 229 537; e 4 330 440, são empregues para a imobilização das proteínas.

Outras modificações incluem a hidroxilação da prolina e da lisina, a fosforilação dos grupos hidroxilo dos resíduos serilo

ou treonilo, a metilação dos grupos alfa-amino das cadeias laterais da lisina, arginina e histidina (T.E. Creighton, Proteins: Structure and Molecule Properties, W.H. freeman & Co., San Francisco, pag. 79-86 (1983)), a acetilação da amina N-terminal e, em alguns casos, a amidação dos grupos carboxilo terminais.

Estas "porções" derivadas podem melhorar a solubilidade, absorção, meia-vida biológica e similares. As "porções" podem, alternativamente, eliminar ou atenuar qualquer efeito lateral indesejável da proteína ou similar. As "porções" capazes de mediar esses efeitos são divulgadas por exemplo, em Remington's Pharmaceutical Sciences, 16ª ed., Mack Publishing Co., Easton, PA (1980).

O hospedeiro bacteriano preferido para este invento é E. coli. Noutras concretizações podem ser usadas outras espécies bacterianas. Ainda noutras concretizações, podem ser usadas células eucarióticas tais como, por exemplo, levedura, fungos filamentosos ou semelhantes. O uso destes tipos de células é bem conhecido da arte. Para expressar a proteína pode usar-se qualquer hospedeiro que seja compatível com o replicão e com as sequências de controlo no plasmídeo de expressão. Em geral, os vectores contendo replicão e as sequências de controlo derivam de espécies compatíveis com uma célula hospedeira são usados para o hospedeiro. O vector contém normalmente um sítio de replicação, bem como genes específicos que são capazes de proporcionar selecção fenotípica em células infectadas ou em células transformadas. A expressão da proteína de fusão também pode ser colocada sob controlo com outras sequências reguladoras que podem ser homólogas ao organismo na sua forma não transformada.

Este invento dirige-se também a um anticorpo específico para um epítopo da proteína GRB-1 ou GRB-2 e ao uso de um tal anticorpo para detectar a presença ou medir a quantidade ou concentração da proteína GRB numa célula, um extracto de célula ou de tecido ou num fluido biológico.

O termo "anticorpo " pretende-se que inclua anticorpos policlonais, anticorpos monoclonais (mAbs), anticorpos quiméricos e anticorpos anti-idiotípicos (anti-Id).

Os anticorpos policlonais são populações heterogéneas de moléculas de anticorpo derivadas dos soros de animais imunizados com um antigénio.

Os anticorpos monoclonais são uma população substancialmente homogénea de anticorpos para antigénios específicos. Os mAbs podem ser obtidos por métodos bem conhecidos dos peritos na arte. Ver, por exemplo Kohler e Milstein, Nature 256:495-497 (1975) e Patente U.S. nº 4 376 110. Estes anticorpos podem ser de qualquer classe das imunoglobulinas incluindo IgG, IgM, IgE, IgA, GILD e qualquer sua subclasse. O hibridoma que produz os mAbs deste invento pode ser cultivado in vitro ou in vivo. A produção de títulos elevados de produção de mAbs in vivo torna este método o método de produção presentemente preferido. Resumidamente, as células dos hibridomas individuais são injectadas intraperitonealmente em ratinhos BALB/c inoculados primariamente com pristano para produzir fluido ascítico contendo elevadas concentrações dos mAbs desejados. Os mAbs de isotipo IgM ou IgG podem ser purificados a partir desses fluidos ascíticos ou dos sobrenadantes da cultura, usando métodos cromatográficos bem conhecidos dos peritos na arte.

Os anticorpos quiméricos são moléculas com diferentes porções derivadas de diferentes espécies animais, tais como as que possuem a região variável derivada de um mAb murino e uma região constante de imunoglobulina humana. Os anticorpos quiméricos e os métodos para a sua produção são conhecidos na arte (Cabilly et al., Proc. Natl. Acad. Sci. USA 81:3273-3277 (1984); Morrison et al., Proc. Natl. Acad. Sci. USA 81:6851-6855 (1984); Boulianne et al., Nature 312:643-646 (1984); Cabilly et al., Pedido de Patente Europeia 125023 (publicado a 14 de Novembro de 1984); Neuberger et al., Nature 314:268-270 (1985); Taniguchi et al., Pedido de Patente Europeia 171496 (publicado a 19 de Fevereiro de 1985); Morrison et al., Pedido de Patente

Europeia 173494 (publicado a 5 de Março de 1986); Neuberger et al. Pedido PCT WO 86/01533, (publicado a 13 de Março de 1986); Kudo et al., Pedido de Patente Europeia 184187 (publicado a 11 de Junho de 1986); Sahagan et al., J. Immunol. 137:1066-1074 (1986); Robinson et al., Publicação de patente Internacional #PCT/US86/02269 (publicada a 7 de Maio de 1987); Liu et al., Proc. Natl. Açad. Sci. USA 84:3439-3443 (1987); Sun et al., Proc. Natl. Acad. Sci. USA 84:214-218 (1987); Better et al., Science 240:1041-1043 (1988)). Estas referências são aqui incorporadas por referência.

Um anticorpo anti-idiotípico (anti-Id) é um anticorpo que reconhece determinantes únicas geralmente associadas ao sítio de ligação ao antigénio de um anticorpo. Um anticorpo Id pode ser preparado por imunização de um animal da mesma espécie e tipo genético (p. exp. uma estirpe de ratinho) da fonte do mAb, com o mAb para o qual está a ser preparado um anti-Id. o animal imunizado reconhecerá e responderá às determinantes idiotípicas do anticorpo imunizante, produzindo um anticorpo para estas determinantes idiotípicas (o anticorpo anti-Id).

O anticorpo anti-Id também pode ser usado como um "imunogénio" para induzir uma resposta imune ainda noutro animal, produzindo um, assim designado, anticorpo anti-anti-Id. O anti-anti-Id pode ser epitopicamente idêntico ao mAb original que induziu o anti-Id. Assim, usando anticorpos para as determinantes idiotípicas de um mAb, é possível identificar outros clones que expressam anticorpos de especificidade idêntica.

Deste modo, os mAbs gerados contra a proteína GRB do presente invento podem ser usados para induzir anticorpos anti-Id em animais adequados, tais como ratinhos BALB/c. As células de baço desses ratinhos imunizados são usadas para produzir hibridomas anti-Id que segregam mAbs anti-Id. Além disso, os mAbs anti-Id podem ser acoplados a um portador tal como hemocianina de lapa do tipo fissurela ("keyhole limpet") (KLH) e usados para imunizar mais ratinhos BALB/c. Os soros destes ratinhos conterá anticorpos anti-anti-Id que possuem as propriedades de ligação do

mAb original específico para um epítopo de proteína GRB.

Os mAbs anti-Id possuem assim os seus próprios epítopos idiotípicos ou "idiotopos" estruturalmente similares ao epítopo a ser investigado, tal como proteína-α GRB.

O termo "anticorpo" também se pretende que inclua tanto as moléculas intactas como os seus fragmentos, tais como, por exemplo, Fab e $F(ab')_2$, os quais são capazes de ligar antigénio. Os fragmentos Fab e $F(ab')_2$ são desprovidos do fragmento Fc do anticorpo intacto, desaparece mais depressa da circulação e pode ter menor ligação de tecido não específica do que um anticorpo intacto (Wahl et al., J. Nucl. Med. 24:316-325 (1983)).

Será de notar que os fragmentos Fab e $F(ab')_2$ e outros fragmentos dos anticorpos úteis do presente invento podem ser usados para a detecção e quantificação da proteína GRB de acordo com os processos aqui revelados para moléculas de anticorpo intactas. Estes fragmentos são tipicamente produzidos por clivagem proteolítica, usando enzimas tais como a papaína (para produzir fragmentos Fab) ou a pepsina (para produzir fragmentos $F(ab')_2$).

Diz-se que um anticorpo é "capaz de ligar" uma molécula se ele é capaz de reagir especificamente com a molécula, para assim ligar a molécula ao anticorpo. O termo "epítopo" pretende designar a porção de qualquer molécula capaz de ser ligada por um anticorpo que pode também ser reconhecida por esse anticorpo. Os epítopos ou "determinantes antigénicas" consistem usualmente em agrupamentos de superfície quimicamente activa de moléculas tais como aminoácidos ou cadeias laterais de açúcar e têm características estruturais tridimensionais específicas bem como características de carga específicas.

Um "antigénio" é uma molécula ou uma porção de uma molécula capaz de ser ligada por um anticorpo que é, adicionalmente, capaz de induzir um animal a produzir anticorpo capaz de se ligar a um epítopo desse antigénio. Um antigénio pode ter um ou mais de um

epítopo. A reacção específica referida acima destina-se a indicar que o antigénio reagirá, de uma maneira altamente selectiva, com o seu anticorpo correspondente e não com os muitos outros anticorpos que podem ter sido evocados ("evoked") por outros antigénios.

Os anticorpos, ou fragmentos de anticorpos, úteis no presente invento podem ser usados para detectar quantitativa ou qualitativamente a presença de células que expressem a proteína GRB. Isto pode ser conseguido por técnicas de imunofluorescência empregando um anticorpo marcado fluorescentemente (ver abaixo) acompanhado de detecção microscópica de luz, citométrica de fluxo ou fluorimétrica.

Os anticorpos (ou seus fragmentos) úteis no presente invento podem ser empregues histologicamente, como em microscopia imunoelectrónica ou imunofluorescência, para a detecção in situ das proteínas GRB. A detecção in situ pode ser conseguida removendo o especímen histológico de um paciente e proporcionando um anticorpo marcado do presente invento a esse especímen. O anticorpo (ou fragmento) é preferivelmente proporcionado por aplicação ou por deposição do anticorpo marcado (ou fragmento) numa amostra biológica. Usando um tal procedimento, é possível determinar não só a presença da proteína GRB mas também a sua distribuição no tecido examinado. Usando o presente invento, os peritos na arte facilmente perceberão que qualquer um da grande variedade de métodos histológicos (tais como procedimentos de coloração) pode ser modificado para se conseguir essa detecção in situ.

Estes ensaios para a proteína GRB compreendem tipicamente a incubação de uma amostra biológica, tal como um fluido biológico, um extracto de tecido, células recém-recolhidas tais como linfócitos ou leucócitos, ou células que foram incubadas em cultura de tecido, na presença de um anticorpo marcado detectavelmente capaz de identificar proteína GRB, e de detectar o anticorpo por qualquer uma de várias técnicas bem conhecidas na arte.



A amostra biológica pode ser tratada com um suporte de fase sólida, tal como nitrocelulose, ou outro suporte sólido que seja capaz de imobilizar células, partículas de células ou proteínas solúveis. O suporte pode então ser lavado com tampões adequados, seguindo-se o tratamento com o anticorpo específico para a proteína GRB, marcado detectavelmente. O suporte de fase sólida pode então ser lavado com o tampão uma segunda vez, para remover o anticorpo não ligado. A quantidade de marcador ligado no referido suporte sólido pode então ser detectado por meios convencionais.

Por "suporte de fase sólida" entende-se qualquer suporte capaz de ligar antigénio ou anticorpos. Suportes ou portadores bem conhecidos incluem vidro, poliestireno, polipropileno, polietileno, dextrana, nilão, amilases, celuloses naturais e modificadas, poliacrilamidas, gabros e magnetite. Para os objectivos do presente invento a natureza do portador pode ser ou solúvel em certa extensão ou insolúvel. O material de suporte pode ter virtualmente qualquer configuração estrutural possível desde que a molécula acoplada seja capaz de se ligar a um antigénio ou anticorpo. Assim, a configuração do suporte pode ser esférica, como no caso de uma conta, ou cilíndrica, como no caso da superfície interior de um tubo de ensaio ou da superfície externa de uma vareta. Alternativamente, a superfície pode ser plana tal como uma folha, uma tira de ensaio, etc. Os suportes preferidos incluem contas de poliestireno. Os peritos na arte saberão de muitos outros portadores adequados para ligar anticorpo ou antigénio, ou serão capazes de os determinar usando experimentação de rotina.

A actividade de ligação de um dado lote de anticorpo anti-GRB-1 e anti-GRB-2 pode ser determinada de acordo com métodos bem conhecidos. Os peritos na arte saberão determinar as condições de ensaio operativas e óptimas para cada determinação usando experimentação de rotina.

Aos ensaios podem juntar-se outros passos tais como lavagem, agitação, vibração, filtração e semelhantes, como é habitual ou

necessário para a situação particular.

Uma das vias para marcar detectavelmente o anticorpo específico para GRB é a sua ligação a uma enzima e a utilização de um imunoensaio enzimático (EIA). Esta enzima, por sua vez, quando posteriormente exposta a um substrato apropriado reagirá com o substrato de maneira a produzir uma "porção" química que pode ser detectada, por exemplo, por meios espectrofotométricos, fluorimétricos ou visuais. As enzimas que podem ser usadas para marcar detectavelmente o anticorpo incluem, não restritivamente, malato-desidrogenase, nuclease estafilocócica, delta-5-esteróide--isomerase, levedura álcool-desidrogenase, alfa-glicerofosfato--desidrogenase, triose fosfato-isomerase, peroxidase de rábano, fosfatase alcalina, asparaginase, glucose-oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-fosfato-desidrogenase, glucoamilase e acetilcolinaesterase. A detecção pode ser conseguida por métodos colorimétricos que empregam um substrato cromogénico para a enzima. A detecção também pode ser conseguida por comparação visual da extensão da reacção enzimática de um substrato em comparação com padrões similarmente preparados.

A detecção pode ser conseguida usando qualquer um de uma variedade de outros imunoensaios. Por exemplo, marcando radioactivamente os anticorpos, ou fragmentos de anticorpos, é possível detectar R-PTPase pelo uso de um radioimunoensaio (RIA). Uma boa descrição de RIA pode ser encontrada em Laboratory Techniques and Biochemistry in Molecular Biology, por Work, T.S., et al., North Holland Publishing Company, NY, (1978) com referência particular ao capítulo entitulado "An Introduction to Radioimmune Assay and Related Techniques" de Chard, T., aqui incorporado por referência. O isótopo radioactivo pode ser detectado por meios tais como o uso de um contador gama ou de um contador de cintilações ou por autoradiografia.

É também possível marcar o anticorpo com um composto fluorescente. Quando o anticorpo marcado fluorescentemente é exposto a luz com o comprimento de onda apropriado, a sua

presença pode então ser detectada devido à fluorescência. Entre os compostos marcadores fluorescentes mais comummente usados estão o isotiocianato de fluoresceína, a rodamina, a ficoeritrina, a ficocianina, a aloficocianina, o o-ftaldeído e a fluorescamina.

O anticorpo também pode ser marcado detectavelmente usando metais emissores de fluorescência tais como ^{152}Eu ou outros metais da série dos lantanídeos. Estes metais podem ser ligados ao anticorpo usando grupos quelantes de metais tais como ácido dietilenotriaminapentaacético (DTPA) ou ácido etilenodiaminatetraacético (EDTA).

O anticorpo também pode ser marcado detectavelmente acoplando-o a um composto quimiluminescente. A presença do anticorpo marcado quimiluminescente é então determinada detectando a presença de luminescência que surge durante o decurso de uma reacção química. Exemplos de compostos marcadores quimiluminescentes particularmente úteis são luminol, isoluminol, éster acridínio teromático, imidazol, sal acridínio e éster oxalato.

De igual modo, pode usar-se um composto bioluminescente para marcar o anticorpo do presente invento. A bioluminescência é um tipo de quimiluminescência encontrada em sistemas biológicos nos quais uma proteína catalítica aumenta a eficiência da reacção quimiluminescente. A presença de uma proteína bioluminescente é determinada detectando a presença de luminescência. Os compostos bioluminescentes importantes para marcação são a luciferina, luciferase e aequorin.

As moléculas de anticorpo do presente invento podem ser adaptadas para utilização num ensaio imunométrico, também conhecido como "dois sítios" ou "sanduíche". Num ensaio imunométrico típico, uma quantidade de anticorpo não marcado (ou fragmento de anticorpo) é ligada a um suporte sólido e uma quantidade de anticorpo solúvel marcado detectavelmente é adicionada para permitir a detecção e/ou quantificação do complexo ternário formado entre anticorpo da fase sólida,

antigénio e anticorpo marcado.

Os ensaios imunométricos típicos e preferidos incluem ensaios "avançados" nos quais o anticorpo ligado à fase sólida é primeiro contactado com a amostra a ser testada para extrair o antigénio da amostra, por formação de um complexo na fase sólida binária de anticorpo-antigénio. Depois de um período de incubação adequado, o suporte sólido é lavado para remover o resíduo da amostra de fluido, incluindo antigénio não reagido, se existir, e em seguida contactado com a solução contendo uma quantidade desconhecida de anticorpo marcado (o qual funciona como "molécula reporter"). Depois de um segundo período de incubação para permitir que o anticorpo marcado se complexe com o antigénio ligado ao suporte sólido através do anticorpo não marcado, o suporte sólido é lavado uma segunda vez para remover o anticorpo marcado não reagido.

Ainda noutro tipo de ensaio "sanduíche", que pode também ser útil com os antigénios do presente invento, usam-se os chamados ensaios "simultâneo" e "inverso". Um ensaio simultâneo envolve um único passo de incubação, visto que o anticorpo ligado ao suporte sólido e o anticorpo marcado são adicionados ao mesmo tempo à amostra a ser testada. Depois da incubação estar terminada, o suporte sólido é lavado para remover o resíduo da amostra de fluido e o anticorpo marcado não complexado. A presença de anticorpo marcado associado ao suporte sólido é então determinada como para um ensaio de sanduíche "avançado" convencional.

No ensaio "inverso", utiliza-se adição passo a passo, primeiro de uma solução de anticorpo marcado, à amostra de fluido seguida de adição de anticorpo não marcado ligado a um suporte sólido, depois de um período de incubação adequado. Depois de uma segunda incubação, a fase sólida é lavada de forma convencional para a libertar do resíduo da amostra a ser testada e da solução de anticorpo marcado não reagido. A determinação de anticorpo marcado associado a um suporte sólido é então feita tal como nos ensaios "simultâneo" e "avançado".

EXEMPLO I

Efectuou-se um estudo para determinar a detectabilidade da ligação do domínio C-terminal de EGFR a uma proteína contendo o domínio SH2, imobilizada em filtros de nitrocelulose. Com este objectivo, foi avaliada a ligação do domínio C-terminal a uma proteína de fusão expressa bacterianamente (ver Figura 1).

A. Isolamento e Marcação do Domínio Carboxiterminal do EGFR

A porção intracelular do EGFR, que inclui o domínio da tirosina-quinase e o domínio do terminal carboxi, foi purificada de baculovírus recombinantes que expressam ADNc complementarmente ao domínio intracelular do EGFR humano, como foi descrito anteriormente (Hsu, C-Y. et al., Cell Growth and Differentiation 1:191-200 (1990)). A proteína recombinante (2 μg) foi então fosforilada com [gama-^{32}P]ATP (200 μCi, 6000 Ci/Mmol), a 4 °C em tampão HNTG (HEPES 20mM, pH 7,5, NaCl 150 mM, Triton X-100 0,1% e glicerol 10%) que continha $MnCl_2$ 5 mM. Para remover [Y-^{32}P]ATP não incorporado, a quinase fosforilada foi diluída a 1 ml com HEPES 20mM, pH 7,5 contendo 100 μg de BSA e em seguida concentrada num Centricon-10 até um volume de 50 μl. Este procedimento foi repetido 3 vezes, resultando na remoção de >99% do ATP não incorporado. Para separar o domínio C-terminal do domínio da quinase, a proteína concentrada foi então digerida com brometo de cianogénio (CNBr) em ácido fórmico a 70% durante 14 horas à temperatura ambiente (ver também Exemplo VI, abaixo). As amostras foram então lavadas três vezes com água, secas e ressuspensas em tampão de ligação até uma concentração de 2 x 10^6 cpm/ml.

B. Ligação do Domínio C-Terminal do EGFR a Proteína de Fusão trpE/GAP-SH2 Expressa Bacterianamente Imobilizada sobre Nitrocelulose

A TrpE e a TrpE/GAP-SH2 foram obtidas do laboratório do Dr. Tony Pawson e preparadas como se descreveu anteriormente (Moran,

M.F. et al., Proc. Natl. Acad. Sci. USA 87: 8622-8626 (1990)). Os estudos de ligação em filtros foram efectuados de acordo com métodos publicados (Schneider, W.J. et al., Proc. Natl. Acad. Sci. 76:5577-5581 (1979); Daniel, T.O. et al., J. Biol. Chem. 258:4606-4611 (1983)) com pequenas modificações. colocaram-se várias concentrações tanto de proteína de fusão de TrpE expressa bacterianamente como de proteína bacteriana sózinha em filtros de nitrocelulose. Depois de se bloquearem os filtros durante 1 hora a 4°C em PBS contendo leite em pó Carnation, adicionou-se domínio C-terminal do EGFR, marcado com ^{32}P, e a incubação continuou durante a noite a 4 °C. Após 24 horas, os filtros de nitrocelulose foram lavados 3 vezes à temperatura ambiente com PBS e expostos a um filme Kodak XAR-5 a -80 °C.

C. Resultados

O método anterior permitiu a detecção de ligação específica do domínio C-terminal de EGFR a menos de 5 ng de uma proteína de fusão GAP-SH2 expressa bacterianamente. A ligação foi específica, uma vez que necessitou da fosforilação da tirosina da sonda e não ocorreu quando se aplicaram proteínas irrelevantes aos filtros de nitrocelulose.

