

(11) Número de Publicação: **PT 637334 E**

(51) Classificação Internacional:  
**C12N 15/12** (2006.01) **C12P 21/08** (2006.01)  
**A61K 39/395** (2006.01) **G01N 33/577** (2006.01)  
**C12Q 1/68** (2006.01)

(12) **FASCÍCULO DE PATENTE DE INVENÇÃO**

(22) Data de pedido: <b>1993.04.19</b>	(73) Titular(es): <b>AVENTIS PHARMA, S.A.</b> <b>20 AVENUE RAYMOND ARON 92160 ANTONY</b> <b>FR</b>
(30) Prioridade(s): <b>1992.04.21 FR 9204827</b>	
(43) Data de publicação do pedido: <b>1995.02.08</b>	
(45) Data e BPI da concessão: <b>2007.06.06</b> <b>075/2007</b>	(72) Inventor(es): <b>KÁROLY TIHANYI</b> HU <b>ISTVÁN TARNAWA</b> HU <b>PÁL KOCKOCSIS</b> HU <b>GYÖRGY NÉMETH</b> HU <b>BALÁZS DALMADI</b> HU
	(74) Mandatário: <b>PEDRO DA SILVA ALVES MOREIRA</b> <b>RUA DO PATROCÍNIO, N.º 94 1399-019 LISBOA</b> PT

(54) Epígrafe: **PÉPTIDOS POSSUINDO UMA ACTIVIDADE DE FACTOR DE PERMUTA DE GDP, SEQUÊNCIAS DE ÁCIDOS NUCLEICOS CODIFICANDO ESSES PÉPTIDOS, PREPARAÇÃO E UTILIZAÇÃO.**

(57) Resumo:

## RESUMO

**"PÉPTIDOS POSSUINDO UMA ACTIVIDADE DE FACTOR DE PERMUTA DE GDP,  
SEQUÊNCIAS DE ÁCIDOS NUCLEICOS CODIFICANDO ESSES PÉPTIDOS,  
PREPARAÇÃO E UTILIZAÇÃO"**

A presente invenção refere-se a péptidos capazes de modular os níveis de permuta do GDP em complexos p21-GDP, as sequências de ácido nucleicos codificando esses péptidos, sua preparação e composições farmacêuticas contendo os mesmos.

## DESCRIÇÃO

### **"PÉPTIDOS POSSUINDO UMA ACTIVIDADE DE FACTOR DE PERMUTA DE GDP, SEQUÊNCIAS DE ÁCIDOS NUCLEICOS CODIFICANDO ESSES PÉPTIDOS, PREPARAÇÃO E UTILIZAÇÃO"**

A presente invenção refere-se a novas sequências peptídicas e nucleotídicas, e a sua utilização farmacêutica. Mais particularmente, a invenção refere-se a péptidos capazes de modular os níveis de permuta do GDP nos complexos p21-GDP.

Os produtos dos genes *ras*, geralmente designados proteínas p21, desempenham um papel chave no controlo da divisão celular em todos os organismos eucarióticos onde foram investigados. Determinadas modificações específicas destas proteínas fazem com que percam o seu controlo normal e conduzam-nas a tornarem-se oncogénicas. Assim, um grande número de tumores humanos foi associado à presença de genes *ras* modificados. Do mesmo modo, uma sobre-expressão destas proteínas p21 pode conduzir a uma desregulação da proliferação celular. A compreensão do papel exacto destas proteínas p21 nas células, do seu modo de funcionamento e das suas características constituem então um aspecto importante para a compreensão e abordagem terapêutica da carcinogénese.

*In vivo*, não se conhece ainda a natureza exacta dos eventos responsáveis da activação das proteínas p21. É sabido que exercem as suas funções oscilando entre dois estados conformacionais; uma forma inactiva ligada ao GDP e uma forma activa ligada ao GTP, mas os factores envolvidos na transição entre estas duas formas não estão claramente identificados.