A demonstração de que o domínio C-terminal de EGFR se podia ligar especificamente a uma proteína contendo SH2 imobilizada sobre filtros de nitrocelulose encorajou os presentes inventores a aplicar esta abordagem à pesquisa de bibliotecas de expressão de lambda gt11 com o objectivo de identificar novas proteínas de ligação a EGFR.

EXEMPLO II

Pesquisa de Bibliotecas de Expressão e Isolamento de um Clone de ADNc Codificando uma Nova Proteína Contendo SH2

A cauda C-terminal fosforilada em tirosina do EGFR foi usada

como uma sonda para pesquisar bibliotecas de expressão de vários tecidos humanos diferentes, como se descreveu acima. A abordagem à pesquisa é delineada na Figura 2. Usando esta abordagem foram até agora isolados vários clones positivos, dois dos quais foram analisados em detalhe.

A. Pesquisa da Biblioteca de ADNc

Uma biblioteca lambda gt11, construída a partir de ARNm isolado de pedúnculo cerebral humano foi obtida de M. Jaye. Para pesquisar a biblioteca, colocaram-se em placas fagos lambda gt11 a uma densidade suficiente para produzir 4 x 10^4 placas por 150 mm de placa de ágar. Foram inicialmente pesquisadas um total de seis placas. Após incubação das placas durante 4 horas a 42 °C, as placas foram cobertas com filtros de nitrocelulose que tinham sido impregnados com isopropil-B-D-tiogalactopiranósido (IPTG), como descrito anteriormente (MacGregor, P.F. et al., Oncogene 5:451-458 (1990)). A incubação foi continuada durante a noite a 37 °C. Os filtros foram então removidos, lavados com tBST (Tris-HCl 10 mM, pH 8, NaCl 150 mM e Triton X-100 0,05%) à temperatura ambiente, e em seguida, bloqueados em tampão HBB (HEPES 20 mM, pH 7,5, Mg/Cl 5 mM, KCl 1 mM) contendo 5% de leite em pó "Carnation" durante 1 hora a 4 °C, como descrito (MacGregor et al., supra). A seguir ao bloqueio, adicionou-se sonda do término carboxi (término C) fosforilada na tirosina e marcada a uma concentração de 1,6 x 10^{-4} µg/ml e a incubação continuou durante a noite. Os filtros foram então lavados 3 vezes à temperatura ambiente em PBS contendo Triton X-100 0,2%. Os filtros foram secos e expostos a um filme Kodak XAR-5 a -80 °C.

Recolheram-se das placas rolos de ágar, correspondendo aos clones positivos, e colocaram-se em 1 ml de meios SM. Depois de se deixar os fagos difundirem-se do ágar, os fagos foram de novo colocados em placas e pesquisados como se descreveu acima. Os fagos que demonstraram enriquecimento na pesquisa subsequente foram isolados e sequenciados. O ADN do fago lambda gt11 foi isolado pelo método do lisado em placa de acordo com Maniatis et al. e subclonado em M13 MP19 digerido com EcoRI (Maniatis et al.,

1982). O ADN de cadeia simples foi isolado e sequenciado pelo método da terminação da cadeia didesoxi usando um conjunto de sequenciação de ADN Sequenase (United States Biochemical).

Numa experiência, pesquisaram-se 240 000 pfu de uma biblioteca lambda gt 11 de pedúnculo cerebral humano. Isolou-se uma única placa, o clone ki4 (Figura 3A). Na pesquisa subsequente, este clone demonstrou enriquecimento e na terceira pesquisa todas as placas ligaram a sonda (Figura 3B). O clone ki4 continha uma inserção de cerca de 900 nucleótidos, os quais, após indução do promotor lac com IPTG, produziram uma proteína de fusão que se pôde ligar a EGFR. O tamanho da proteína de fusão fez prever que a inserção de ADNc codificava uma proteína de cerca de 300 aminoácidos, que era o tamanho esperado se o ADNc contivesse uma única estrutura de leitura aberta grande.

Para analisar o clone ki4 em maior detalhe, o ADN foi isolado e o fragmento EcoRI, correspondendo à inserção de ADNc humano, foi subclonado em M13 e sequenciado. A tradução da sequência desta inserção demonstrou uma única estrutura de leitura aberta grande, a qual após análise usando a base de dados Genbank se verificou conter um segmento simples de cerca de 100 aminoácidos com homologia de sequência com os domínios SH2 de outras proteínas conhecidas (Figuras 4 e 5A). Contudo, noutras regiões, não foi observada homologia de sequência. Assim, usando esta abordagem para a pesquisa, foi identificada uma nova proteína contendo SH2 que se podia ligar ao EGFR.

B. Isolado de ADNc de Comprimento Total

O clone inicial isolado codificava um domínio SH2, mas não continha as extremidades 3' ou 5' do gene. Para isolar o ADNc de comprimento total, a biblioteca foi de novo pesquisada usando ADN isolado do fago positivo inicial. O ADN, de bacteriófago M13 recombinante que expressou o clone positivo, foi amplificado usando um ciclizador térmico, polimerase Taq1 e oligonucleótidos complementares às regiões flanqueadoras de EcoR1 do vector M13. Para se obterem clones contendo mais sequência 5' na informação,

gerou-se também um segundo produto de ADN amplificado, correspondendo aos 250 nucleótidos mais 5' do fago isolado inicial, usando oligonucleótidos complementares às sequências de ambas as extremidades desta região. Foram então preparadas sondas de ADN marcadas com [^{32}P], por tradução por corte dos produtos amplificados.

Para pesquisar de novo a biblioteca de ADNc, a biblioteca foi de novo colocada em placas como se descreveu acima. Depois da incubação das placas durante 8 horas a 37 °C, as placas foram arrefecidas durante 1 hora a 4 °C e em seguida o ADN de fago foi transferido para filtros de nitrocelulose. Os filtros foram desnaturados numa solução de NaOH 0,2 N e NaCl 1,5 M e em seguida cozidos in vacuo durante 2 horas a 80 °C (Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, 2ª Edição, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)). Depois da pré-hibridação dos filtros durante 1 hora a 42 °C, foi adicionada a sonda de ADN marcada com ^{32}P e a hibridação continuou durante a noite a 42 °C numa solução contendo 5X Denhardt, formamida 50%, 5X SSC, SDS 0,1%, Tris-HCl 200 mM, pH 7,6 e 100 µg/ml de ADN de esperma de salmão. Os filtros foram então lavados numa solução contendo 0,1X SSC e SDS 0,1%, secos e expostos a filme Kodak XAR-5 a -70 °C. os clones positivos foram então isolados e sequenciados como se descreveu acima.

Visto que a inserção do clone ki4 era desprovida das extremidades 3' e 5' do gene, a biblioteca foi de novo pesquisada usando duas sondas de ADN que tinham sido geradas por amplificação de ADN de clone ki4. Esta abordagem permitiu a identificação de cinco clones adicionais. Três dos clones estendiam-se 3' a partir do clone ki4 inicial, dois dos quais, os clones ki2,2 e ki2,4, continham um sinal de poliadenilação e uma região 3' não traduzida longa (>1000 nucleótidos). Além disso, estes clones codificavam uma proteína que continha um segundo domínio SH2 (Figuras 4 e 5A).

Os outros dois clones, ki3,0 e ki5,3, estendiam-se 5' a partir do clone ki4. Ambos os clones continham estruturas de leitura

aberta grandes e um codão AUG que concordava com os critérios de iniciação de tradução tal como definidos por Kozak (Kozak, M. J. Cell. Biol. 108:229-241 (1989)). Contudo, apenas o clone ki3,0, quando traduzido em proteína e comparado com sequências conhecidas do Genbank, se verificou conter um domínio de 50 aminoácidos que era homólogo aos domínios SH3 presentes noutras proteínas conhecidas. O peso molecular previsto da proteína de comprimento total codificada pelos clones sobrepostos, ki2,2 e ki3,0, era de cerca de 84 kDa. Esta nova proteína foi designada por GRB-1.

EXEMPLO III

A Proteína GRB-1 Contém domínios SH2 e SH3

A análise da sequência da proteína GRB-1, por comparação com as sequências da base de dados Genbank, revelou a presença de dois segmentos de cerca de 100 aminoácidos, iniciando-se nos aminoácidos 333 e 624, com homologia de sequência com os domínios SH2 de outras proteínas que se sabe interagirem com o EGFR (Figura 5A). Embora a GRB-1 apresentasse homologia de batida com outros domínios SH2 ao nível da proteína, ela não revelou uma homologia significativa ao nível do ADN. A GRB-1 continha também um segmento de cerca de 50 aminoácidos, localizado na região N-terminal, que tinha homologia de sequência com os domínios SH3 (Figura 4 e 5B).

Na Figura 6 apresenta-se uma comparação entre a organização estrutural de GRB-1 e a de várias outras proteínas contendo SH2/SH3. Deste esquema é evidente que a localização dos domínios SH2 e SH3 varia de proteína para proteína. Apesar disto, existem algumas semelhanças e diferenças entre estas proteínas contendo SH2. A GRB-1 é semelhante a outros substratos que se verificou interagirem com EGFR, tais como PLC-gama e GAP, por a GRB-1 conter dois domínios SH2 e um só domínio SH3. Contudo, ao contrário destes substratos, a GRB-1 não apresenta homologia com qualquer domínio catalítico conhecido, e neste aspecto assemelha-se à proteína codificada pelo vírus do sarcoma de ave, v-crk.

Além destas regiões não existe homologia de sequência com outras sequências presentes no Genbank. Em particular, a GRB-1 era desprovida de um domínio de ligação a ATP de consenso e não apresentava homologia de sequência com nenhuma serina/ /treonina-quinase ou tirosina-quinase.

Pensa-se que o domínio SH2 proporcione um motivo comum com o qual as moléculas sinalizadoras enzimaticamente distintas podem ser ligadas a receptores activados com actividade de tirosina--quinase (Moran, M.F. et al., Proc. Natl. Acad. Sci. USA 87: 8622-8626 (1990); Anderson, D. et al., Science 250: 979-982 (1990)).

A presença dos domínios SH2 em GRB-1 (Figura 4) e em GRB-2 reforça ainda a importância desta domínio na mediação da interacção destas proteínas com a cauda C-terminal do EGFR. Além disso, visto que estão por identificar muitas proteínas capazes de interagir com receptores de tirosina , isto sugere que estão por descobrir outros membros desta família de proteínas.

Além de conter dois domínios SH2, a GRB-1 também contém um domínio SH3. O domínio SH3 é um domínio não catalítico de cerca de 50 resíduos de aminoácido que é partilhado por muitas proteínas contendo SH2. Uma vez que os domínios SH3 também são encontrados em proteínas citosqueléticas, tais como espectrina e fodrina, a função deste domínio poderia ser a de localizar estas proteínas na membrana ou submembrana do citoesqueleto onde elas interagiriam com outras moléculas.

A comparação entre a sequência de aminoácidos deduzida de GRB-1 e o produto proteico codificado pelo oncogene de ave v-crk pode trazer luz à função de GRB-1. O gene v-crk codifica uma proteína que é composta principalmente por uma proteína viral gag fundida com um domínio SH2 e SH3 (Mayer, B.J. et al., Nature 332: 272-275 (1988)). Ambas as proteínas GRB-1 e $p47^{gag-crk}$ não têm homologia com quaisquer domínios catalíticos conhecidos. Contudo, os fibroblastos de embrião de galinha transformados com

$p47^{gag-crk}$ apresentam níveis elevados de proteínas contendo fosfotirosina (Mayer, B.J. et al., supra; Proc. Natl. Acad. Sci. USA 87: 2638-2642 (1990); Matsuda, M. et al., Science 248: 1537-1539 (1990)).

Uma vez que o produto v-crk mostrou ligar várias proteínas contendo fosfotirosina em células transformadas com v-crk pode ser que a função de c-crk seja a de actuar como uma ponte entre quinases e substratos. Relativamente a este ponto, é intrigante que a GRB-1, tal como GAP e PLC-gama, contenha dois domínios SH2, a combinação dos quais pode ser idealmente adequada para ligar outras proteínas para receptores de tirosina-quinase activados.

EXEMPLO IV

Análise Northern da Expressão de GRB-1

A. Métodos

Preparou-se ARN celular total a partir de tecido de macaco pelo método de isotiocianato de guanidínio/cloreto de césio descrito por Sambrook, J. et al., (supra). Preparou-se Poly $(A)^+$ ARN por cromatografia de oligo(dT) celulose. Para a análise Northern, o ARN foi fraccionado por tamanhos por electroforese num gel de agarose 1,2% /formaldeído 2,2M, transferido para uma membrana de nilão por acção da capilaridade e cozido a 80 °C durante 2 horas. A seguir à pré-hibridação, a mancha foi hibridada com uma sonda de ADN traduzido por corte, marcada com $[^{32}P]$, que foi preparada como se descreveu anteriormente. A hibridação foi realizada durante a noite a 42 °C na presença de formamida 50%, 5X SSC, SDS 0,1% e 5X Denhardt's. A membrana foi então lavada em 0,1X SSC, SDS 0,1% a 42 °C e resposta a um filme Kodak XAR a -70 °C durante 12 horas usando um filtro intensificador.

B. Resultados

Para testar a expressão de ARNm correspondente ao ADNc recém

isolado, a análise por transferência Northern de diferentes ARNm de tecido de macaco, sondados com ADN correspondente à inserção do clone ki4, demonstrou a presença de duas bandas principais de 4,4 kb e 7,0 kb na maioria dos tecidos examinados (Figura 7). A expressão foi mais elevada no cérebro, mostrando o coração, baço, fígado e timo níveis decrescentes de expressão. A mensagem de 4,4 kb corresponde ao tamanho esperado do transcripto que codificaria os clones isolados. Em contraste com os transcriptos de 4,4 kb e 7,0 kb observados na maioria dos tecidos, a pele continha dois ARNm, com um tamanho ligeiramente mais pequeno, de 3,6 e 6,6 kb.

Os transcriptos de 3,6, 6,6 e 7,0 kb podem representar formas de ARNm alternativamente divididas ("sliced") ou podem codificar espécies de ARNm distintas mas afins.

EXEMPLO V

Produção de Anticorpos anti-GRB-1 e Análise da Proteína de Fusão GRB-1

A. Métodos

Produziram-se anticorpos policlonais imunizando coelhos com a proteína de fusão de β-galactosidase expressa pelo clone do fago isolado inicial, ki4. Infectaram-se bactérias E. coli CAG 456 (obtidas do Dr. Michael Snyder, Yale University) com fago recombinante ki4 a uma multiplicidade de infecção de 10 e recuperou-se proteína de fusão de β-galactosidase a partir da pelota de proteína após 1,5 horas. Prepararam-se extractos de proteína, separaram-se num gel de SDS 6% e excisou-se do gel a banda correspondente à proteína de fusão e usou-se para imunização.

Cultivaram-se até à confluência a linha de células de glioblastoma humano U1242, a linha de células de carcinoma da bexiga de ratazana NBT II e células NIH3T3 em meio DMEM suplementado com soro bovino fetal a 10%. As células foram marcadas com [^{35}S]-metionina (50 μCi/ml) em soro bovino fetal a 0,5% e lisadas

após 12 horas como descrito anteriormente (Margolis, B. et al., Cell 57: 1101-1107 (1989)). Depois da imunoprecipitação com 10 µl de anticorpo acoplado a proteína A-Sepharose, as contas foram lavadas três vezes com uma solução contendo HEPES 20mM, pH 7,5, NaCl 300mM, glicerol 10%, Triton X-100 1%, SDS 0,1% e desoxicolato de sódio 1%. Depois de se ferver em tampão de amostra, as proteínas foram separadas num gel de SDS 8%.

B. Resultados

Induziram-se anticorpos policlonais contra a proteína de fusão de β-galactosidase expressa pelo fago isolado inicial. As experiências de imunoprecipitação, usando células marcadas biossinteticamente, demonstrou que estes anticorpos reconhecem uma proteína de 85 kDa em três linhas de células diferentes (Figura 8, faixas indicadas por "I"). O reconhecimento da proteína de 85 kDa por este anti-soro foi específico pois o soro pré-imune não reconheceu esta proteína (faixas indicadas por "P"). Estes resultados proporcionam suporte para o peso molecular previsto, baseado na sequência de aminoácidos da GRB-1 clonada.

C. Discussão

A verificação de que o gene para GRB-1 codifica uma proteína com um peso molecular esperado de 85 kDa, em conjunto com a demonstração de que anticorpos para GRB-1 imunoprecipitaram uma proteína de 85 kDa a partir de três diferentes linhas de células, sugere que a GRB-1 pode representar uma proteína particular que se mostrou anteriormente associar-se com receptores do factor de crescimento activado, nomeadamente p85. Embora a função exacta de p85 seja desconhecida, presumiu-se ser fosfatidilinositol (PI3)--quinase, uma vez que a actividade PI3-quinase copurificou com uma proteína de 85 kDa encontrada em células estimuladas com PDGF assim como transformadas com antigénio T médio (MTAg) (Kaplan, D.R., Cell 50: 1021-1029 (1987); Whitman, M. et al., Nature 315: 239-242 (1985); Coughlin, S.R. et al., Science 243:1191-1194 (1989)). A ausência de um sítio de ligação a ATP sustenta que a GRB-1 muito provavelmente não é uma quinase fosfolipídica. A

GRB-1 exibe uma identidade de sequência de 97% com a p85 murina e bovina. Assim, a GRB-1 é a contraparte de p85. A p85 recombinante é capaz de se ligar ao PDGFR ou ao EGFR, mas ela própria não contém actividade de quinase PI3 intrínseca. A p85, contudo, é encontrada associada com uma proteína fosforilada na tirosina de 110 kDa que pode ser a subunidade catalítica da quinase PI3. Embora a relação exacta entre a quinase PI3 e a p85 não seja conhecida, a sobreexpressão de p85 modula a interacção entre a quinase PI3 e o PDGFR. A p85 poderia funcionar como uma subunidade reguladora ou como uma ponte entre receptores activados e a quinase PI3.

EXEMPLO VI

O Terminus Carboxi Fosforilado na Tirosina do Receptor EGF é um Sítio de Ligação para GAP e PLC-gama

Os estudos descritos abaixo confirmam que a ligação de PLC-gama e de uma proteína de fusão contendo os domínios SH2 e SH3 de GAP (trpE/GAP SH2) são especificamente controlados pela autofosforilação do EGFR. Os resultados mostram que a fosforilação de PLC-gama reduz, na realidade, a sua associação com EGFR. Apresenta-se evidência de que tanto PLC-gama como a proteína de fusão trpE/GAP SH2 se ligam especificamente ao término C fosforilado na tirosina do EGFR. Em suma, estes resultados indicam que os domínios SH2/SH3 interactuam directamente com regiões contendo fosfotirosina do receptor EGF.

A. Materiais e Métodos

1. Linhas de células, receptores mutantes e proteínas de fusão

As linhas de células CD126 (Margolis, B.L., et al., J.

Biol. Chem. 246: 10667-10671 (1989a), HER14, K721 (Honegger, A.M. et al., Cell 51:199-209 (1987); Honegger, A.M. et al., Mol. Cell. Biol. 7: 4567-4571 (1987)) foram usadas como fontes para o receptor EGF do tipo selvagem, receptor EGF (kin^-) negativo para a quinase e o receptor EGF truncado (C-terminal) C-terminal, respectivamente. o domínio intracelular do receptor EGF (EGFR-C) foi purificado a partir de um sistema de expressão de baculovírus (Hsu, C-Y. et al., Cell Growth Differ. 1:191-200 (1990)) (Figura 9A). A 3TP1, uma linha de células que sobreexpressa ADNc de PLC-gama transfectado mas que não tem receptor EGF foi usada como uma fonte de PLC-gama (Margolis, B. et al., Science 248: 607-610 (1990b)).

A preparação de proteínas de fusão de trpE contendo o domínio SH2 de GAP (resíduos de GAP 171-448, Figura 9B) foi descrita por (Moran, M.F. et al., Proc. Natl. Acad. Sci. USA 87: 8622-8626 (1990)). Os lisados bacterianos contendo proteínas de fusão trpE/GAP SH2 foram preparados ressuspendendo 1 g de bactérias em 3 ml de Tris 50 mM pH7,5, EDTA 0,5 mM, PMSF 0,1 mM. Depois da incubação a 4 °C em 1 mg/ml de lisosima e 0,2 % de NP-40, as células foram sonicadas 5 vezes durante 5 segundos e o lisado foi clarificado por centrifugação durante 30 min a 10 000 g. Os lisados bacterianos foram diluídos 1:100 no tampão de lise Triton 1% com inibidores de proteínase e fosfatase como se descreveu acima e foram pré-depurados com proteína A-Sepharose.

2. Anticorpos, imunoprecipitação e imunotransferência

Usaram-se os seguintes anticorpos anti-EGFR (Figura 9A): (a) mAb108, um anticorpo monoclonal dirigido contra o domínio III do domínio extracelular (Lax, I. et al., EMBO J. 8:421-427 (1989)); (b) anticorpo antipeptídico RK2 específico para resíduos 984-996; (c) anticorpo antipeptídico C específico para resíduos 1176-1186; e (d) anticorpo antipeptídico F específico para resíduos 656-676. Para imunoprecipitação das proteínas de fusão trpE, utilizou-se um anticorpo monoclonal de ratinho contra trpE (Oncogene Science) ligado a IgG anti-ratinho ligada a agarose (Sigma). Para a imuno-transferência usou-se um anticorpo policlonal de coelho contra

trpE (Moran, M.F. et al., Proc. Natl. Acad. Sci. USA 87: 8622-8626 (1990)). Submeteu-se a imunotransferência PLC-gama e imunoprecipitou-se com um anticorpo antipeptídico policlonal de coelho já descrito (Margolis, B. et al., Cell 57: 1101-1107 (1989b)).

As técnicas usadas são descritas em vários referências do laboratório dos presentes inventores (Margolis, B.L. et al., J. Biol. Chem. 264: 10667-10671 (1989); Cell 57: 1101-1107 (1989)). As células não estimuladas foram feitas crescer até à confluência em Meio de Eagle Modificado por Dulbecco com soro de vitela a 10% e não alimentadas durante a noite em soro de vitela fetal a 1% antes da lise num tampão de lise Triton X-100 1% contendo inibidores de proteínase e fosfatase. os receptores EGF foram imunoprecipitados utuilizando anticorpos ligados a proteína A-Sepharose. Depois da lavagem do material receptor com HNTG (HEPES 20mM, pH 7,5, NaCl 150 mM, Triton X-100 0,1% e glicerol 10%), a autofosforilação foi induzida pela adição de $MnCl_2$ 5mM e ATP 30 μM. Os controlos foram incubados só com Mn^{2+}. Depois de mais lavagens com HNTG, adicionou-se lisado contendo PLC-gama (das células 3TP1) ou as proteínas de fusão bacterianas. Depois de se deixar que a ligação prossiga durante 90 min, realizaram-se mais lavagens com HNTG e processaram-se amostras em gel de SDS e submeteram-se a imunotransferência.

3. Clivagem com Brometo de Cianogénio (CNBr)

Fosforilou-se EGFR-C a 4 °C com $MnCl_2$ e ATP por vezes na pressença de [gama-^{32}P]ATP (NEN/Dupont, 6000 Ci/mmol). A preparação do receptor foi então ressuspensa em HEPES 20mM, pH 7,5 com 100 μg de BSA e concentrada num Centricon 10 (Amicon) até 50 μl. Adicionaram-se então 240 μl de ácido fórmico a 88% com dois grãos de CNBr e as amostras foram armazenadas sob azoto no escuro durante 14 h à temperatura ambiente. As amostras foram secas e lavadas três vezes com água num Speed-Vac (Savant) e em seguida ressuspensas em tampão de lise Triton 1%.

B. Resultados

Fez-se uma comparação entre a ligação de PLC-gama a EGFRs do tipo selvagem e mutantes (Figura 9A). Primeiro, os receptores do tipo selvagem e mutantes de células NIH-3T3 transfectadas foram imunoprecipitados e alguns dos imunoprecipitados de receptor foram deixados sofrer autofosforilação in vitro com ATP e Mn^{2+} (Margolis, B. et al., Mol. Cell. Biol. 10: 435-441 (1990a)). Em seguida, adicionaram-se lisados de células NIH-3T3 que sobre-expressam PLC-gama (Margolis, B. et al., Science 248: 607-610 (1990b)) e deixou-se a ligação prosseguir durante 90 min a 4 °C. Depois de se lavarem os imunoprecipitados com HNTG, determinou-se a quantidade de PLC-gama ligada por imunotransferência.

Como se ilustra na Figura 10, a PLC-gama liga-se apenas ao receptor do tipo selvagem fosforilado na tirosina e não ao receptor não fosforilado.

Para determinar a importância da autofosforilação, foram feitos dois estudos com receptores mutantes. A primeira coisa a ser examinada foi a ligação de PLC-gama a um receptor EGF truncado ao qual faltam 126 aminoácidos do C-término (CD126, Figura 9A) e desprovido de quatro sítios de autofosforilação principais (Downward, J. et al., Nature 311:483-485 (1984)). Este receptor truncado foi autofosforilado, provavelmente na tirosina 992 (Walton, G.M. et al., J. Biol. Chem. 265:1750-1754 (1990)). Contudo, apesar deste nível de autofosforilação na tirosina, a ligação de PLC-gama foi marcadamente reduzida quando comparada com a do receptor de comprimento total. Observou-se também associação reduzida com CD63, um receptor EGF mutante de delecção ao qual faltam 63 resíduos C-terminais contendo dois sítios de autofosforilação. Estes resultados sugerem um papel para o C-término do receptor na ligação ou modulação da ligação de PLC-gama ao receptor EGF.

A Figura 10 também demonstra qua a PLC-gama não se pode ligar ao receptor mutante kin^-. Para explorar a importância da autofosforilação neste efeito, fosforilou-se cruzadamente o receptor kin^- com o receptor CD126 (Honegger, A.M. et al., Proc. Natl. Acad. Sci. USA 86:925-929 (1989)). O resultado foi a

normalização da ligação de PLC-gama a níveis do tipo selvagem. Isto sugere que a fosforilação de receptor kin⁻ foi suficiente para normalizar a ligação a PLC-gama.

Para confirmar que o receptor kin⁻ sózinho podia ligar PLC-gama depois da fosforilação, este receptor foi fosforilado cruzadamente com um domínio citoplásmico de EGFR expresso por baculovírus, solúvel, (EGFR-C) que não se liga a mAb 108 (Figura 9A).

Embora a fosforilação cruzada não fosse tão forte como com o mutante CD126, a fosforilação na tirosina do mutante K721A e a ligação de PLC-gama foram claramente detectadas. Estas verificações confirmam que a fosforilação na tirosina do EGFR promove a ligação a PLC-gama.

O papel da fosforilação em tirosina em PLC-gama na interacção entre EGFR do tipo selvagem e PLC-gama foi examinado. A PLC-gama fosforilada na tirosina pôde ser dissociada do EGFR mais facilmente do que a PLC-gama não fosforilada (Figura 11), sugerindo uma afinidade mais baixa da PLC-gama fosforilada em tirosina pelo EGFR.

Estas verificações foram estendidas ao exame da ligação de uma proteína de fusão contendo domínio SH2 de trpE/GAP (Figura 9B) ao EGFR-C expressso por baculovírus. Tal como com o EGFR de comprimento total e a PLC-gama, o domínio da proteína de fusão trpE/GAP SH2 liga-se apenas ao EGFR-C fosforilado na tirosina (Figura 12A). A proteína trpE sózinha não se liga a EGFR-C. Similarmente, o EGFR-C fosforilado liga-se apenas a trpE/GAP SH2; contudo, a ligação não específica de EGFR-C não fosforilada foi elevada (Figura 12B). estes rasultados demonstram que o sítio de ligação do EGFR está situado no seu domínio intracelular.

Em geral, a proteína de fusão trpE/GAP SH2 liga-se com uma esteoquiometria mais elevada a EGFR de comprimento total do que a PLC-gama. Contudo, a proteína de fusão não era fosforilada em tirosina pelo EGFR. A proteína trpE/GAP SH2 liga-se muito melhor ao receptor de comprimento total fosforilado do que ao mutante de

delecção CD126 (Figura 13A). Como se mostra na Figura 13B, a fosforilação cruzada do receptor EGF de comprimento total kin⁻ pelo EGFR-C permite-lhe ligar-se à proteína trpE/GAP SH2.

Nos grupos de controlo, EGFR-C mostrou não melhorar a ligação ao receptor CD126, provavelmente porque este receptor estava já fosforilado ao máximo em tirosina (Figura 13A). Também não foi observada ligação quando se testou EGFR-C na presença de imunoprecipitado de mAb 108 de células não contendo receptor EGF (Figura 13B). Isto indica que os efeitos de EGFR-C não podem ser atribuídos à ligação não específica de EGFR-C fosforilado em tirosina a Sepharose. Estes estudos confirmam a importância da autofosforilação na mediação da ligação e mostram que para a ligação do receptor EGF, o domínio SH2 de GAP se comporta analogamente a PLC-gama intacta.

A fraca ligação ao mutante de delecção CD126 sugere que pelo menos parte do sítio de ligação para a molécula era no término C. Contudo, não pode ser excluído um efeito, possivelmente alostérico, desta delecção na conformação global do receptor. Assim, examinou-se a ligação de PLC-gama e trpE/GAP SH2 a um fragmento C-terminal do EGFR. No EGFR, o resíduo metionina mais C-terminal é encontrado na posição 983; a clivagem com CNBr gera portanto um fragmento de 203 aminoácidos o qual contém todos os sítios de autofosforilação conhecidos. Este fragmento de proteína é reconhecido por um anticorpo específico para o C-término de EGFR, anti-C (Figura 9A).

Quando este fragmento C-terminal foi especificamente imunoprecipitado e fosforilado em tirosina, ele ligou-se a PLC-gama e à proteína de fusão trpE/GAP SH2 (Figura 14). A clivagem com CNBr foi completa; não se detectou EGFR-C de comprimento total depois da proteólise, o que pode ser devido à ligação. Não se observou, de novo, ligação ao fragmento por CNBr, C-terminal não fosforilado. A clivagem com CNBr de EGFR-C gerou também um péptido N-terminal de 97 aminoácidos identificado pelo anticorpo F (Figura 9A, resíduos de EGFR 645-742). Este fragmento, imunoprecipitado pelo anticorpo F, não ligou trpE/GAP SH2. Adicionalmente, o EGFR-C foi

autofosforilado com [gama-^{32}P]ATP e gerou-se um fragmento C-terminal com CNBr marcado com ^{32}P. Como se mostra na Figura 15, este fragmento liga-se à proteína de fusão trpE/GAP SH2 mas não a trpE. No total estas verificações demonstram que a ligação directa ao C-término fosforilado em tirosina contribui pelo menos em parte para a ligação específica de proteínas com domínios SH2 e SH3 ao EGFR.

C. Discussão

Quando observadas em conjunto, as verificações anteriores e várias linhas adicionais de evidência sustentam fortemente que os resíduos fosfotirosina são parte do sítio de ligação real do EGFR para os domínios SH2. Primeiro verificou-se que a P47$^{gag-crk}$ se ligava a praticamente todas as proteínas contendo fosfotirosina em células transformadas v-crk (Matsuda, M. et al., Science 248:1537-1539 (1990)). Em segundo lugar, as mutações de dois sítios de autofosforilação no receptor PDGF diminuem muito a ligação de GAP (Kazlauskas, A. et al., Science 247:1578-1581 (1990)). Finalmente, os resultados apresentados acima demonstram que há ligação específica ao C-término do EGFR só quando está presente fosfotirosina.

Assim, conclui-se que os resíduos fosfotirosina ou constituem parte do sítio de ligação ou alteram localmente a conformação desta região, permitindo a ligação. É improvável que a fosfotirosina por si só constitua o sítio de ligação. Por exemplo, a fosfotirosina sózinha não consegue interferir na ligação de P47$^{gag-crk}$ a proteínas contendo fosfotirosina (Matsuda et al., supra). Adicionalmente, a PLC-gama não se liga a todas as moléculas activadas que contêm resíduos fosfotirosina, tais como o receptor CSF-1 (Downing, J.R. et al., EMBO J. 8:3345-3350 (1989)). Similarmente, a ligação de PLC-gama a PDGFR não parece ser idêntica à ligação a GAP; diferentes proteínas contendo domínios SH2 e SH3 podem ter diferentes especificidades de ligação (Kazlauskas et al., supra).

As referências citadas acima são todas aqui incorporadas por

referência, quer estejam especificamente incorporadas quer não.

Tendo-se descrito este invento na sua totalidade, será evidente para os peritos na arte que o mesmo pode ser executado com uma vasta gama de parâmetros, concentrações e condições equivalentes, sem se desviar do espírito e âmbito do invento e sem experimentação indevida.

Embora este invento tenha sido descrito relativamente a suas concretizações específicas deverá ser entendido que se lhe podem fazer outras modificações. Este pedido pretende cobrir quaisquer variações, usos ou adaptações dos inventos que seguem, no geral, os princípios do invento e incluir esses inventos derivados da presente revelação como provindo da prática conhecida ou habitual na arte à qual o invento pertence e como pode ser aplicada às características essenciais anteriormente estabelecidas que se observam do âmbito das reivindicações anexas. Deve entender-se que a fraseologia ou terminologia aqui empregue tem o fim de descrever e não de limitar o invento.

REIVINDICAÇÕES

1 - Processo para detectar a expressão, numa célula, de uma proteína capaz de se ligar a uma porção polipeptídica, fosforilada na tirosina, de uma molécula de tirosina-quinase de receptor, caracterizado por compreender:

(a) fazer contactar a referida célula, um seu extracto, um seu lisado ou um seu sobrenadante com um portador em fase sólida, provocando a ligação da referida proteína ao referido portador;

(b) incubar a referida proteína ligada ao portador com o referido polipéptido fosforilado na tirosina, permitindo que o referido polipéptido se ligue à referida proteína ligada ao portador;

(c) remover do referido portador os materiais não ligados; e

(d) detectar a presença ou medir a quantidade do referido polipéptido fosforilado na tirosina ligado ao referido portador;

detectando-se, assim, a expressão da referida proteína.

2 - Processo de acordo com a reivindicação 1, caracterizado por a referida tirosina-quinase de receptor ser seleccionada do grupo constituído por receptor do factor de crescimento epidérmico, receptor do factor de crescimento derivado de plaquetas, receptor do factor de crescimento de fibroblastos, receptor do factor-1 estimulante de colónias e receptor da insulina.

3 - Processo de acordo com a reivindicação 1, caracterizado por a referida proteína ser de origem humana.

4 - Processo de acordo com a reivindicação 1, caracterizado por o referido polipéptido fosforilado ser marcado detectavelmente.

5 - Processo de acordo com a reivindicação 4, caracterizado por o referido marcador detectável ser um radiomarcador.

6 - Processo de acordo com a reivindicação 5, caracterizado por o referido radiomarcador ser ^{32}P.

7 - Processo de mapeamento, num cromossoma humano, de um gene que codifica uma proteína que é capaz de se ligar a uma porção polipeptídica, fosforilada na tirosina, de uma molécula de tirosina-quinase de receptor, caracterizado por compreender:

(a) infectar células bacterianas com uma biblioteca de expressão de genes humanos;

(b) detectar um clone que expresse a referida proteína, usando um processo de acordo com a reivindicação 1;

(c) sequenciar o ADN do referido clone;

(d) mapear a referida sequência num cromossoma humano.

8 - Sonda polipeptídica útil para detectar a expressão de uma proteína capaz de se ligar a uma porção polipeptídica, fosforilada na tirosina, de uma molécula de tirosina-quinase de receptor, caracterizada por compreender uma sequência de aminoácidos derivada da porção fosforilada na tirosina da referida molécula de receptor, ou um seu derivado funcional, tendo a referida sequência, pelo menos, um resíduo de tirosina fosforilado e sendo desprovida da porção tirosina-quinase do referido receptor, e sendo a referida sonda marcada detectavelmente.

9 - Sonda de acordo com a reivindicação 8, caracterizada por possuir entre um e cinco resíduos de tirosina fosforilados.

10 - Sonda de acordo com a reivindicação 8, caracterizada por possuir entre 25 e 250 resíduos de aminoácido.

11 - Sonda de acordo com a reivindicação 8, caracterizada por o marcador detectável ser um átomo de fósforo radioactivo do referido resíduo de tirosina fosforilado.

12 - Processo de preparação de uma sonda de acordo com a reivindicação 8, caracterizado por compreender:

(a) proporcionar o referido receptor, ou um derivado do tipo

receptor construído por engenharia genética, numa forma substancialmente pura, possuindo o referido receptor ou o referido derivado do tipo receptor, um domínio de tirosina-quinase e um domínio fosforilado na tirosina, incluindo o referido domínio fosforilado na tirosina, pelo menos, um resíduo tirosina capaz de ser fosforilado pela referida tirosina-quinase;

(b) incubar o referido receptor ou derivado do tipo receptor com [gama-^{32}P]adenosina-trifosfato, sob condições que permitam a fosforilação do referido resíduo de tirosina, provocando a fosforilação do referido resíduo de tirosina;

produzindo-se, assim, a referida sonda.

13 - Processo de acordo com a reivindicação 12, caracterizado por compreender após o passo (b),

(c) adicionalmente, o tratamento da referida molécula de receptor fosforilada com um agente capaz de clivar a referida molécula, entre o domínio de tirosina-quinase e o domínio fosforilado na tirosina.

14 - Processo de acordo com a reivindicação 13, caracterizado por o referido agente ser brometo de cianogénio.

15 - Processo de acordo com a reivindicação 12, caracterizado por o referido derivado do tipo receptor ser um polipéptido codificado por uma molécula de ADN compreendendo uma sequência de ADN que codifica tirosina-quinase, ligada a uma sequência de ADN que codifica um sítio de clivagem enzimática selectivo, ligada a uma sequência de ADN que codifica o referido domínio carboxi-terminal, e por o referido agente ser uma enzima capaz de clivar no referido sítio de clivagem.

16 - Processo para purificar, a partir de uma mistura complexa, uma proteína capaz de se ligar a uma porção polipeptídica, fosforilada na tirosina, de uma molécula de tirosina-quinase de receptor, caracterizado por compreender:

(a) fazer contactar a referida mistura complexa com um

portador em fase sólida, ao qual está ligada uma sonda de acordo com a reivindicação 8, permitindo que a referida proteína se ligue à referida sonda;

(b) remover do referido portador os materiais não ligados;

(c) eluir do referido portador a referida proteína ligada;

purificando-se, assim, a referida proteína.

17 - Proteína, GRB-1, caracterizada por possuir a sequência de aminoácidos:

(segue Sequência)



```
     TACAACCAGGCTCAACTGTTGCATGGTAGCAGATTTGCAAACATGAGTGCTGAGGGGTAC
   1 ---------+---------+---------+---------+---------+---------+  60
     ATGTTGGTCCGAGTTGACAACGTACCATCGTCTAAACGTTTGTACTCACGACTCCCCATG

                                                M  S  A  E  G  Y  -

     CAGTACAGAGCGCTGTATGATTATAAAAAGGAAAGAGAAGAAGATATTGACTTGCACTTG
  61 ---------+---------+---------+---------+---------+---------+ 120
     GTCATGTCTCGCGACATACTAATATTTTTCCTTTCTCTTCTTCTATAACTGAACGTGAAC

     Q  Y  R  A  L  Y  D  Y  K  K  E  R  E  E  D  I  D  L  H  L

     GGTGACATATTGACTGTGAATAAAGGGTCCTTAGTAGCTCTTGGATTCAGTGATGGACAG
 121 ---------+---------+---------+---------+---------+---------+ 180
     CCACTGTATAACTGACACTTATTTCCCAGGAATCATCGAGAACCTAAGTCACTACCTGTC

     G  D  I  L  T  V  N  K  G  S  L  V  A  L  G  F  S  D  G  Q  -

     GAAGCCAGGCCTGAAGAAATTGGCTGGTTAAATGGCTATAATGAAACCACAGGGGAAAGG
 181 ---------+---------+---------+---------+---------+---------+ 240
     CTTCGGTCCGGACTTCTTTAACCGACCAATTTACCGATATTACTTTGGTGTCCCCTTTCC

     E  A  R  P  E  E  I  G  W  L  N  G  Y  N  E  T  T  G  E  R

     GGGGACTTTCCGGGAACTTACGTAGAATATATTGGAAGGAAAAAAATCTCGCCTCCCACA
 241 ---------+---------+---------+---------+---------+---------+ 300
     CCCCTGAAAGGCCCTTGAATGCATCTTATATAACCTTCCTTTTTTTAGAGCGGAGGGTGT

     G  D  F  P  G  T  Y  V  E  Y  I  G  R  K  K  I  S  P  P  T

     CCAAAGCCCCGGCCACCTCGGCCTCTTCCTGTTGCACCAGGTTCTTCGAAAACTGAAGCA
 301 ---------+---------+---------+---------+---------+---------+ 360
     GGTTTCGGGGCCGGTGGAGCCGGAGAAGGACAACGTGGTCCAAGAAGCTTTTGACTTCGT

     P  K  P  R  P  P  R  P  L  P  V  A  P  G  S  S  K  T  E  A  -

     GATGTTGAACAACAAGCTTTGACTCTCCCGGATCTTGCAGAGCAGTTTGCCCCTCCTGAC
 361 ---------+---------+---------+---------+---------+---------+ 420
     CTACAACTTGTTGTTCGAAACTGAGAGGGCCTAGAACGTCTCGTCAAACGGGGAGGACTG

     D  V  E  Q  Q  A  L  T  L  P  D  L  A  E  Q  F  A  P  P  D  -

     ATTGCCCCGCCTCTTCTTATCAAGCTCGTGGAAGCCATTGAAAAGAAAGGTCTGGAATGT
 421 ---------+---------+---------+---------+---------+---------+ 480
     TAACGGGGCGGAGAAGAATAGTTCGAGCACCTTCGGTAACTTTTCTTTCCAGACCTTACA

     I  A  P  P  L  L  I  K  L  V  E  A  I  E  K  K  G  L  E  C  -

 481 TCAACTCTATACAGAACACAGAGCTCCAGCAACCTGGCAGAATTACGACAGCTTCTTGAT
     ---------+---------+---------+---------+---------+---------+ 540
     AGTTGAGATATGTCTTGTGTCTCGAGGTCGTTGGACCGTCTTAATGCTGTCGAAGAACTA

     S  T  L  Y  R  T  Q  S  S  S  N  L  A  E  L  R  Q  L  L  D  -

 541 TGTGATACACCCTCCGTGGACTTGGAAATGATCGATGTGCACGTTTTGGCTGACGCTTTC
     ---------+---------+---------+---------+---------+---------+ 600
     ACACTATGTGGGAGGCACCTGAACCTTTACTAGCTACACGTGCAAAACCGACTGCGAAAG

     C  D  T  P  S  V  D  L  E  M  I  D  V  H  V  L  A  D  A  F  -
```



```
601  AAACGCTATCTCCTGGACTTACCAAATCCTGTCATTCCAGCAGCCGTTTACAGTGAAATG
     ---------+---------+---------+---------+---------+---------+ 660
     TTTGCGATAGAGGACCTGAATGGTTTAGGACAGTAAGGTCGTCGGCAAATGTCACTTTAC

     K  R  Y  L  L  D  L  P  N  P  V  I  P  A  A  V  Y  S  E  M  -

661  ATTTCTTTAGCTCCAGAAGTACAAAGCTCCGAAGAATATATTCAGCTATTGAAGAAGCTT
     ---------+---------+---------+---------+---------+---------+ 720
     TAAAGAAATCGAGGTCTTCATGTTTCGAGGCTTCTTATATAAGTCGATAACTTCTTCGAA

     I  S  L  A  P  E  V  Q  S  S  E  E  Y  I  Q  L  L  K  K  L  -

     ATTAGGTCGCCTAGCATACCTCATCAGTATTGGCTTACGCTTCAGTATTTGTTAAAACAT
721  ---------+---------+---------+---------+---------+---------+ 780
     TAATCCAGCGGATCGTATGGAGTAGTCATAACCGAATGCGAAGTCATAAACAATTTTGTA

     I  R  S  P  S  I  P  H  Q  Y  W  L  T  L  Q  Y  L  L  K  H

     TTCTTCAAGCTCTCTCAAACCTCCAGCAAAAATCTGTTGAATGCAAGAGTACTCTCTGAA
781  ---------+---------+---------+---------+---------+---------+ 840
     AAGAAGTTCGAGAGAGTTTGGAGGTCGTTTTTAGACAACTTACGTTCTCATGAGAGACTT

     F  F  K  L  S  Q  T  S  S  K  N  L  L  N  A  R  V  L  S  E  -

     ATTTTCAGCCCTATGCTTTTCAGATTCTCAGCAGCCAGCTCTGATAATACTGAAAACCTC
841  ---------+---------+---------+---------+---------+---------+ 900
     TAAAAGTCGGGATACGAAAAGTCTAAGAGTCGTCGGTCGAGACTATTATGACTTTTGGAG

     I  F  S  P  M  L  F  R  F  S  A  A  S  S  D  N  T  E  N  L  -

     ATAAAAGTTATAGAAATTTTAATCTCAACTGAATGGAATGAACGACAGCCTGCACCAGCA
901  ---------+---------+---------+---------+---------+---------+ 960
     TATTTTGAATATGTTTAAAATTAGAGTTGACTTACCTTACTTGCTGTCGGACGTGGTCGT

     I  K  V  I  E  I  L  I  S  T  E  W  N  E  R  Q  P  A  P  A

     CTGCCTCCTAAACCACCAAAACCTACTACTGTAGCCAACAACGGTATGAATAACAATATG
961  ---------+---------+---------+---------+---------+---------+ 1020
     GACGGAGGATTTGGTGGTTTTGGATGATGACATCGGTTGTTGCCATACTTATTGTTATAC

     L  P  P  K  P  P  K  P  T  T  V  A  N  N  G  M  N  N  N  M  -

     TCCTTACAAAATGCTGAATGGTACTGGGGAGATATCTCGAGGGAAGAAGTGAATGAAAAA
1021 ---------+---------+---------+---------+---------+---------+ 1080
     AGGAATGTTTTACGACTTACCATGACCCCTCTATAGAGCTCCCTTCTTCACTTACTTTTT

     S  L  Q  N  A  E  W  Y  W  G  D  I  S  R  E  E  V  N  E  K  -

     CTTCGAGATACAGCAGACGGGACCTTTTTGGTACGAGATGCGTCTACTAAAATGCATGGT
1081 ---------+---------+---------+---------+---------+---------+
     GAAGCTCTATGTCGTCTGCCCTGGAAAAACCATGCTCTACGCAGATGATTTTACGTACCA 1140

     L  R  D  T  A  D  G  T  F  L  V  R  D  A  S  T  K  M  H  G  -

     GATTATACTCTTACACTAAGGAAAGGGGGAAATAACAAATTAATCAAAATATTTCATCGA
1141 ---------+---------+---------+---------+---------+---------+ 1200
     CTAATATGAGAATGTGATTCCTTTCCCCCTTTATTGTTTAATTAGTTTTATAAAGTAGCT

     D  Y  T  L  T  L  R  K  G  G  N  N  K  L  I  K  I  F  H  R  -

     GATGGGAAATATGGCTTCTCTGACCCATTAACCTTCAGTTCTGTGGTTGAATTAATAAAC
1201 ---------+---------+---------+---------+---------+---------+ 1260
     CTACCCTTTATACCGAAGAGACTGGGTAATTGGAAGTCAAGACACCAACTTAATTATTTG

     D  G  K  Y  G  F  S  D  P  L  T  F  S  S  V  V  E  L  I  N  -
```



```
                CACTACCGGAATGAATCTCTAGCTCAGTATAATCCCAAATTGGATGTGAAATTACTTTAT
1261  ---------+---------+---------+---------+---------+---------+---------+ 1320
                GTGATGGCCTTACTTAGAGATCGAGTCATATTAGGGTTTAACCTACACTTTAATGAAATA

                H  Y  R  N  E  S  L  A  Q  Y  N  P  K  L  D  V  K  L  L  Y  -

1321  CCAGTATCCAAATACCAACAGGATCAAGTTGTCAAAGAAGATAATATTGAAGCTGTAGGG
      ---------+---------+---------+---------+---------+---------+---------+ 1380
      GGTCATAGGTTTATGGTTGTCCTAGTTCAACAGTTTCTTCTATTATAACTTCGACATCCC

      P  V  S  K  Y  Q  Q  D  Q  V  V  K  E  D  N  I  E  A  V  G  -

      AAAAAATTACATGAATATAACACTCAGTTTCAAGAAAAAAGTCGAGAATATGATAGATTA
1381  ---------+---------+---------+---------+---------+---------+---------+ 1440
      TTTTTTAATGTACTTATATTGTGAGTCAAAGTTCTTTTTTCAGCTCTTATACTATCTAAT

      K  K  L  H  E  Y  N  T  Q  F  Q  E  K  S  R  E  Y  D  R  I  -

      TATGAAGAATATACCCGCACATCCCAGGAAATCCAAATGAAAAGGACAGCTATTGAAGCA
1441  ---------+---------+---------+---------+---------+---------+---------+ 1500
      ATACTTCTTATATGGGCGTGTAGGGTCCTTTAGGTTTACTTTTCCTGTCGATAACTTCGT

      Y  E  E  Y  T  R  T  S  Q  E  I  Q  M  K  R  T  A  I  E  A  -

      TTTAATGAAACCATAAAAATATTTGAAGAACAGTGCCAGACCCAAGAGCGGTACAGCAAA
1501  ---------+---------+---------+---------+---------+---------+---------+ 1560
      AAATTACTTTGGTATTTTTATAAACTTCTTGTCACGGTCTGGGTTCTCGCCATGTCGTTT

      F  N  E  T  I  K  I  F  E  E  Q  C  Q  T  Q  E  R  Y  S  K

      GAATACATAGAAAAGTTTAAACGTGAAGGCAATGAGAAAGAAATACAAAGGATTATGCAT
1561  ---------+---------+---------+---------+---------+---------+---------+ 1620
      CTTATGTATCTTTTCAAATTTGCACTTCCGTTACTCTTTCTTTATGTTTCCTAATACGTA

      E  Y  I  E  K  F  K  R  E  G  N  E  K  E  I  Q  R  I  M  H  -

      AATTATGATAAGTTGAAGTCTCGAATCAGTGAAATTATTGACAGTAGAAGAAGATTGGAA
1621  ---------+---------+---------+---------+---------+---------+---------+ 1680
      TTAATACTATTCAACTTCAGAGCTTAGTCACTTTAATAACTGTCATCTTCTTCTAACCTT

      N  Y  D  K  L  K  S  R  I  S  E  I  I  D  S  R  R  R  L  E  -

      GAAGACTTGAAGAAGCAGGCAGCTGAGTATCGAGAAATTGACAAACGTATGAACAGCATT
1681  ---------+---------+---------+---------+---------+---------+---------+ 1740
      CTTCTGAACTTCTTCGTCCGTCGACTCATAGCTCTTTAACTGTTTGCATACTTGTCGTAA

      E  D  L  K  K  Q  A  A  E  Y  R  E  I  D  K  R  M  N  S  I  -

      AAACCAGACCTTATCCAGCTGAGAAAGACGAGAGACCAATACTTGATGTGGTTGACTCAA
1741  ---------+---------+---------+---------+---------+---------+---------+ 1800
      ATTGGTCTGGAATAGGTCGACTCTTTCTGCTCTCTGGTTATGAACTACACCAACTGAGTT

      K  P  D  L  I  Q  L  R  K  T  R  D  Q  Y  L  M  W  L  T  Q  -

      AAAGGTGTTCGGCAAAAGAAGTTGAACGAGTGGTTGGGCAATGAAAACACTGAAGACCAA
1801  ---------+---------+---------+---------+---------+---------+---------+ 1860
      TTTCCACAAGCCGTTTTCTTCAACTTGCTCACCAACCCGTTACTTTTGTGACTTCTGGTT

      K  G  V  R  Q  K  K  L  N  E  W  L  G  N  E  N  T  E  D  Q
```



```
     TATTCACTGGTGGAAGATGATGAAGATTTGCCCCATCATGATGAGAAGACATGGAATGTT
1861 ---------+---------+---------+---------+---------+---------+ 1920
     ATAAGTGACCACCTTCTACTACTTCTAAACGGGGTAGTACTACTCTTCTGTACCTTACAA

     Y  S  L  V  E  D  D  E  D  L  P  H  H  D  E  K  T  W  N  V  -

     GGAAGCAGCAACCGAAACAAAGCTGAAAACCTGTTGCGAGGGAAGCGAGATGGCACTTTT
1921 ---------+---------+---------+---------+---------+---------+ 1980
     CCTTCGTCGTTGGCTTTGTTTCGACTTTTGGACAACGCTCCCTTCGCTCTACCGTGAAAA

     G  S  S  N  R  N  K  A  E  N  L  L  R  G  K  R  D  G  T  F  -

     CTTGTCCGGGAGAGCAGTAAACAGGGCTGCTATGCCTGCTCTGTAGTGGTGGACGGCGAA
1981 ---------+---------+---------+---------+---------+---------+ 2040
     GAACAGGCCCTCTCGTCATTTGTCCCGACGATACGGACGAGACATCACCACCTGCCGCTT

     L  V  R  E  S  S  K  Q  G  C  Y  A  C  S  V  V  V  D  G  E  -

     GTAAAGCATTGTGTCATAAACAAAACAGCAACTGGCTATGGCTTTGCCGAGCCCTATAAC
2041 ---------+---------+---------+---------+---------+---------+ 2100
     CATTTCGTAACACAGTATTTGTTTTGTCGTTGACCGATACCGAAACGGCTCGGGATATTG

     V  K  H  C  V  I  N  K  T  A  T  G  Y  G  F  A  E  P  Y  N  -

     TTGTACAGCTCTCTGAAAGAACTGGTGCTACATTACCAACACACCTCCCTTGTGCAGCAC
2101 ---------+---------+---------+---------+---------+---------+ 2160
     AACATGTCGAGAGACTTTCTTGACCACGATGTAATGGTTGTGTGGAGGGAACACGTCGTG

     L  Y  S  S  L  K  E  L  V  L  H  Y  Q  H  T  S  L  V  Q  H  -

     AACGACTCCCTCAATGTCACACTAGCCTACCCAGTATATGCACAGCAGAGGCGATGAAGC
2161 ---------+---------+---------+---------+---------+---------+ 2220
     TTGCTGAGGGAGTTACAGTGTGATCGGATGGGTCATATACGTGTCGTCTCCGCTACTTCG

     N  D  S  L  N  V  T  L  A  Y  P  V  Y  A  Q  Q  R  R        -

     GCTTACTCTTTGATCCTTCTCCTGAAGTTCAGCCACCCTGAGGCCTCTGGAAAGCAAAGG
2221 ---------+---------+---------+---------+---------+---------+ 2280
     CGAATGAGAAACTAGGAAGAGGACTTCAAGTCGGTGGGACTCCGGAGACCTTTCGTTTCC

     GCTCCTCTCCAGTCTGATCTGTGAATTGAGCTGCAGAAACGAAGCCATCTTTCTTTGGAT
2281 ---------+---------+---------+---------+---------+---------+ 2340
     CGAGGAGAGGTCAGACTAGACACTTAACTCGACGTCTTTGCTTCGGTAGAAAGAAACCTA

     GGGACTAGAGCTTTCTTTCACAAAAAAGAAGTAGGGGAAGACATGCAGCCTAAGGCTGTA
2341 ---------+---------+---------+---------+---------+---------+ 2400
     CCCTGATCTCGAAAGAAAGTGTTTTTTCTTCATCCCCTTCTGTACGTCGGATTCCGACAT

     TGATGACCACACGTTCCTAAGCTGGAGTGCTTATCCCTTCTTTTTCTTTTTTTCTTTGGT
2401 ---------+---------+---------+---------+---------+---------+ 2460
     ACTACTGGTGTGCAAGGATTCGACCTCACGAATAGGGAAGAAAAAGAAAAAAAGAAACCA

     TTAATTTAAAGCCACAACCACATACAACACAAAGAGAAAAAGAAATGCAAAAATCTCTGC
2461 ---------+---------+---------+---------+---------+---------+ 2520
     AATTAAATTTCGGTGTTGGTGTATGTTGTGTTTCTCTTTTTCTTTACGTTTTTAGAGACG

     GTGCAGGGACAAAGAGGCCTTTAACCATGGTGCTTGTTAATGCTTTCTGAAGCTTTACCA
2521 ---------+---------+---------+---------+---------+---------+ 2580
     CACGTCCCTGTTTCTCCGGAAATTGGTACCACGAACAATTACGAAAGACTTCGAAATGGT
```



```
      GCTGAAAGTTGGGACTCTGGAGAGCGGAGGAGAGAGAGGCAGAAGAACCCTGGCCTGAGA
2581  ---------+---------+---------+---------+---------+---------+ 2640
      CGACTTTCAACCCTGAGACCTCTCGCCTCCTCTCTCTCCGTCTTCTTGGGACCGGACTCT

      AGGTTTGGTCCAGCCTGGTTTAGCCTGGATGTTGCTGTGCACGGTGGACCCAGACACATC
2641  ---------+---------+---------+---------+---------+---------+ 2700
      TCCAAACCAGGTCGGACCAAATCGGACCTACAACGACACGTGCCACCTGGGTCTGTGTAG

      GCACTGTGGATTATTTCATTTTGTAACAAATGAACGATATGTAGCAGAAAGGCACGTCCA
2701  ---------+---------+---------+---------+---------+---------+ 2760
      CGTGACACCTAATAAAGTAAAACATTGTTTACTTGCTATACATCGTCTTTCCGTGCAGGT

      CTCACAAGGGACGCTTTGGGAGAATGTCAGTTCATGTATGTTCAGAAGAAATTCTGTCAT
2761   ---------+---------+---------+---------+---------+---------+ 2820
      GAGTGTTCCCTGCGAAACCCTCTTACAGTCAAGTACATACAAGTCTTCTTTAAGACAGTA

      AGAAAGTGCCAGAAAGTGTTTAACTTGTCAAAAAACAAAAACCCAGCAACAGAAAAATGG
2821  ---------+---------+---------+---------+---------+---------+ 2880
      TCTTTCACGGTCTTTCACAAATTGAACAGTTTTTTGTTTTTGGGTCGTTGTCTTTTTACC

      AGTTTGGAAAACAGGACTTAAAATGACATTCAGTATATAAAATATGTACATAATATTGGA
2881  ---------+---------+---------+---------+---------+---------+ 2940
      TCAAACCTTTTGTCCTGAATTTTACTGTAAGTCATATATTTTATACATGTATTATAACCT

      TGACTAACTATCAAATAGATGGATTTGTATCAATACCAAATAGCTTCTGTTTTGTTTTGC
2941  ---------+---------+---------+---------+---------+---------+ 3000
      ACTGATTGATAGTTTATCTACCTAAACATAGTTATGGTTTATCGAAGACAAAACAAAACG

       TGAAGGCTAAATTCACAGCGCTATGCAATTCTTAATTTTCATTAAGTTGTTATTTCAGTT
3001  ---------+---------+---------+---------+---------+---------+ 3060
      ACTTCCGATTTAAGTGTCGCGATACGTTAAGAATTAAAAGTAATTCAACAATAAAGTCAA

      TTAAATGTACCTTCAGAATAAGCTTCCCCACCCCAGTTTTTGTTGCTTGAAAATATTGTT
3061  ---------+---------+---------+---------+---------+---------+ 3120
      AATTTACATGGAAGTCTTATTCGAAGGGGTGGGGTCAAAAACAACGAACTTTTATAACAA

      GTCCCGGATTTTTGTTAATATTCATTTTTGTTATCCTTTTTTAAAAATAAATGTACAGGA
3121  ---------+---------+---------+---------+---------+---------+ 3180
      CAGGGCCTAAAAAGAATTATAAGTAAAAACAATAGGAAAAAATTTTTATTTACATGTCCT

      TGCCAGTAAAAAAAAAAATGGCTTCAGAATTAAAACTATGAAATATTTTACAGTTTTTCT
3181  ---------+---------+---------+---------+---------+---------+ 3240
      ACGGTCATTTTTTTTTTTACCGAAGTCTTAATTTTGATACTTTATAAAATGTCAAAAAGA

      TGTACAGAGTACTTGCTGTTAGCCCAAGGTTAAAAAGTTCATAACAGATTTTTTTTGGAC
3241  ---------+---------+---------+---------+---------+---------+ 3300
      ACATGTCTCATGAACGACAATCGGGTTCCAATTTTTCAAGTATTGTCTAAAAAAAACCTG

      TGTTTTGTTGGGCAGTGCCTGATAAGCTTCAAAGCTGCTTTATTCAATAAAAAAAAAACC
3301  ---------+---------+---------+---------+---------+---------+ 3360
      ACAAAACAACCCGTCACGGACTATTCGAAGTTTCGACGAAATAAGTTATTTTTTTTTTGG

      CGAATTCACTGG
3361  ---------+-- 3372
      GCTTAAGTGACC
```