Estudos recentes descrevem situações fisiológicas durante as quais a proporção de proteínas ras ligadas a GTP aumenta na célula. Estas situações compreendem a activação de linfócitos T e a estimulação de fibroblastos 3T3 por factores de crescimento, incluindo o EGF e o PDGF (Downward *et al.*, *Nature* 346 (1990) 719; Gibbs *et al.*, *J. Biol. Chem.* 265 (1990) 20437). O aumento na proporção de p21-GTP pode ser explicado, pelo menos em parte, pela acção de uma proteína que desempenha um papel análogo àquele de um receptor das proteínas G de transdução. A este respeito foram identificadas determinadas proteínas capazes de promover a permuta do GDP nas proteínas p21, a partir de cérebro de boi [West *et al.*, *FEBS Lett.* 259 (1990) 245] e cérebro de rato [Wolfman e Macara, *Science* 248 (1990) 67]. A localização celular distinta destes factores e as condições experimentais muito diferentes daquelas que foram obtidas, sugerem que se tratam de proteínas diferentes. Elas são também activas nas proteínas ras normais e naquelas que são oncogénicas. Estas actividades são reagrupadas sob o termo GEF: Factor de Permuta de nucleótidos de Guanidina ou GRF.

Na levedura *Saccharomyces cerevisiae*, a actividade de GRF foi atribuída ao produto do gene CDC25 [Camonis *et al.*, *EMBO J.* 5 (1986) 375], e foram realizados estudos a fim de compreender a via de sinalização em que intervém o produto dos genes CDC25, RAS1 e RAS2 por um lado e a adenilato ciclase por outro, em levedura *Saccharomyces cerevisiae*. Em particular, numerosos estudos focalizaram-se na caracterização do produto do gene CDC25 que era o elemento menos conhecido desta cadeia. O produto do gene CDC25 constitui o elemento mais a montante na cascata de reacções que conduzem à activação de p21 em levedura. Os trabalhos realizados neste domínio contribuíram para demonstrar que o produto deste gene deve agir como factor de permuta de GDP → GTP para activar as proteínas ras. Um segundo gene da

levedura *S. cerevisiae*, SDC25, estruturalmente muito semelhante a CDC25, foi isolado e caracterizado. O domínio activo de SDC25 parece ser um factor de permuta capaz de actuar *in vitro* e *in vivo* nas proteínas ras. Este domínio constitui o primeiro constituinte molecular descrito dotado desta actividade.

Muito recentemente, uma proteína de tipo GRF foi igualmente posta em evidência em murganhos [Vanoni e Martegani, J. Cell. Bioch. Suppl. 16B (1992) 220]. Foi descrita a purificação de uma proteína de origem bovina, denominada *rho* GDI (Inibidor da Dissociação de GDP) que regula a permuta GDP-GTP da proteína *rho* (Fukumoto *et al.*, Oncogene (1990), 5, pp. 1321), assim como a clonagem de uma proteína de origem bovina, denominada *smg* p25A GDI (Inibidor da Dissociação de GDP) que regula a permuta GDP-GTP da proteína *smg* p25A (Matsui *et al.*, Molecular and Cellular Biology (Ago. 1990) p. 41116-41122).

Contudo, até à data, nenhuma actividade de GRF foi isolada e caracterizada no homem. A presente invenção resulta precisamente da demonstração pela requerente da existência de um factor humano de permuta do GDP. A presente invenção resulta, mais particularmente, da identificação, do isolamento e da caracterização de péptidos e de sequências de nucleótidos de origem humana, designadas hGRF e hSOS, capazes de modular o estado de activação das proteínas p21.

Um primeiro aspecto da invenção consiste, portanto, em péptidos que podem ser utilizados farmacêuticamente. Mais especialmente, um objectivo da invenção encontra-se em péptidos capazes de modular os níveis de permuta do GDP nos complexos p21-GDP. É compreendido que p21 designa qualquer produto da expressão de um gene ras normal ou oncogénico.

Mais particularmente, os péptidos da invenção são escolhidos a partir de toda ou parte das sequências SEQ ID N° 2, 3, 4, 6 ou 8, ou de um derivado das mesmas.

No âmbito da presente invenção, o termo derivado designa qualquer molécula obtida por modificação de natureza genética e/ou química destas sequências e que conserva a actividade desejada. Por modificação de natureza genética e/ou química, deve ser entendida qualquer mutação, substituição, deleção, adição e/ou modificação de um ou mais resíduos. Esses derivados podem ser produzidos para objectivos diferentes, tais como particularmente para aumentar a afinidade do péptido para o seu sítio de interacção, melhorar os seus níveis de produção, aumentar a sua resistência a proteases, aumentar a sua eficácia terapêutica ou reduzir os seus efeitos secundários, ou conferir novas propriedades farmacocinéticas e/ou biológicas.