18 - Proteína, GRB-2, caracterizada por possuir a sequência de aminoácidos:

```
    GCCAGTGAATTCGGGCCCGAATTGGCAGAGCTTAATGGAAAAGACGGCTTCATTCCCAAG
  1 ---------+---------+---------+---------+---------+---------+  60
    CGGTCACTTAAGCCCGGGCTTAACCGTCTCGAATTACCTTTTCTGCCGAAGTAAGGGTTC

    A  S  E  F  G  P  E  L  A  E  L  N  G  K  D  G  F  I  P  K  -

    AACTACATAGAAATGAAACCACATCCGTGGTTTTTTGGCAAAATCCCCAGAGCCAAGGCA
 61 ---------+---------+---------+---------+---------+---------+ 120
    TTGATGTATCTTTACTTTGGTGTAGGCACCAAAAAACCGTTTTAGGGGTCTCGGTTCCGT
                                        Domínio SH2
    N  Y  I  E  M  K  P  H  P  W  F  F  G  K  I  P  R  A  K  A  -

    GAAGAAATGCTTAGCAAACAGCGGCACGATGGGGCCTTTCTTATCCGAGAGAGTGAGAGC
121 ---------+---------+---------+---------+---------+---------+ 180
    CTTCTTTACGAATCGTTTGTCGCCGTGCTACCCCGGAAAGAATAGGCTCTCTCACTCTCG

    E  E  M  L  S  K  Q  R  H  D  G  A  F  L  I  R  E  S  E  S  -

    GCTCCTGGGGACTTCTCCCTCTCTGTCAAGTTTGGAACGATGTGCAGCACTTTCAAGGTG
181 ---------+---------+---------+---------+---------+---------+ 240
    CCACCACCCCTCAACACGGAGAGACAGTTCAAACCTTGCTACACGTCGTGAAAGTTCCAC

    A  P  G  D  F  S  L  S  V  K  F  G  T  M  C  S  T  F  K  V  -

    CTCCCGAGATGGAGCCGGGAAGTACTTCCTCTGGTGGTGAAGTTCAATTCTTTGAATGAG
241 ---------+---------+---------+---------+---------+---------+ 300
    GAGGGCTCTACCTCGGCCCTTCATGAAGGAGACCACCACTTCAAGTTAAGAAACTTACTC

    L  P  R  W  S  R  E  V  L  P  L  V  V  K  F  N  S  L  N  E  -

    CTGGTGGATTATCACAGATCTACATCTGTCTCCAGAAACCAGCAGATATTCCTGCGGGAC
301 ---------+---------+---------+---------+---------+---------+ 360
    GACCACCTAATAGTGTCTAGATGTAGACAGAGGTCTTTGGTCGTCTATAAGGACGCCCTG

    L  V  D  Y  H  R  S  T  S  V  S  R  N  Q  Q  I  F  L  R  D  -


    ATAGAACAGGTGCCACAGCAGCCGACATACGTCCAGGCCCTCTTTGACTTTGATCCCCAG
361 ---------+---------+---------+---------+---------+---------+ 420
    TATCTTGTCCACGGTGTCGTCGGCTGTATGCAGGTCCGGGAGAAACTGAAACTAGGGGTC
                                        Domínio SH3
    I  E  Q  V  P  Q  Q  P  T  Y  V  Q  A  L  F  D  F  D  P  Q  -

    GAGGATGGAGAGCTGGGCTTCCGCCGGGGAGATTTTATCCATGTCATGGATAACTCAGAC
421 ---------+---------+---------+---------+---------+---------+ 480
    CTCCTACCTCTCGACCCGAAGGCGGCCCCTCTAAAATAGGTACAGTACCTATTGAGTCTG

    E  D  G  E  L  G  F  R  R  G  D  F  I  H  V  M  D  N  S  D  -

    CCCAACTGGTGGAAAGGAGCTTGCCACGGGCAGACCGGCATGTTTCCCCGCGAATTATGT
481 ---------+---------+---------+---------+---------+---------+ 540
    GGGTTGACCACCTTTCCTCGAACGGTGCCCGTCTGGCCGTACAAAGGGGCGCTTAATACA

    P  N  W  W  K  G  A  C  H  G  Q  T  G  M  F  P  R  E  L  C  -

    CTCCCCCXGTGAACCGGAACGTCTAAGAGTCAAGAAGCAATTATTTAAAGAAAGTGAAAA
541 ---------+---------+---------+---------+---------+---------+ 600
    GAGGGGGXCACTTGGCCTTGCAGATTCTCAGTTCTTCGTTAATAAATTTCTTTCACTTTT

    L  P  ?  *  T  G  T  S  K  S  Q  E  A  I  I  *  R  K  *  K  -
```



```
    ATGTAAAACACATACAAAAGAATTAAACCCACAAGCTGCCTCTGACAGCAGCCTGTGAGG
601 ---------+---------+---------+---------+---------+---------+ 660
    TACATTTTGTGTATGTTTTCTTAATTTGGGTGTTCGACGGAGACTGTCGTCGGACACTCC

    M * N T Y K R I K P T S C L * Q Q P V R -

    GAGTGCAGAACACCTGGCCGGGTCACCCTGTGACCCTCTCACTTTGGTTGGAACTTTAGG
661 ---------+---------+---------+---------+---------+---------+ 720
    CTCACGTCTTGTGGACCGGCCCAGTGGGACACTGGGAGAGTGAAACCAACCTTGAAATCC

    E C R T P G R V T L * P S H F G W N F R -

    GGGTGGGAGGGGGCGTTGGATTTAAAAATGCCAAAACTTACCTATAAATTAAGAAGAGTT
721 ---------+---------+---------+---------+---------+---------+ 780
    CCCACCCTCCCCCGCAACCTAAATTTTTACGGTTTTGAATGGATATTTAATTCTTCTCAA

    G W E G A L D L K M P K L T Y K L R R V -

    TTTATTACAAATTTTCACTGCTGCTCCTCTTTCCCCTCCTTTGTCTTTTTTTTTCATCCT
781 ---------+---------+---------+---------+---------+---------+ 840
    AAATAATGTTTAAAAGTGACGACGAGGAGAAAGGGGAGGAAACAGAAAAAAAAAAGTAGGA

    F I T N F H C C S S F P S F V F F F H P -

    TTTTTCTCTTCTGTCCATCAGTGCATGACGTTTAAGGCCACGTATAGTCCTAGCTGACGC
841 ---------+---------+---------+---------+---------+---------+ 900
    AAAAAGAGAAGACAGGTAGTCACGTACTGCAAATTCCGGTGCATATCAGGATCGACTGCG

    F F S S V H Q C M T F K A T Y S P S * R -

    CAATAATAAAAACCGAATTCGAGCTCGGATCCGGGGATCCTCTAGAGTC
901 ---------+---------+---------+---------+--------- 949
    GTTATTATTTTTGGCTTAAGCTCGAGCCTAGGCCCCTAGGAGATCTCAG

    Q * * K P N S S S D P G I L * S ? -
```

19 - Molécula de ADN que codifica uma proteína de acordo com a reivindicação 17, ou um seu derivado funcional, caracterizada por, quando a referida molécula de ADN ocorre naturalmente, estar substancialmente isenta das sequências nucleotídicas às quais está nativamente associada.

20 - Molécula de ADN que codifica uma proteína de acordo com a reivindicação 18, ou um seu derivado funcional, caracterizada por, quando a referida molécula de ADN ocorre naturalmente, estar substancialmente isenta das sequências nucleotídicas às quais está nativamente associada.

21 - Molécula de ADN, caracterizada por consistir essencialmente na sequência nucleotídica apresentada na reivindicação 17.

22 - Molécula de ADN, caracterizada por codificar uma proteína de acordo com a reivindicação 18, ou codificar um seu derivado funcional.

23 - Molécula de ADN, caracterizada por codificar uma proteína de acordo com a reivindicação 19, ou codificar um seu derivado funcional.

24 - Molécula de ADN de acordo com a reivindicação 20, caracterizada por ser um veículo de expressão.

25 - Molécula de ADN de acordo com a reivindicação 24, caracterizada por o referido veículo ser um plasmídeo.

26 - Hospedeiro transformado, caracterizado por estar transformado com a molécula da reivindicação 25.

27 - Processo de preparação da proteína da reivindicação 17, substancialmente isenta de outras proteínas às quais está nativamente associada, ou de um seu derivado funcional, caracterizado por compreender:

(a) cultivar uma célula hospedeira capaz de expressar a referida proteína sob condições de cultura;
(b) expressar a referida proteína ou derivado funcional; e
(c) recuperar da referida cultura a referida proteína ou seu derivado funcional.

28 - Processo de preparação da proteína da reivindicação 18, substancialmente isenta de outras proteínas às quais está nativamente associada, ou de um seu derivado funcional, caracterizado por compreender:

(a) cultivar uma célula hospedeira capaz de expressar a referida proteína sob condições de cultura;
(b) expressar a referida proteína ou derivado funcional; e
(c) recuperar, da referida cultura, a referida proteína ou seu derivado funcional.

73 578
SCHLESSINGER 2 PORT

Lisboa,

Por NEW YORK UNIVERSITY
=O AGENTE OFICIAL=

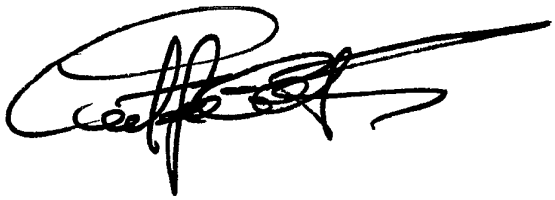

FIG. 1



FIG. 2

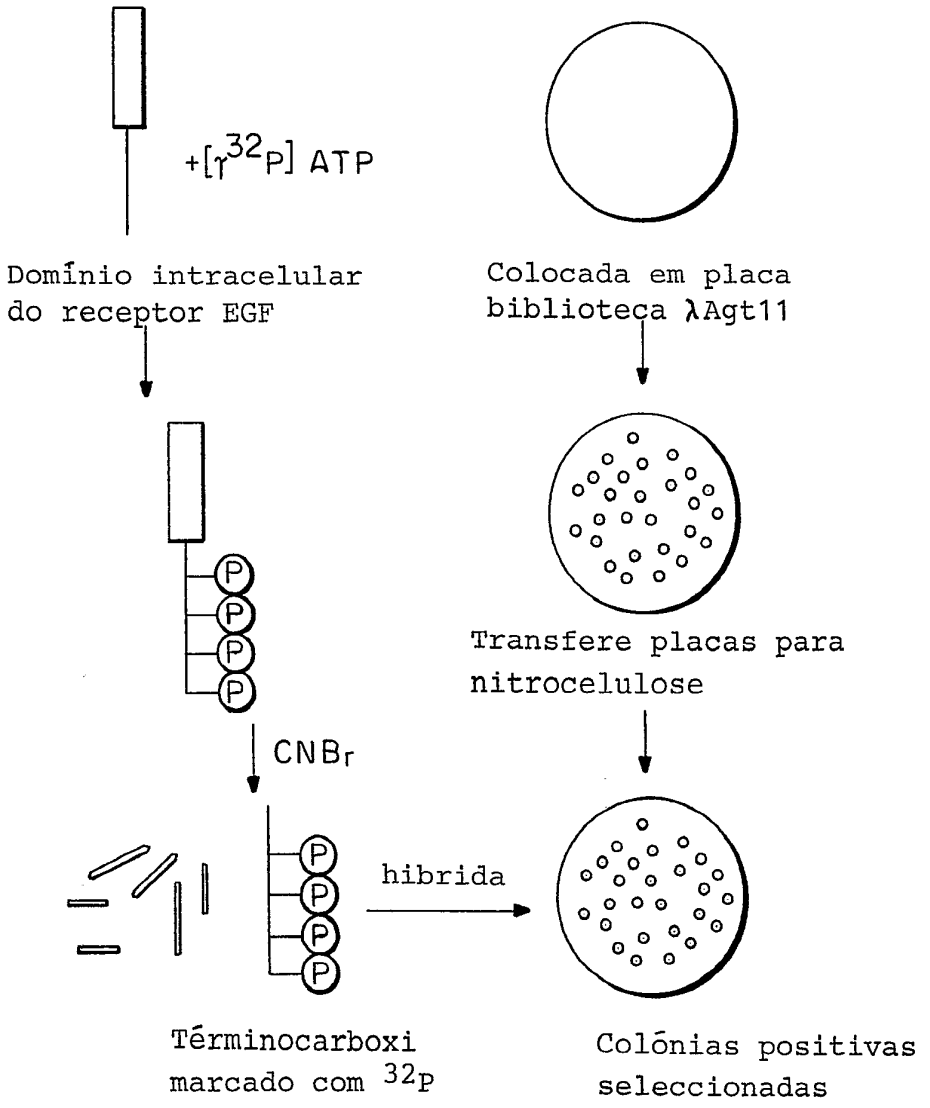
+[γ32P] ATP
Domínio intracelular do receptor EGF
Colocada em placa biblioteca λAgt11
Transfere placas para nitrocelulose
CNBr
hibrida
Términocarboxi marcado com 32P
Colónias positivas seleccionadas

FIG. 3A

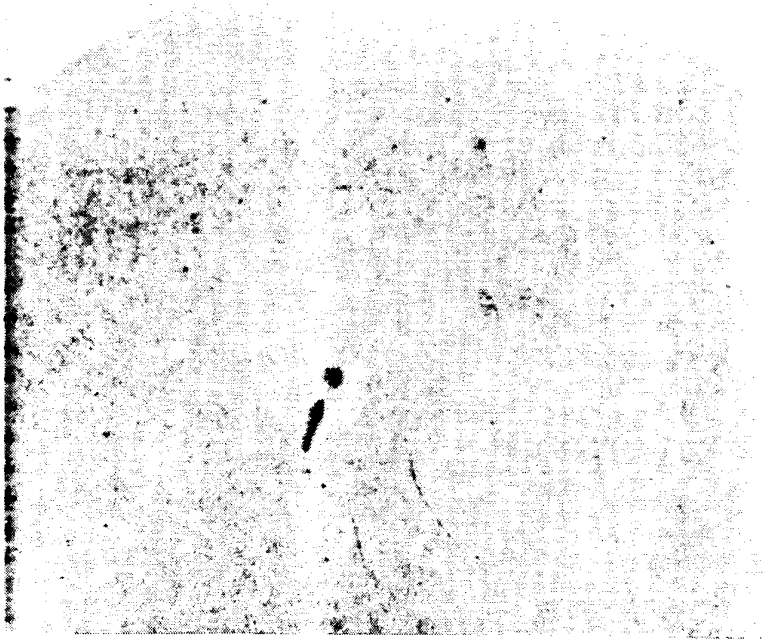

FIG. 3B



FIG. 4A

```
      TACAACCAGGCTCAACTGTTGCATGGTAGCAGATTTGCAAACATGAGTGCTGAGGGGTAC
    1 ---------+---------+---------+---------+---------+---------+  60
      ATGTTGGTCCGAGTTGACAACGTACCATCGTCTAAACGTTTGTACTCACGACTCCCCATG

o                                                 M  S  A  E  G  Y  -

      CAGTACAGAGCGCTGTATGATTATAAAAAGGAAAGAGAAGAAGATATTGACTTGCACTTG
   61 ---------+---------+---------+---------+---------+---------+ 120
      GTCATGTCTCGCGACATACTAATATTTTTCCTTTCTCTTCTTCTATAACTGAACGTGAAC

o     Q  Y  R  A  L  Y  D  Y  K  K  E  R  E  E  D  I  D  L  H  L

      GGTGACATATTGACTGTGAATAAAGGGTCCTTAGTAGCTCTTGGATTCAGTGATGGACAG
  121 ---------+---------+---------+---------+---------+---------+ 180
      CCACTGTATAACTGACACTTATTTCCCAGGAATCATCGAGAACCTAAGTCACTACCTGTC

o     G  D  I  L  T  V  N  K  G  S  L  V  A  L  G  F  S  D  G  Q  -

      GAAGCCAGGCCTGAAGAAATTGGCTGGTTAAATGGCTATAATGAAACCACAGGGGAAAGG
  181 ---------+---------+---------+---------+---------+---------+ 240
      CTTCGGTCCGGACTTCTTTAACCGACCAATTTACCGATATTACTTTGGTGTCCCCTTTCC

o     E  A  R  P  E  E  I  G  W  L  N  G  Y  N  E  T  T  G  E  R

      GGGGACTTTCCGGGAACTTACGTAGAATATATTGGAAGGAAAAAAATCTCGCCTCCCACA
  241 ---------+---------+---------+---------+---------+---------+ 300
      CCCCTGAAAGGCCCTTGAATGCATCTTATATAACCTTCCTTTTTTTAGAGCGGAGGGTGT

o     G  D  F  P  G  T  Y  V  E  Y  I  G  R  K  K  I  S  P  P  T

      CCAAAGCCCCGGCCACCTCGGCCTCTTCCTGTTGCACCAGGTTCTTCGAAAACTGAAGCA
  301 ---------+---------+---------+---------+---------+---------+ 360
      GGTTTCGGGGCCGGTGGAGCCGGAGAAGGACAACGTGGTCCAAGAAGCTTTTGACTTCGT

o     P  K  P  R  P  P  R  P  L  P  V  A  P  G  S  S  K  T  E  A  -
```

FIG. 4B

```
       GATGTTGAACAACAAGCTTTGACTCTCCCGGATCTTGCAGAGCAGTTTGCCCCTCCTGAC
  361  ---------+---------+---------+---------+---------+---------+  420
       CTACAACTTGTTGTTCGAAACTGAGAGGGCCTAGAACGTCTCGTCAAACGGGGAGGACTG

a         D  V  E  Q  Q  A  L  T  L  P  D  L  A  E  Q  F  A  P  P  D  -

       ATTGCCCCGCCTCTTCTTATCAAGCTCGTGGAAGCCATTGAAAAGAAAGGTCTGGAATGT
  421  ---------+---------+---------+---------+---------+---------+  480
       TAACGGGGCGGAGAAGAATAGTTCGAGCACCTTCGGTAACTTTTCTTTCCAGACCTTACA

a         I  A  P  P  L  L  I  K  L  V  E  A  I  E  K  K  G  L  E  C  -

  481  TCAACTCTATACAGAACACAGAGCTCCAGCAACCTGGCAGAATTACGACAGCTTCTTGAT
       ---------+---------+---------+---------+---------+---------+  540
       AGTTGAGATATGTCTTGTGTCTCGAGGTCGTTGGACCGTCTTAATGCTGTCGAAGAACTA

a         S  T  L  Y  R  T  Q  S  S  S  N  L  A  E  L  R  Q  L  L  D  -

  541  TGTGATACACCCTCCGTGGACTTGGAAATGATCGATGTGCACGTTTTGGCTGACGCTTTC
       ---------+---------+---------+---------+---------+---------+  600
       ACACTATGTGGGAGGCACCTGAACCTTTACTAGCTACACGTGCAAAACCGACTGCGAAAG

a         C  D  T  P  S  V  D  L  E  M  I  D  V  H  V  L  A  D  A  F  -

  601  AAACGCTATCTCCTGGACTTACCAAATCCTGTCATTCCAGCAGCCGTTTACAGTGAAATG
       ---------+---------+---------+---------+---------+---------+  660
       TTTGCGATAGAGGACCTGAATGGTTTAGGACAGTAAGGTCGTCGGCAAATGTCACTTTAC

a         K  R  Y  L  L  D  L  P  N  P  V  I  P  A  A  V  Y  S  E  M  -

  661  ATTTCTTTAGCTCCAGAAGTACAAAGCTCCGAAGAATATATTCAGCTATTGAAGAAGCTT
       ---------+---------+---------+---------+---------+---------+  720
       TAAAGAAATCGAGGTCTTCATGTTTCGAGGCTTCTTATATAAGTCGATAACTTCTTCGAA

a         I  S  L  A  P  E  V  Q  S  S  E  E  Y  I  Q  L  L  K  K  L  -
```

FIG. 4C

```
      ATTAGGTCGCCTAGCATACCTCATCAGTATTGGCTTACGCTTCAGTATTTGTTAAAACAT
  721 ---------+---------+---------+---------+---------+---------+ 780
      TAATCCAGCGGATCGTATGGAGTAGTCATAACCGAATGCGAAGTCATAAACAATTTTGTA

a         I  R  S  P  S  I  P  H  Q  Y  W  L  T  L  Q  Y  L  L  K  H

      TTCTTCAAGCTCTCTCAAACCTCCAGCAAAAATCTGTTGAATGCAAGAGTACTCTCTGAA
  781 ---------+---------+---------+---------+---------+---------+ 840
      AAGAAGTTCGAGAGAGTTTGGAGGTCGTTTTTAGACAACTTACGTTCTCATGAGAGACTT

a         F  F  K  L  S  Q  T  S  S  K  N  L  L  N  A  R  V  L  S  E  -

      ATTTTCAGCCCTATGCTTTTCAGATTCTCAGCAGCCAGCTCTGATAATACTGAAAACCTC
  841 ---------+---------+---------+---------+---------+---------+ 900
      TAAAAGTCGGGATACGAAAAGTCTAAGAGTCGTCGGTCGAGACTATTATGACTTTTGGAG

a         I  F  S  P  M  L  F  R  F  S  A  A  S  S  D  N  T  E  N  L  -

      ATAAAAGTTATAGAAATTTTAATCTCAACTGAATGGAATGAACGACAGCCTGCACCAGCA
  901 ---------+---------+---------+---------+---------+---------+ 960
      TATTTTGAATATGTTTAAAATTAGAGTTGACTTACCTTACTTGCTGTCGGACGTGGTCGT

a         I  K  V  I  E  I  L  I  S  T  E  W  N  E  R  Q  P  A  P  A

      CTGCCTCCTAAACCACCAAAACCTACTACTGTAGCCAACAACGGTATGAATAACAATATG
  961 ---------+---------+---------+---------+---------+---------+ 1020
      GACGGAGGATTTGGTGGTTTTGGATGATGACATCGGTTGTTGCCATACTTATTGTTATAC

a         L  P  P  K  P  P  K  P  T  T  V  A  N  N  G  M  N  N  N  M  -

      TCCTTACAAAATGCTGAATGGTACTGGGGAGATATCTCGAGGGAAGAAGTGAATGAAAAA
 1021 ---------+---------+---------+---------+---------+---------+ 1080
      AGGAATGTTTTACGACTTACCATGACCCCTCTATAGAGCTCCCTTCTTCACTTACTTTTT

a         S  L  Q  N  A  E  W  Y  W  G  D  I  S  R  E  E  V  N  E  K  -
```




FIG. 4D

```
      CTTCGAGATACAGCAGACGGGACCTTTTTGGTACGAGATGCGTCTACTAAAATGCATGGT
1081  ---------+---------+---------+---------+---------+---------+
      GAAGCTCTATGTCGTCTGCCCTGGAAAAACCATGCTCTACGCAGATGATTTTACGTACCA 1140

a         L  R  D  T  A  D  G  T  F  L  V  R  D  A  S  T  K  M  H  G  -

      GATTATACTCTTACACTAAGGAAAGGGGGAAATAACAAATTAATCAAAATATTTCATCGA
1141  ---------+---------+---------+---------+---------+---------+ 1200
      CTAATATGAGAATGTGATTCCTTTCCCCCTTTATTGTTTAATTAGTTTTATAAAGTAGCT

a         D  Y  T  L  T  L  R  K  G  G  N  N  K  L  I  K  I  F  H  R  -

      GATGGGAAATATGGCTTCTCTGACCCATTAACCTTCAGTTCTGTGGTTGAATTAATAAAC
1201  ---------+---------+---------+---------+---------+---------+ 1260
      CTACCCTTTATACCGAAGAGACTGGGTAATTGGAAGTCAAGACACCAACTTAATTATTTG

a         D  G  K  Y  G  F  S  D  P  L  T  F  S  S  V  V  E  L  I  N  -

      CACTACCGGAATGAATCTCTAGCTCAGTATAATCCCAAATTGGATGTGAAATTACTTTAT
1261  ---------+---------+---------+---------+---------+---------+ 1320
      GTGATGGCCTTACTTAGAGATCGAGTCATATTAGGGTTTAACCTACACTTTAATGAAATA

a         H  Y  R  N  E  S  L  A  Q  Y  N  P  K  L  D  V  K  L  L  Y  -

1321  CCAGTATCCAAATACCAACAGGATCAAGTTGTCAAAGAAGATAATATTGAAGCTGTAGGG
      ---------+---------+---------+---------+---------+---------+ 1380
      GGTCATAGGTTTATGGTTGTCCTAGTTCAACAGTTTCTTCTATTATAACTTCGACATCCC

a         P  V  S  K  Y  Q  Q  D  Q  V  V  K  E  D  N  I  E  A  V  G  -

      AAAAAATTACATGAATATAACACTCAGTTTCAAGAAAAAAGTCGAGAATATGATAGATTA
1381  ---------+---------+---------+---------+---------+---------+ 1440
      TTTTTTAATGTACTTATATTGTGAGTCAAAGTTCTTTTTTCAGCTCTTATACTATCTAAT

a         K  K  L  H  E  Y  N  T  Q  F  Q  E  K  S  R  E  Y  D  R  I  -
```

FIG. 4E

```
      TATGAAGAATATACCCGCACATCCCAGGAAATCCAAATGAAAAGGACAGCTATTGAAGCA
1441  ---------+---------+---------+---------+---------+---------+ 1500
      ATACTTCTTATATGGGCGTGTAGGGTCCTTTAGGTTTACTTTTCCTGTCGATAACTTCGT

a         Y   E   E   Y   T   R   T   S   Q   E   I   Q   M   K   R   T   A   I   E   A   -

      TTTAATGAAACCATAAAAATATTTGAAGAACAGTGCCAGACCCAAGAGCGGTACAGCAAA
1501  ---------+---------+---------+---------+---------+---------+ 1560
      AAATTACTTTGGTATTTTTATAAACTTCTTGTCACGGTCTGGGTTCTCGCCATGTCGTTT

a         F   N   E   T   I   K   I   F   E   E   Q   C   Q   T   Q   E   R   Y   S   K

      GAATACATAGAAAAGTTTAAACGTGAAGGCAATGAGAAAGAAATACAAAGGATTATGCAT
1561  ---------+---------+---------+---------+---------+---------+ 1620
      CTTATGTATCTTTTCAAATTTGCACTTCCGTTACTCTTTCTTTATGTTTCCTAATACGTA

a         E   Y   I   E   K   F   K   R   E   G   N   E   K   E   I   Q   R   I   M   H   -

      AATTATGATAAGTTGAAGTCTCGAATCAGTGAAATTATTGACAGTAGAAGAAGATTGGAA
1621  ---------+---------+---------+---------+---------+---------+ 1680
      TTAATACTATTCAACTTCAGAGCTTAGTCACTTTAATAACTGTCATCTTCTTCTAACCTT

a         N   Y   D   K   L   K   S   R   I   S   E   I   I   D   S   R   R   R   L   E   -

      GAAGACTTGAAGAAGCAGGCAGCTGAGTATCGAGAAATTGACAAACGTATGAACAGCATT
1681  ---------+---------+---------+---------+---------+---------+ 1740
      CTTCTGAACTTCTTCGTCCGTCGACTCATAGCTCTTTAACTGTTTGCATACTTGTCGTAA

a         E   D   L   K   K   Q   A   A   E   Y   R   E   I   D   K   R   M   N   S   I   -

      AAACCAGACCTTATCCAGCTGAGAAAGACGAGAGACCAATACTTGATGTGGTTGACTCAA
1741  ---------+---------+---------+---------+---------+---------+ 1800
      ATTGGTCTGGAATAGGTCGACTCTTTCTGCTCTCTGGTTATGAACTACACCAACTGAGTT

          K   P   D   L   I   Q   L   R   K   T   R   D   Q   Y   L   M   W   L   T   Q   -
```

FIG. 4F

```
      AAAGGTGTTCGGCAAAAGAAGTTGAACGAGTGGTTGGGCAATGAAAACACTGAAGACCAA
1801  ---------+---------+---------+---------+---------+---------+ 1860
      TTTCCACAAGCCGTTTTCTTCAACTTGCTCACCAACCCGTTACTTTTGTGACTTCTGGTT

a         K   G   V   R   Q   K   K   L   N   E   W   L   G   N   E   N   T   E   D   Q

      TATTCACTGGTGGAAGATGATGAAGATTTGCCCCATCATGATGAGAAGACATGGAATGTT
1861  ---------+---------+---------+---------+---------+---------+ 1920
      ATAAGTGACCACCTTCTACTACTTCTAAACGGGGTAGTACTACTCTTCTGTACCTTACAA

a         Y   S   L   V   E   D   D   E   D   L   P   H   H   D   E   K   T   W   N   V   -

      GGAAGCAGCAACCGAAACAAAGCTGAAAACCTGTTGCGAGGGAAGCGAGATGGCACTTTT
1921  ---------+---------+---------+---------+---------+---------+ 1980
      CCTTCGTCGTTGGCTTTGTTTCGACTTTTGGACAACGCTCCCTTCGCTCTACCGTGAAAA

a         G   S   S   N   R   N   K   A   E   N   L   L   R   G   K   R   D   G   T   F   -

      CTTGTCCGGGAGAGCAGTAAACAGGGCTGCTATGCCTGCTCTGTAGTGGTGGACGGCGAA
1981  ---------+---------+---------+---------+---------+---------+ 2040
      GAACAGGCCCTCTCGTCATTTGTCCCGACGATACGGACGAGACATCACCACCTGCCGCTT

a         L   V   R   E   S   S   K   Q   G   C   Y   A   C   S   V   V   V   D   G   E   -

      GTAAAGCATTGTGTCATAAACAAAACAGCAACTGGCTATGGCTTTGCCGAGCCCTATAAC
2041  ---------+---------+---------+---------+---------+---------+ 2100
      CATTTCGTAACACAGTATTTGTTTTGTCGTTGACCGATACCGAAACGGCTCGGGATATTG

a         V   K   H   C   V   I   N   K   T   A   T   G   Y   G   F   A   E   P   Y   N   -

      TTGTACAGCTCTCTGAAAGAACTGGTGCTACATTACCAACACACCTCCCTTGTGCAGCAC
2101  ---------+---------+---------+---------+---------+---------+ 2160
      AACATGTCGAGAGACTTTCTTGACCACGATGTAATGGTTGTGTGGAGGGAACACGTCGTG

a         L   Y   S   S   L   K   E   L   V   L   H   Y   Q   H   T   S   L   V   Q   H   -
```

FIG. 4G

```
      AACGACTCCCTCAATGTCACACTAGCCTACCCAGTATATGCACAGCAGAGGCGATGAAGC
2161  ---------+---------+---------+---------+---------+---------+ 2220
      TTGCTGAGGGAGTTACAGTGTGATCGGATGGGTCATATACGTGTCGTCTCCGCTACTTCG

a         N   D   S   L   N   V   T   L   A   Y   P   V   Y   A   Q   Q   R   R       -

      GCTTACTCTTTGATCCTTCTCCTGAAGTTCAGCCACCCTGAGGCCTCTGGAAAGCAAAGG
2221  ---------+---------+---------+---------+---------+---------+ 2280
      CGAATGAGAAACTAGGAAGAGGACTTCAAGTCGGTGGGACTCCGGAGACCTTTCGTTTCC
a

      GCTCCTCTCCAGTCTGATCTGTGAATTGAGCTGCAGAAACGAAGCCATCTTTCTTTGGAT
2281  ---------+---------+---------+---------+---------+---------+ 2340
      CGAGGAGAGGTCAGACTAGACACTTAACTCGACGTCTTTGCTTCGGTAGAAAGAAACCTA
a

      GGGACTAGAGCTTTCTTTCACAAAAAAGAAGTAGGGGAAGACATGCAGCCTAAGGCTGTA
2341  ---------+---------+---------+---------+---------+---------+ 2400
      CCCTGATCTCGAAAGAAAGTGTTTTTTCTTCATCCCCTTCTGTACGTCGGATTCCGACAT
a

      TGATGACCACACGTTCCTAAGCTGGAGTGCTTATCCCTTCTTTTTCTTTTTTTCTTTGGT
2401  ---------+---------+---------+---------+---------+---------+ 2460
      ACTACTGGTGTGCAAGGATTCGACCTCACGAATAGGGAAGAAAAAGAAAAAAAGAAACCA
a

      TTAATTTAAAGCCACAACCACATACAACACAAAGAGAAAAAGAAATGCAAAAATCTCTGC
2461  ---------+---------+---------+---------+---------+---------+ 2520
      AATTAAATTTCGGTGTTGGTGTATGTTGTGTTTCTCTTTTTCTTTACGTTTTTAGAGACG
a

      GTGCAGGGACAAAGAGGCCTTTAACCATGGTGCTTGTTAATGCTTTCTGAAGCTTTACCA
2521  ---------+---------+---------+---------+---------+---------+ 2580
      CACGTCCCTGTTTCTCCGGAAATTGGTACCACGAACAATTACGAAAGACTTCGAAATGGT
```

FIG. 4H

```
          GCTGAAAGTTGGGACTCTGGAGAGCGGAGGAGAGAGAGGCAGAAGAACCCTGGCCTGAGA
2581      ---------+---------+---------+---------+---------+---------+ 2640
          CGACTTTCAACCCTGAGACCTCTCGCCTCCTCTCTCTCCGTCTTCTTGGGACCGGACTCT
a

          AGGTTTGGTCCAGCCTGGTTTAGCCTGGATGTTGCTGTGCACGGTGGACCCAGACACATC
2641      ---------+---------+---------+---------+---------+---------+ 2700
          TCCAAACCAGGTCGGACCAAATCGGACCTACAACGACACGTGCCACCTGGGTCTGTGTAG
a

          GCACTGTGGATTATTTCATTTTGTAACAAATGAACGATATGTAGCAGAAAGGCACGTCCA
2701      ---------+---------+---------+---------+---------+---------+ 2760
          CGTGACACCTAATAAAGTAAAACATTGTTTACTTGCTATACATCGTCTTTCCGTGCAGGT
a

          CTCACAAGGGACGCTTTGGGAGAATGTCAGTTCATGTATGTTCAGAAGAAATTCTGTCAT
2761       ---------+---------+---------+---------+---------+---------+ 2820
          GAGTGTTCCCTGCGAAACCCTCTTACAGTCAAGTACATACAAGTCTTCTTTAAGACAGTA
a

          AGAAAGTGCCAGAAAGTGTTTAACTTGTCAAAAAACAAAAACCCAGCAACAGAAAAATGG
2821      ---------+---------+---------+---------+---------+---------+ 2880
          TCTTTCACGGTCTTTCACAAATTGAACAGTTTTTTGTTTTTGGGTCGTTGTCTTTTTACC
a

          AGTTTGGAAAACAGGACTTAAAATGACATTCAGTATATAAAATATGTACATAATATTGGA
2881      ---------+---------+---------+---------+---------+---------+ 2940
          TCAAACCTTTTGTCCTGAATTTTACTGTAAGTCATATATTTTATACATGTATTATAACCT
a

          TGACTAACTATCAAATAGATGGATTTGTATCAATACCAAATAGCTTCTGTTTTGTTTTGC
2941      ---------+---------+---------+---------+---------+---------+ 3000
          ACTGATTGATAGTTTATCTACCTAAACATAGTTATGGTTTATCGAAGACAAAACAAAACG
```

FIG. 4I

```
      TGAAGGCTAAATTCACAGCGCTATGCAATTCTTAATTTTCATTAAGTTGTTATTTCAGTT
3001  ---------+---------+---------+---------+---------+---------+ 3060
      ACTTCCGATTTAAGTGTCGCGATACGTTAAGAATTAAAAGTAATTCAACAATAAAGTCAA
a

      TTAAATGTACCTTCAGAATAAGCTTCCCCACCCCAGTTTTTGTTGCTTGAAAATATTGTT
3061  ---------+---------+---------+---------+---------+---------+ 3120
      AATTTACATGGAAGTCTTATTCGAAGGGGTGGGGTCAAAAACAACGAACTTTTATAACAA
a

      GTCCCGGATTTTTGTTAATATTCATTTTTGTTATCCTTTTTTAAAAATAAATGTACAGGA
3121  ---------+---------+---------+---------+---------+---------+ 3180
      CAGGGCCTAAAAAGAATTATAAGTAAAAACAATAGGAAAAAATTTTTATTTACATGTCCT
a

      TGCCAGTAAAAAAAAAAATGGCTTCAGAATTAAAACTATGAAATATTTTACAGTTTTTCT
3181  ---------+---------+---------+---------+---------+---------+ 3240
      ACGGTCATTTTTTTTTTTTACCGAAGTCTTAATTTTGATACTTTATAAAATGTCAAAAAGA
a

      TGTACAGAGTACTTGCTGTTAGCCCAAGGTTAAAAAGTTCATAACAGATTTTTTTTGGAC
3241  ---------+---------+---------+---------+---------+---------+ 3300
      ACATGTCTCATGAACGACAATCGGGTTCCAATTTTTCAAGTATTGTCTAAAAAAAACCTG
a

      TGTTTTGTTGGGCAGTGCCTGATAAGCTTCAAAGCTGCTTTATTCAATAAAAAAAAAACC
3301  ---------+---------+---------+---------+---------+---------+ 3360
      ACAAAACAACCCGTCACGGACTATTCGAAGTTTCGACGAAATAAGTTATTTTTTTTTTGG
a

      CGAATTCACTGG
3361  ---------+-- 3372
      GCTTAAGTGACC
```




FIG. 5

```
GRB-1 N 333  WYWGDIS -- R EE--- VN E- KL RDTAD------ GTFLVRDST KM HGDY T LT LRK-----GG--NN LIKI
GRB-1 C 624  WNVGSSN -- R NK--- AE N- LL RGKRD------ GTFLVRESS K- QGCY A CS VVV-----DG--EV KHCV

c-src   150  WYFGKIT -- R RE--- SE RL LL NPENPR----- GTFLVRESE TT KGAY C LS VSDF-DNAKGLNVK HYKI
v-abl   248  WYHGPVS -- R NA--- AE YK KS SGIN------- GSFLVRESE SS PG-Q R -S ISLRYE---G-RVY HYRI
PLC N   550  WFHGKLG AG R DGRHI AE R- LL TEYCIETGAPD GSFLVRESE TF VGDY T LS F--WR-N--G-KVQ HCRI
PLC C   668  WYHASLT -- R AQ--- AE H- ML MRVPRD----- GAFLVRKRN -E PNSY A IS F--RAE---G-KIK HCRV
GAP N   178  WYHGKLD -- R TI--- AE E- RL RQAGKS----- GSYLIRESD RR PGSF V LS FRSQM-N-V---VN HPRI
GAP C   348  WYHGKIS -- K QE--- AY N- LL MTVGQVC---- -SFLVRPSD NT PGDY S LY F-RTNENIQ--R-- -FKI
v-crk   248  WYWGRLS -- R GD--- AV S- LL QGQRH------ GTFLVRDSG SI PGDF V LS VSES---S---RVS HYIV

GRB-1 N 384   --FHRD G KYGFSDPLT---------------- F S SV V ELI N HY RNESLAQYNPKLDV- KL LY PVSK
GRB-1 C 672   INKTAT G -YGFAEPYNL--------------- Y S SL K ELV L HY QHTSLVQHNDSLNV- TL AY PVYA

c-src   207   RKLDSG G FYITSRTQ----------------- F S SL Q QLV A YY SKHADGLCH------ RL TN -VCP
v-abl   298   -NTASD G KLYVSSESR---------------- F N TL A ELV H HH STVADGLITT----- LH YP --AP
PLC N   611   HSRQDA G TPKFFLTDNLV-------------- F D SL Y DLI T HY QQVPLRCA-EFEM-- RL SL PV-P
PLC C   718   ---QQE G QTVMLGNSE---------------- F D SL V DLI S YY EKHPLYRK----M-- KL RY PI--
GAP N   230   --IAMC G DYYIGGRR----------------- F S SL S DLI G YY SHVSCLLKGE----- KL LY PVAP
GAP C   399   -CPTPN N QFMMGGRY----------------- Y N SI G DII D HY RKEQIVEG-YY---- -L KE PV-P
v-crk   298   NSLGPA G GRRAGGEGPFAPGLNPTRFLIGDNV F D SL P SLL E FY KIHYLDT--TT---- -L IE PV--

GRB-1   10   ALYDY KKEREE D IDLHLGDI LT VNK G SLVALGFSDPEARPEDIG WL NGYNETTGER GDFP GT YVE YIGRK

c-src   88   ALYDY ESRTET D -------- LA FKK G ERLQIV------MNTEGD WW LAHSLTTGQT GYIP SN YVA PS-DS
v-abl   68   ALYDF VASGDN T -------- LS ITK G EKLRVLG-----YNHNGE WC EAQTK-NGQ- GWVP SN YIT PV-NS
PLC    148   ALFDY KAGRED E -------- LT FTK S AIIQNV------EKQEGG WW RGDYHHKKQ- LWFP SN YVE EMV-S
GAP    284   AILDY TKVPDT D E------- IS FLK G DMFIVN------NELEDG WM WVTNLRTDEQ GLIV ED LVE EV-GR
v-crk  375   ALFDF KGNDDG D -------- LP FKK G DILKIR------DKPEEQ WW NAEDMDGKR- GMIP VP YVE KCRPS
```

FIG. 6

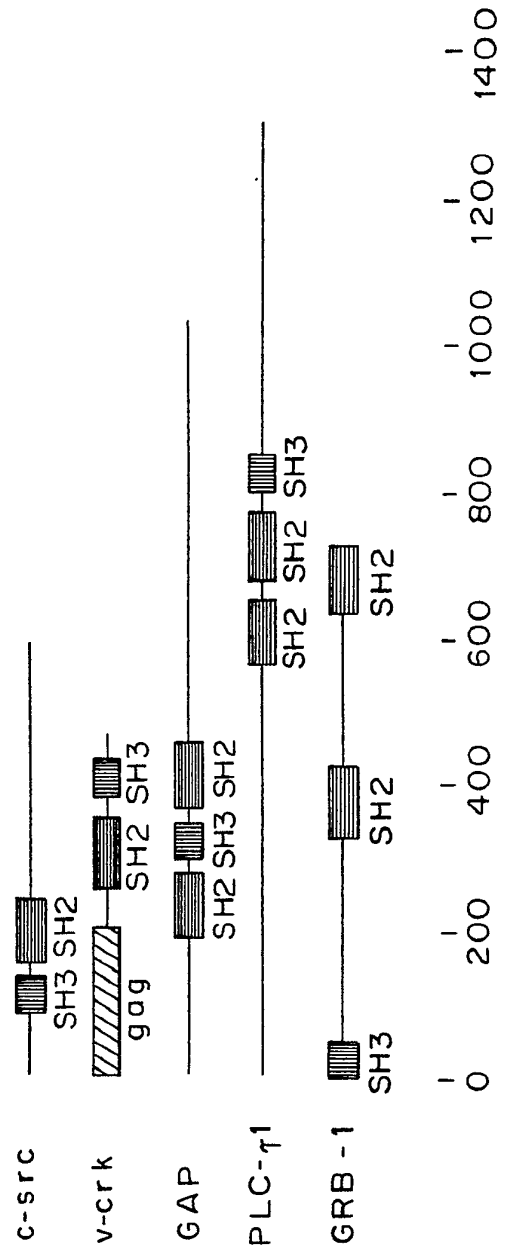
c-src
SH3 SH2
v-crk
gag
SH2 SH3
GAP
SH2 SH3 SH2
PLC-γ1
SH2 SH2 SH3
GRB-1
SH3
SH2
SH2
0
200
400
600
800
1000
1200
1400

FIG. 7

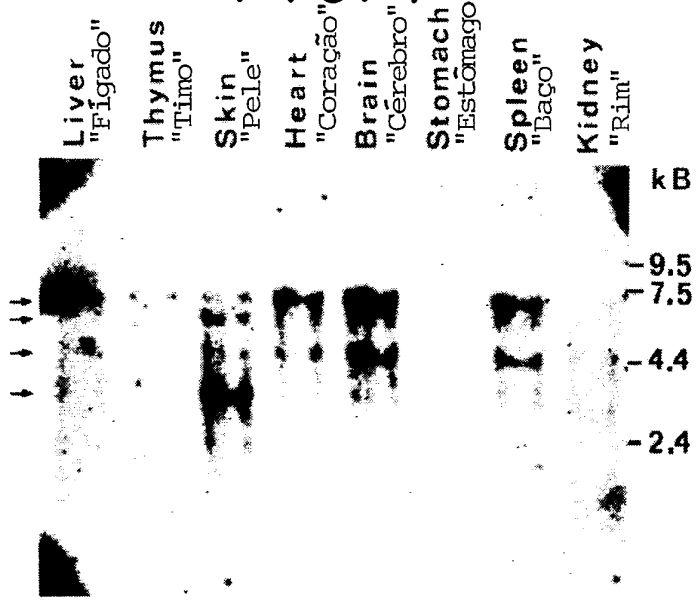
Liver
"Fígado"
Thymus
"Timo"
Skin
"Pele"
Heart
"Coração"
Brain
"Cérebro"
Stomach
"Estômago"
Spleen
"Baço"
Kidney
"Rim"
kB
9.5
7.5
4.4
2.4

FIG. 8

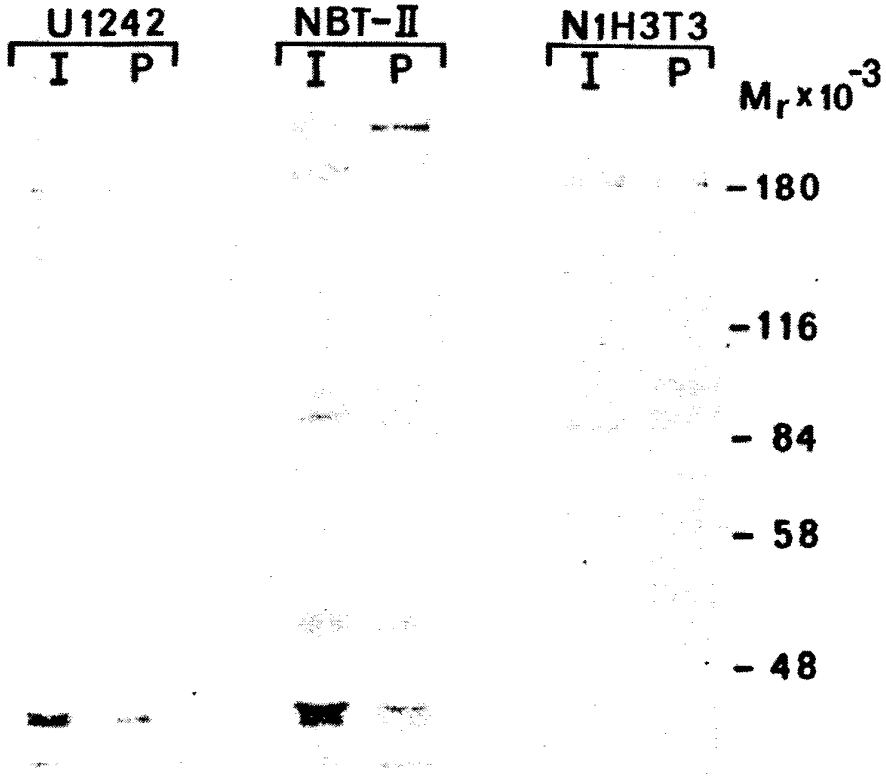
U1242
I P
NBT-II
I P
N1H3T3
I P
Mr x 10-3
180
116
84
58
48

FIG. 9A

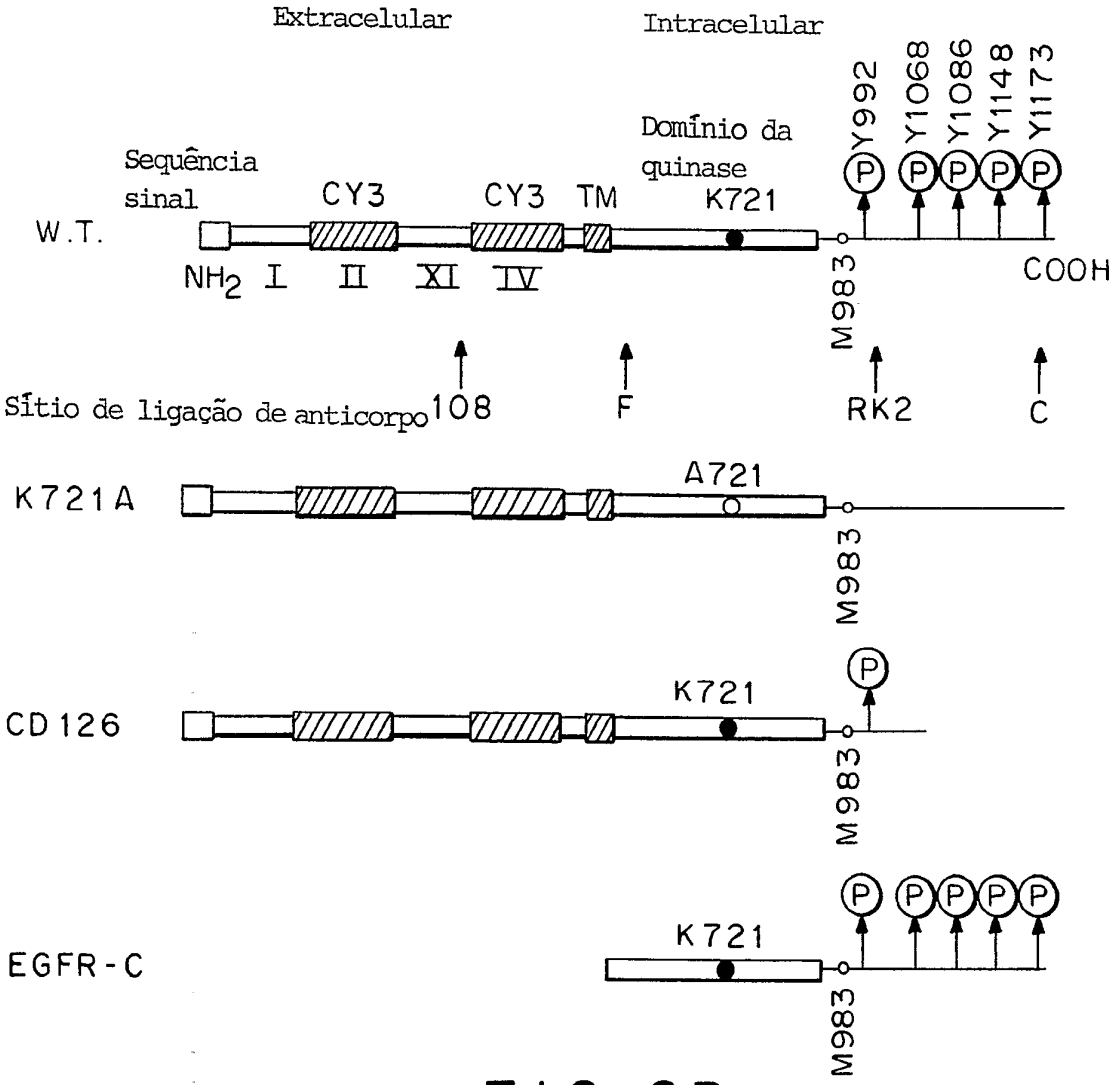
Extracelular
Intracelular
Domínio da quinase
Sequência sinal
CY3
CY3
TM
K721
Y992
Y1068
Y1086
Y1148
Y1173
W.T.
NH2
I
II
XI
IV
M983
COOH
Sítio de ligação de anticorpo
108
F
RK2
C
K721A
A721
M983
CD126
K721
M983
EGFR-C
K721
M983

FIG. 9B

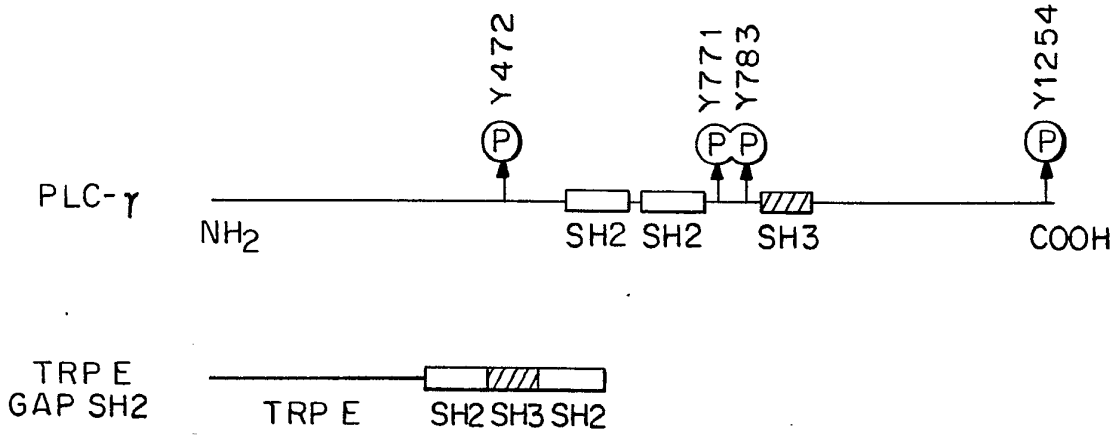
Y472
Y771
Y783
Y1254
PLC-γ
NH2
SH2
SH2
SH3
COOH
TRP E
GAP SH2
TRP E
SH2
SH3
SH2

FIG. 10

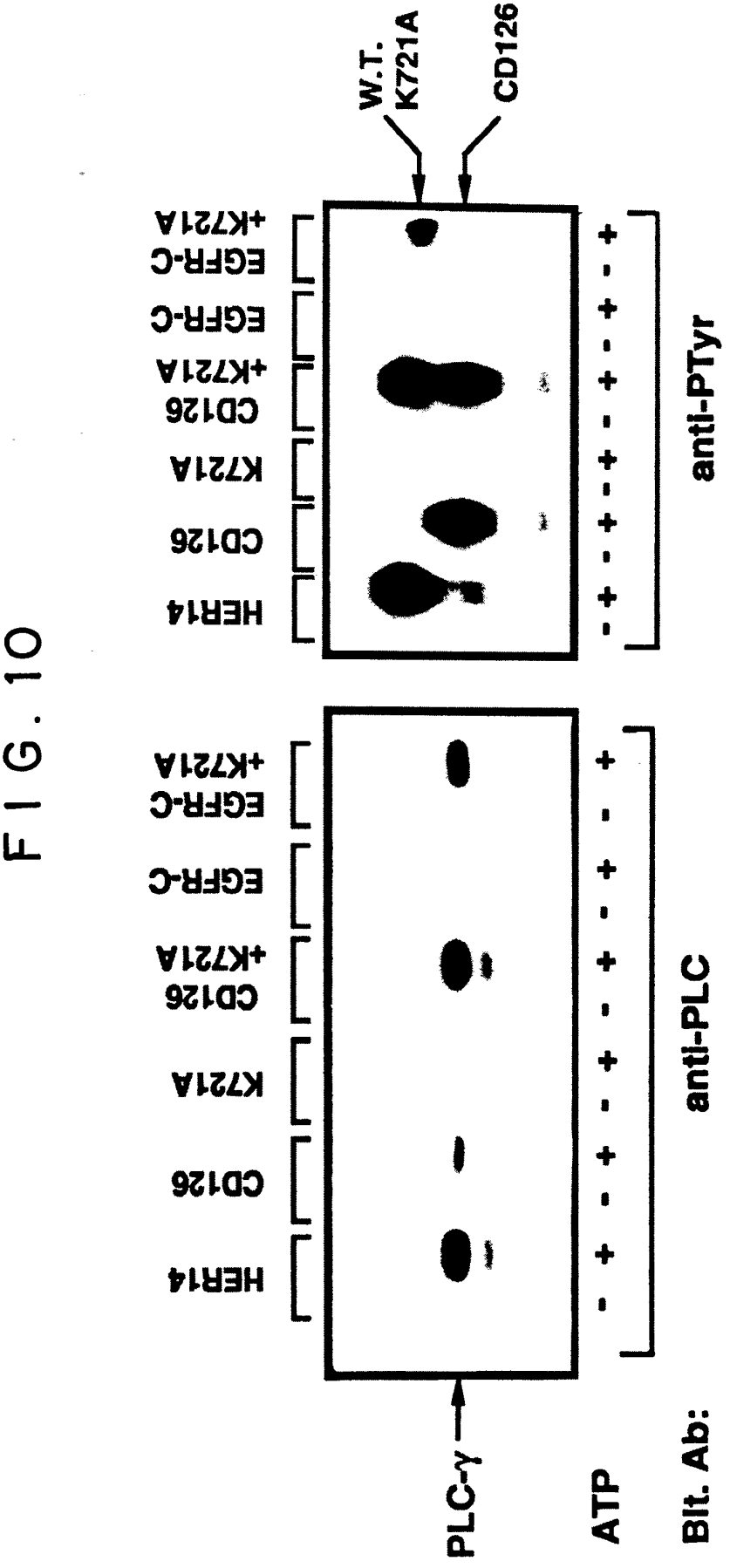
W.T.
K721A
CD126
EGFR-C +K721A
EGFR-C
CD126 +K721A
K721A
CD126
HER14
anti-PTyr
EGFR-C +K721A
EGFR-C
CD126 +K721A
K721A
CD126
HER14
anti-PLC
PLC-γ
ATP
Blt. Ab:

FIG. 11

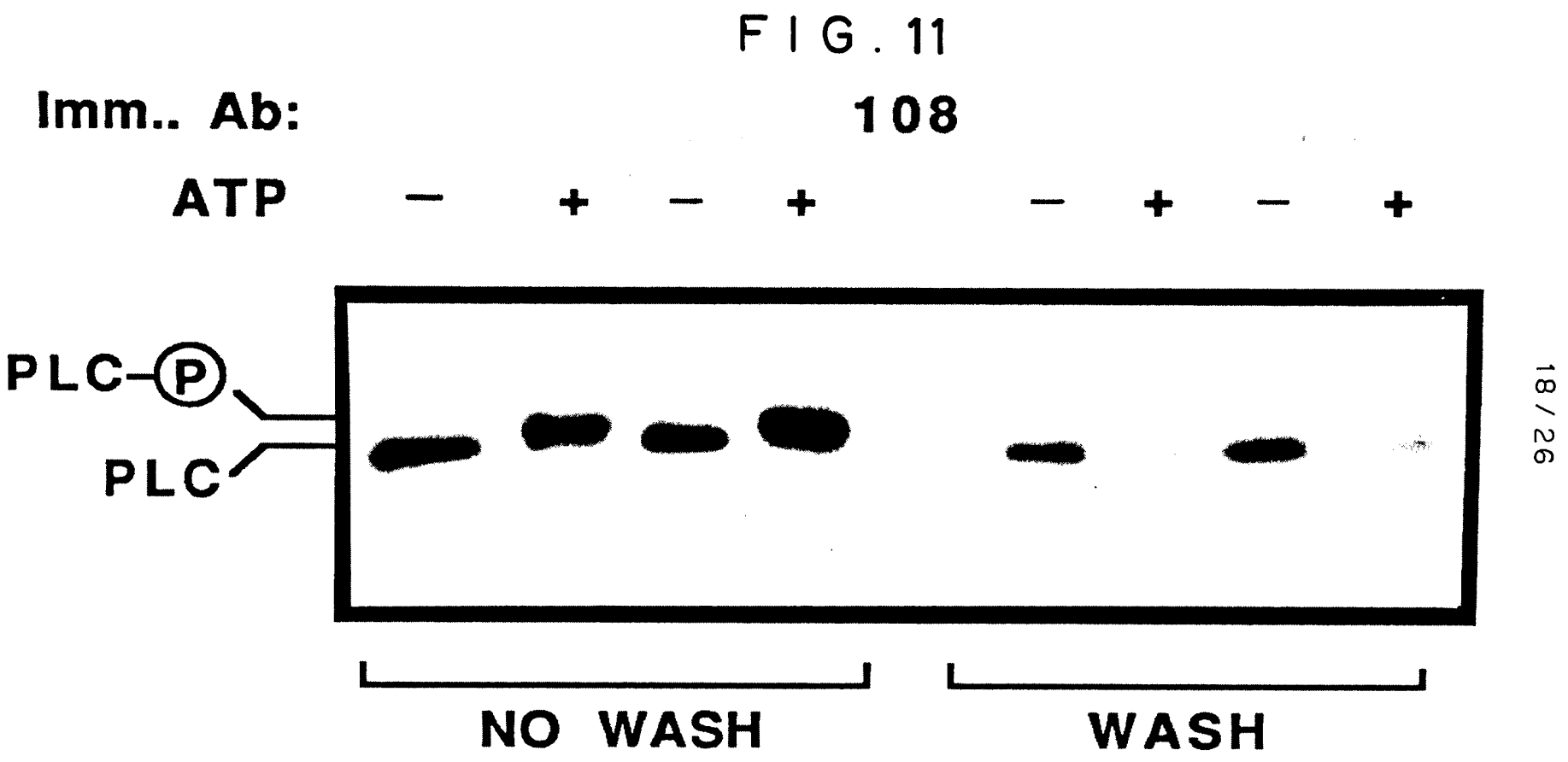
Imm.. Ab:
108
ATP
– + – + – + – +
PLC-P
PLC
NO WASH
"Sem lavagem"
WASH
"Com lavagem"
Blt. Ab:
PLC

FIG. 12A

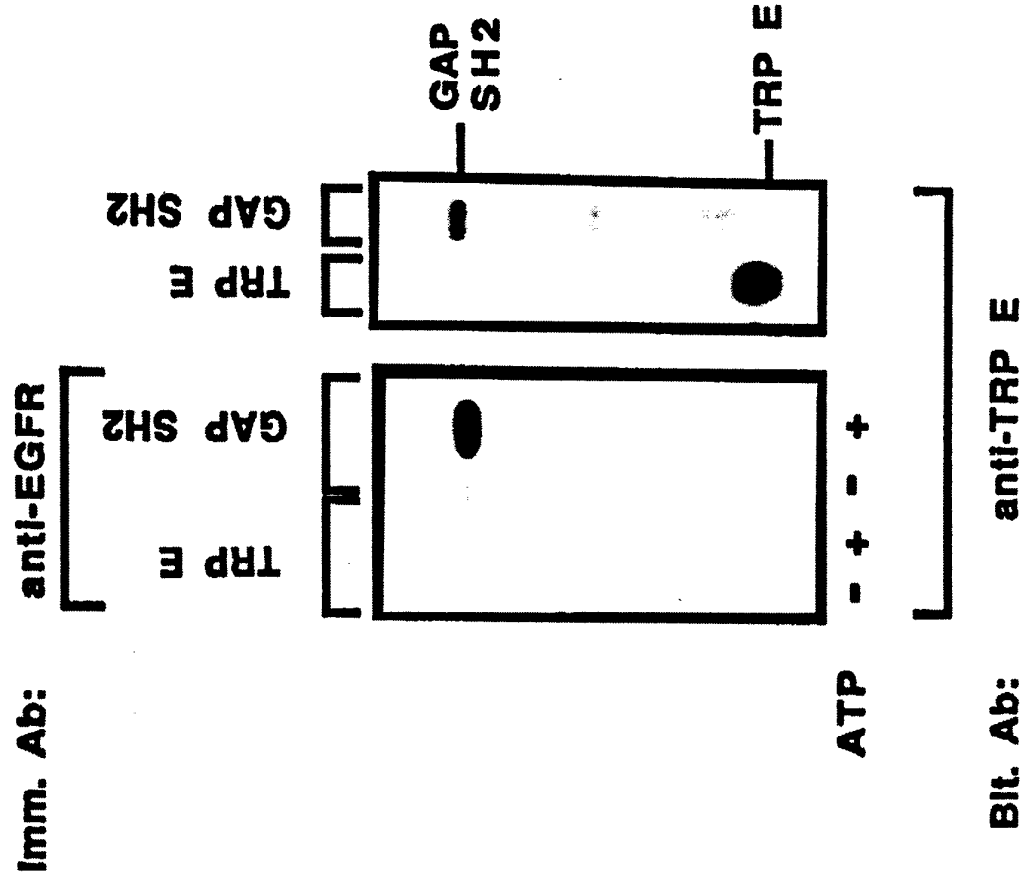

FIG. 12B

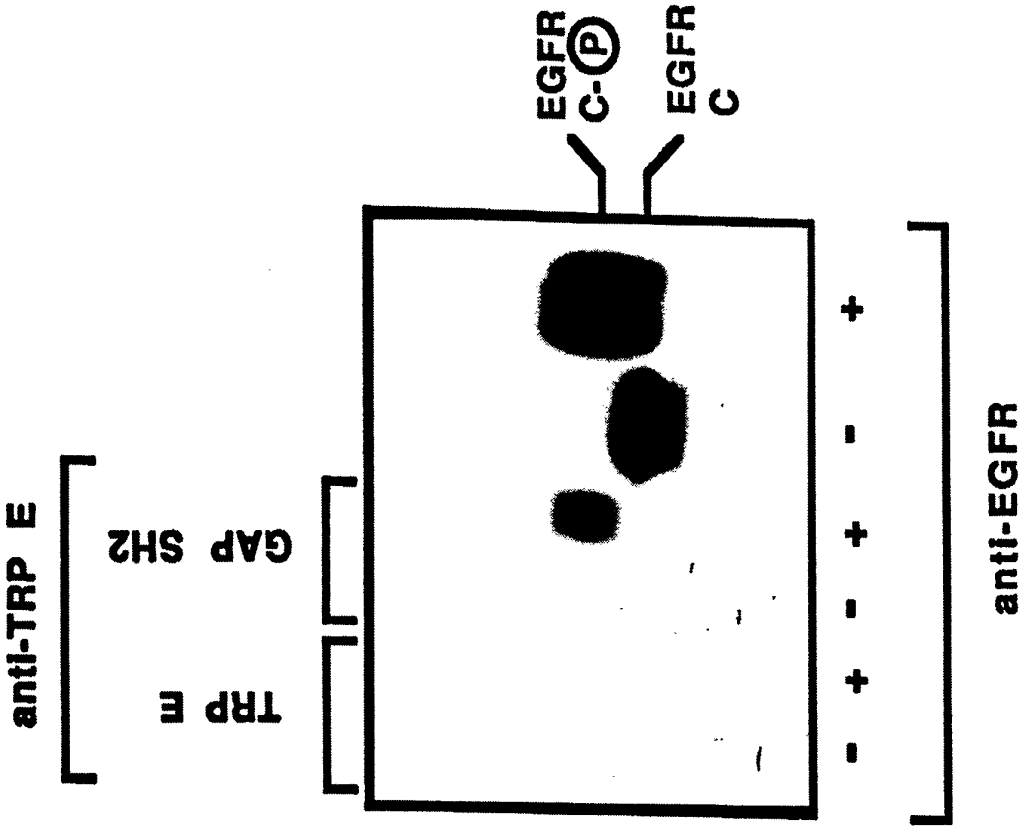

FIG. 13 A

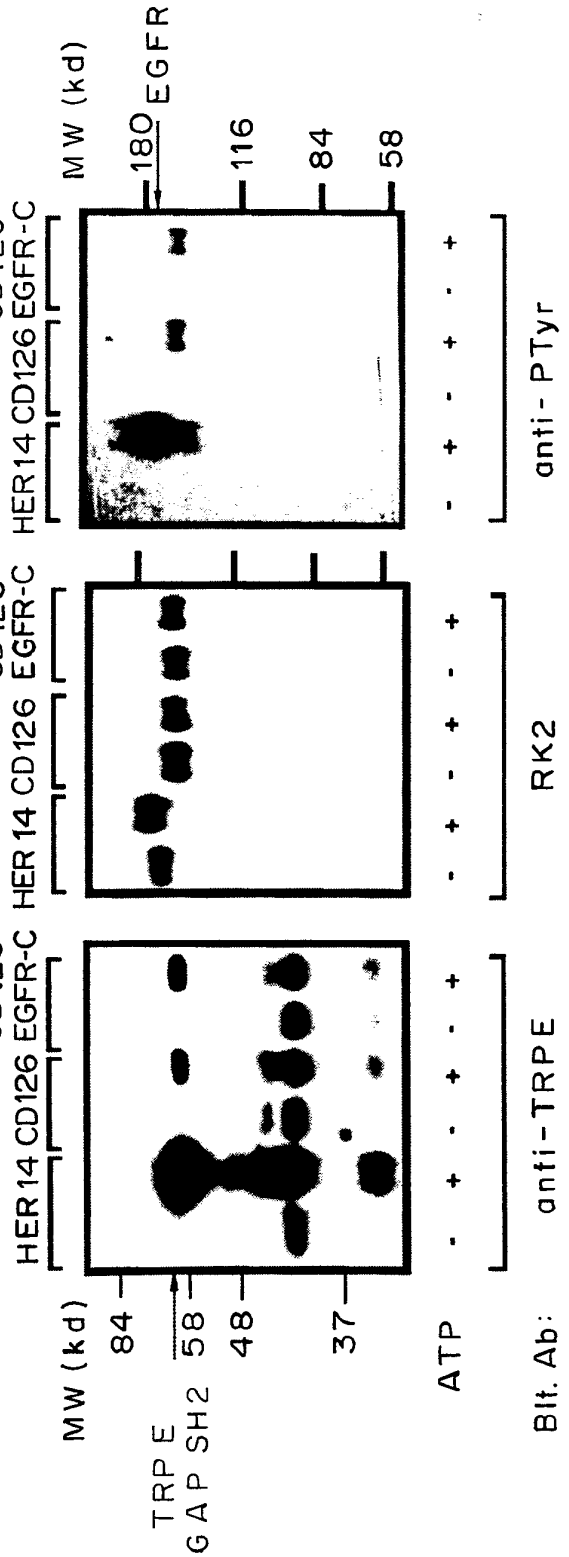
MW (kd)
84
58
48
37
TRPE GAP SH2
HER14 CD126 CD126+ EGFR-C
ATP - + - + - +
Blt. Ab: anti-TRPE
HER14 CD126 CD126+ EGFR-C
- + - + - +
RK2
MW (kd)
180 EGFR
116
84
58
HER14 CD126 CD126+ EGFR-C
- + - + - +
anti-PTyr

FIG.13B

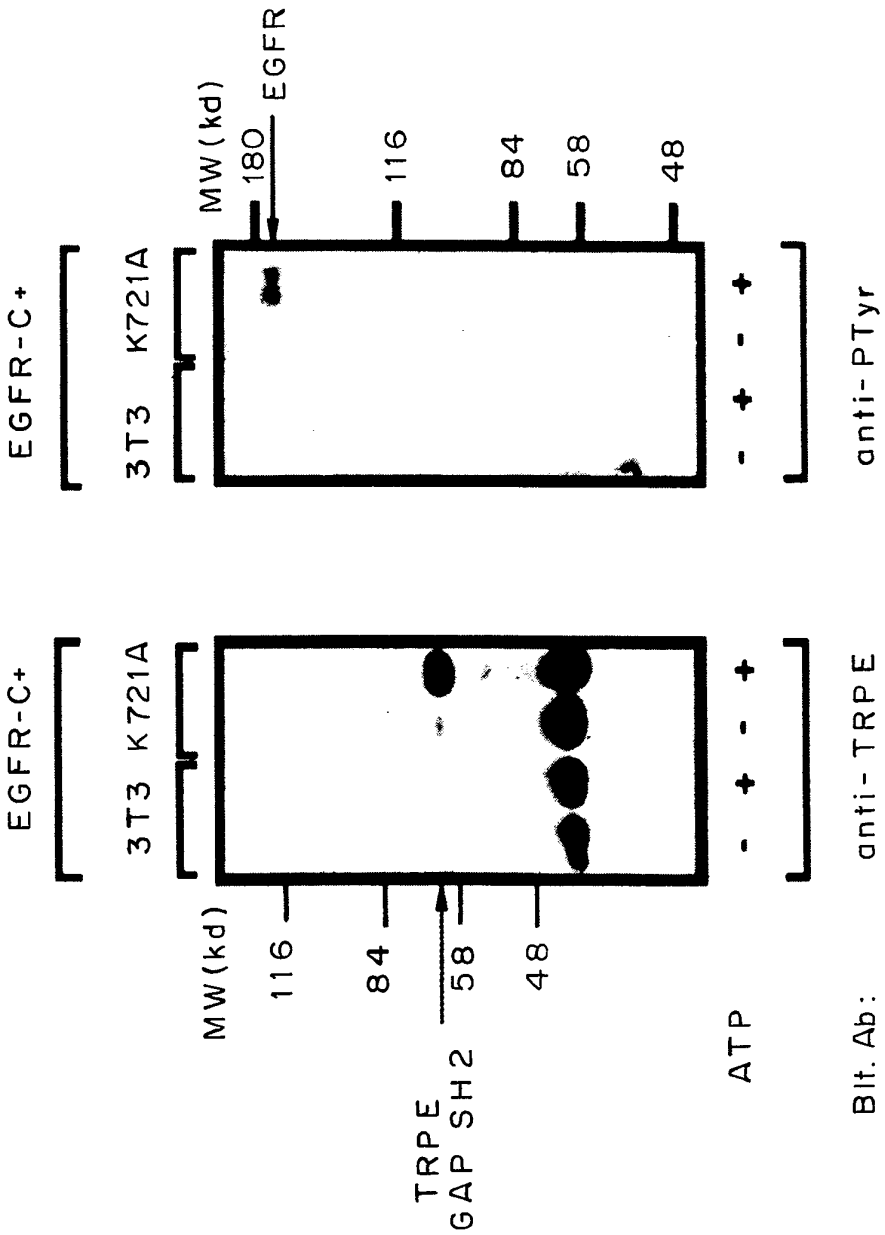
EGFR-C+
3T3 K721A
MW(kd)
180
EGFR
116
84
58
48
- + - +
anti-PTyr
EGFR-C+
3T3 K721A
MW(kd)
116
84
58
48
TRPE
GAP SH2
- + - +
ATP
Blt. Ab:
anti-TRPE

FIG. 14

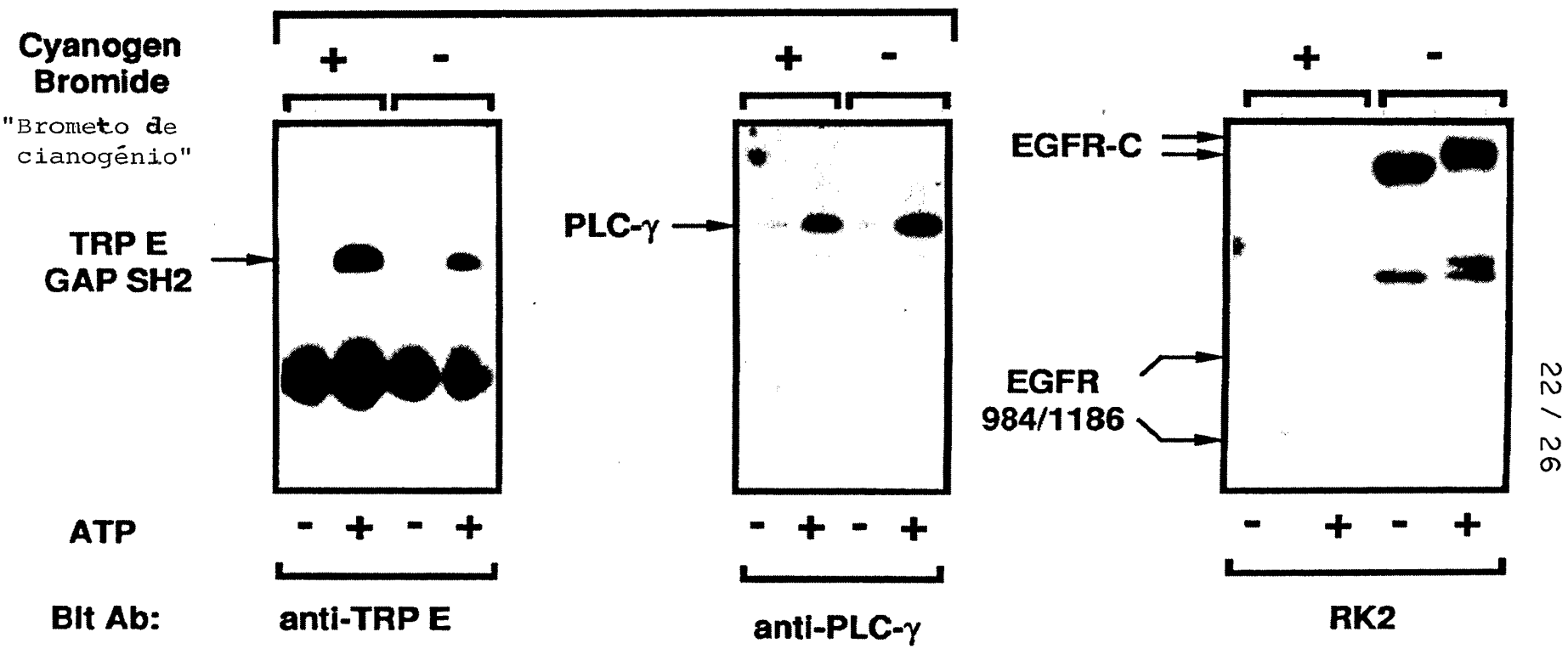
Imm Ab:
c
Cyanogen Bromide
"Brometo de cianogénio"
+ -
+ -
+ -
TRP E
GAP SH2
PLC-γ
EGFR-C
EGFR
984/1186
ATP
- + - +
- + - +
- + - +
Blt Ab:
anti-TRP E
anti-PLC-γ
RK2

FIG. 15

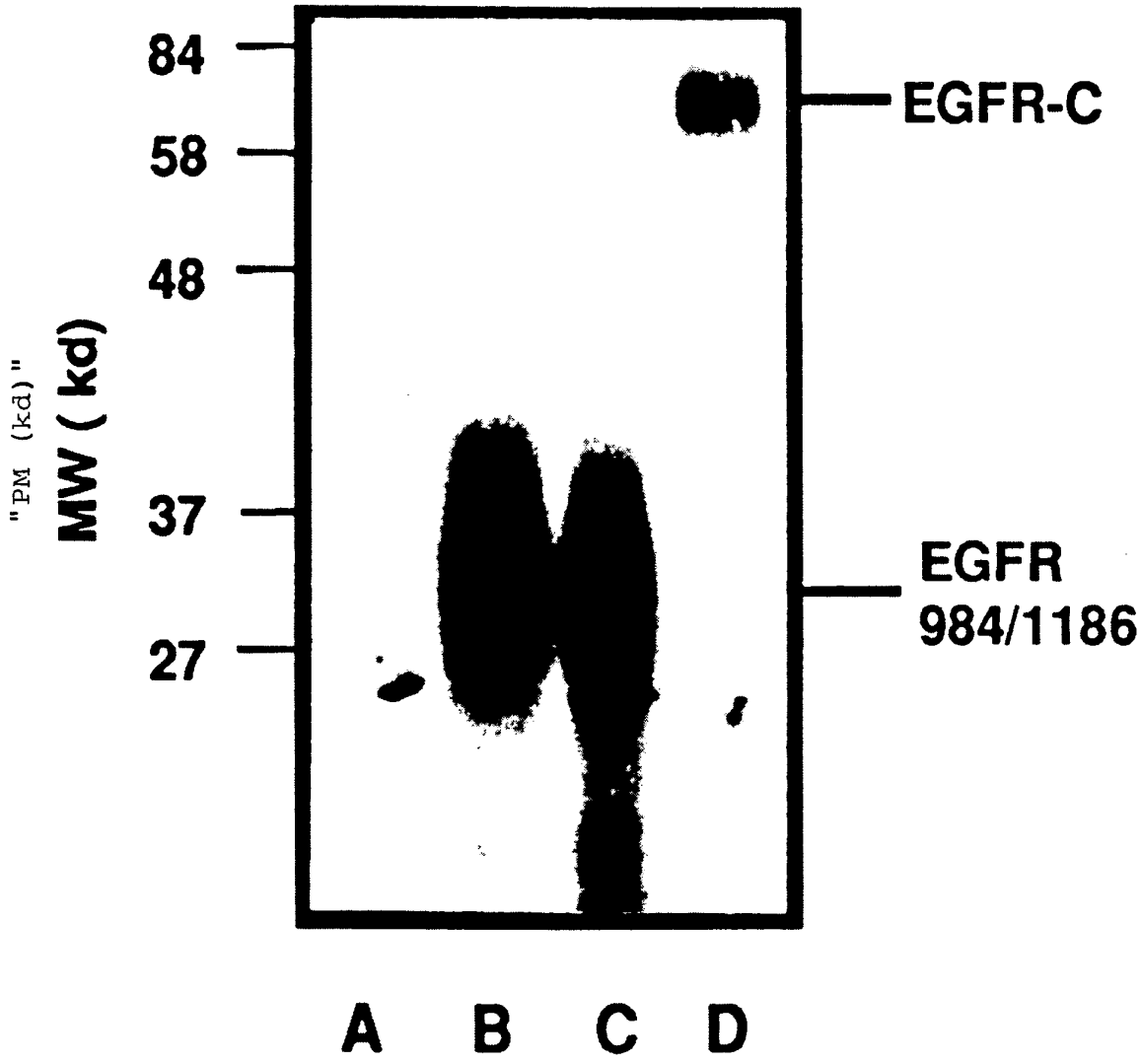

FIG. 16 A

```
     GCCAGTGAATTCGGGCCCGAATTGGCAGAGCTTAATGGAAAAGACGGCTTCATTCCCAAG
   1 ---------+---------+---------+---------+---------+---------+  60
     CGGTCACTTAAGCCCGGGCTTAACCGTCTCGAATTACCTTTTCTGCCGAAGTAAGGGTTC

a    A  S  E  F  G  P  E  L  A  E  L  N  G  K  D  G  F  I  P  K  -

     AACTACATAGAAATGAAACCACATCCGTGGTTTTTTGGCAAAATCCCCAGAGCCAAGGCA
  61 ---------+---------+---------+---------+---------+---------+ 120
     TTGATGTATCTTTACTTTGGTGTAGGCACCAAAAAACCGTTTTAGGGGTCTCGGTTCCGT
                                        Domínio SH2
a    N  Y  I  E  M  K  P  H  P  W  F  F  G  K  I  P  R  A  K  A  -

     GAAGAAATGCTTAGCAAACAGCGGCACGATGGGGCCTTTCTTATCCGAGAGAGTGAGAGC
 121 ---------+---------+---------+---------+---------+---------+ 180
     CTTCTTTACGAATCGTTTGTCGCCGTGCTACCCCGGAAAGAATAGGCTCTCTCACTCTCG

a    E  E  M  L  S  K  Q  R  H  D  G  A  F  L  I  R  E  S  E  S  -

     GCTCCTGGGGACTTCTCCCTCTCTGTCAAGTTTGGAACGATGTGCAGCACTTTCAAGGTG
 181 ---------+---------+---------+---------+---------+---------+ 240
     CCACCACCCCTCAACACGGAGAGACAGTTCAAACCTTGCTACACGTCGTGAAAGTTCCAC

a    A  P  G  D  F  S  L  S  V  K  F  G  T  M  C  S  T  F  K  V  -

     CTCCCGAGATGGAGCCGGGAAGTACTTCCTCTGGTGGTGAAGTTCAATTCTTTGAATGAG
 241 ---------+---------+---------+---------+---------+---------+ 300
     GAGGGCTCTACCTCGGCCCTTCATGAAGGAGACCACCACTTCAAGTTAAGAAACTTACTC

a    L  P  R  W  S  R  E  V  L  P  L  V  V  K  F  N  S  L  N  E  -
```

FIG. 16B

```
    CTGGTGGATTATCACAGATCTACATCTGTCTCCAGAAACCAGCAGATATTCCTGCGGGAC
301 ---------+---------+---------+---------+---------+---------+ 360
    GACCACCTAATAGTGTCTAGATGTAGACAGAGGTCTTTGGTCGTCTATAAGGACGCCCTG

a     L  V  D  Y  H  R  S  T  S  V  S  R  N  Q  Q  I  F  L  R  D  -


    ATAGAACAGGTGCCACAGCAGCCGACATACGTCCAGGCCCTCTTTGACTTTGATCCCCAG
361 ---------+---------+---------+---------+---------+---------+ 420
    TATCTTGTCCACGGTGTCGTCGGCTGTATGCAGGTCCGGGAGAAACTGAAACTAGGGGTC
                                                Domínio SH3
a     I  E  Q  V  P  Q  Q  P  T  Y  V  Q  A  L  F  D  F  D  P  Q  -

    GAGGATGGAGAGCTGGGCTTCCGCCGGGGAGATTTTATCCATGTCATGGATAACTCAGAC
421 ---------+---------+---------+---------+---------+---------+ 480
    CTCCTACCTCTCGACCCGAAGGCGGCCCCTCTAAAATAGGTACAGTACCTATTGAGTCTG

a     E  D  G  E  L  G  F  R  R  G  D  F  I  H  V  M  D  N  S  D  -

    CCCAACTGGTGGAAAGGAGCTTGCCACGGGCAGACCGGCATGTTTCCCCGCGAATTATGT
481 ---------+---------+---------+---------+---------+---------+ 540
    GGGTTGACCACCTTTCCTCGAACGGTGCCCGTCTGGCCGTACAAAGGGGCGCTTAATACA

a     P  N  W  W  K  G  A  C  H  G  Q  T  G  M  F  P  R  E  L  C   -

    CTCCCCCXGTGAACCGGAACGTCTAAGAGTCAAGAAGCAATTATTTAAAGAAAGTGAAAA
541 ---------+---------+---------+---------+---------+---------+ 600
    GAGGGGGXCACTTGGCCTTGCAGATTCTCAGTTCTTCGTTAATAAATTTCTTTCACTTTT

a     L  P  ?  *  T  G  T  S  K  S  Q  E  A  I  I  *  R  K  *  K  -
```

FIG. 16C

```
    ATGTAAAACACATACAAAAGAATTAAACCCACAAGCTGCCTCTGACAGCAGCCTGTGAGG
601 ---------+---------+---------+---------+---------+---------+ 660
    TACATTTTGTGTATGTTTTCTTAATTTGGGTGTTCGACGGAGACTGTCGTCGGACACTCC

a       M  *  N  T  Y  K  R  I  K  P  T  S  C  L  *  Q  Q  P  V  R  -

    GAGTGCAGAACACCTGGCCGGGTCACCCTGTGACCCTCTCACTTTGGTTGGAACTTTAGG
661 ---------+---------+---------+---------+---------+---------+ 720
    CTCACGTCTTGTGGACCGGCCCAGTGGGACACTGGGAGAGTGAAACCAACCTTGAAATCC

a       E  C  R  T  P  G  R  V  T  L  *  P  S  H  F  G  W  N  F  R  -

    GGGTGGGAGGGGGCGTTGGATTTAAAAATGCCAAAACTTACCTATAAATTAAGAAGAGTT
721 ---------+---------+---------+---------+---------+---------+ 780
    CCCACCCTCCCCCGCAACCTAAATTTTTACGGTTTTGAATGGATATTTAATTCTTCTCAA

a       G  W  E  G  A  L  D  L  K  M  P  K  L  T  Y  K  L  R  R  V  -

    TTTATTACAAATTTTCACTGCTGCTCCTCTTTCCCCTCCTTTGTCTTTTTTTTTCATCCT
781 ---------+---------+---------+---------+---------+---------+ 840
    AAATAATGTTTAAAAGTGACGACGAGGAGAAAGGGGAGGAAACAGAAAAAAAAAGTAGGA

a       F  I  T  N  F  H  C  C  S  S  F  P  S  F  V  F  F  F  H  P  -

    TTTTTCTCTTCTGTCCATCAGTGCATGACGTTTAAGGCCACGTATAGTCCTAGCTGACGC
841 ---------+---------+---------+---------+---------+---------+ 900
    AAAAAGAGAAGACAGGTAGTCACGTACTGCAAATTCCGGTGCATATCAGGATCGACTGCG

a       F  F  S  S  V  H  Q  C  M  T  F  K  A  T  Y  S  P  S  *  R  -

    CAATAATAAAAACCGAATTCGAGCTCGGATCCGGGGATCCTCTAGAGTC
901 ---------+---------+---------+---------+--------- 949
    GTTATTATTTTTGGCTTAAGCTCGAGCCTAGGCCCCTAGGAGATCTCAG

a       Q  *  *  K  P  N  S  S  S  D  P  G  I  L  *  S  ?  -
```