Numa forma de realização particular da invenção, os péptidos da invenção são os péptidos capazes de estimular a permuta do GDP no complexo p21-GDP.

Noutra forma de realização particular da invenção, os péptidos da invenção são os péptidos capazes de retardar ou inibir a permuta do GDP no complexo p21-GDP. Esses péptidos são, de um modo preferido, os péptidos capazes de antagonizar a interacção do factor de permuta do GDP no complexo p21-GDP. Podem-se assim tratar de fragmentos das sequências acima mencionadas ou derivados das mesmas. Esses fragmentos podem ser produzidos de formas diferentes. Em particular, podem ser sintetizados por via química, com base nas sequências apresentadas no presente pedido, utilizando sintetizadores de péptidos conhecidos do especialista na técnica. Podem ser igualmente sintetizados por via genética, por expressão num hospedeiro celular de uma sequência nucleotídica codificante do

péptido desejado. Neste caso, a sequência nucleotídica pode ser preparada quimicamente utilizando um sintetizador de oligonucleótidos, com base na sequência peptídica apresentada no presente pedido e no código genético. A sequência nucleotídica pode igualmente ser preparada a partir de sequências apresentadas no presente pedido (SEQ ID N° 1, 5 e 7), por clivagens enzimáticas, ligação, clonagem, etc., de acordo com as técnicas conhecidas do especialista na técnica, ou através de rastreio de bibliotecas de ADN com sondas produzidas a partir destas sequências. Além disso, os péptidos da invenção capazes de retardar ou de inibir a permuta de GDP no complexo p21-GDP podem igualmente ser péptidos que têm uma sequência correspondente ao sítio de interacção do factor de permuta no complexo p21-GDP.

Um outro objectivo da invenção encontra-se em anticorpos ou fragmentos de anticorpos policlonais ou monoclonais dirigidos contra um péptido como acima definido. Esses anticorpos podem ser produzidos através de métodos conhecidos do especialista na técnica, tendo em conta os ensinamentos apresentados no presente pedido. Em particular, estes anticorpos podem ser preparados através de imunização de um animal contra um péptido da invenção, recolha de sangue e isolamento dos anticorpos. Estes anticorpos podem igualmente ser produzidos através da preparação de hibridomas de acordo com as técnicas conhecidas do especialista na técnica.

De um modo mais preferido, os anticorpos ou fragmentos de anticorpos da invenção possuem a capacidade de inibir, pelo menos parcialmente, a interacção do factor de permuta com o complexo p21-GDP. Podem ser assim utilizados para regular o estado de activação do produto de genes ras.

Além disso, estes anticorpos podem igualmente ser utilizados para detectar e/ou dosear o factor de permuta do GDP humano em amostras biológicas, e conseqüentemente proporciona informação acerca do estado de activação do produto de genes ras.

A presente invenção permite portanto produzir péptidos derivados de sequências SEQ ID N° 2-4, 6 e 8, assim como anticorpos dirigidos contra estes péptidos, possuindo propriedades biológicas interessantes na perspectiva de uma utilização farmacêutica. A actividade biológica dos diferentes péptidos e anticorpos da invenção sobre a permuta do GDP pode ser avaliada de formas diferentes, como ilustrado nos exemplos.

A invenção proporciona igualmente compostos não peptídicos ou não exclusivamente peptídicos utilizáveis do ponto de vista farmacêutico. É de facto possível, a partir de motivos proteicos activos descritos no presente pedido, produzir moléculas inibidoras da via de sinalização dependente de proteína rãs, não exclusivamente peptídicas e compatíveis com uma utilização farmacêutica. A este respeito, a invenção refere-se à utilização de um polipéptido da invenção, tal como acima descrito, para a preparação de moléculas não peptídicas ou não exclusivamente peptídicas, farmacologicamente activas sobre os níveis de permuta do GDP, para determinação dos elementos estruturais deste polipéptido que são importantes para a sua actividade e reprodução destes elementos por estruturas não peptídicas ou não exclusivamente peptídicas. A invenção tem também como objecto, composições farmacêuticas compreendendo uma ou mais moléculas assim preparadas.

A presente invenção tem igualmente como objecto qualquer sequênciade ácidos nucleicos codificante de um polipéptido, tal como acima definido. De um modo mais preferido, trata-se de uma sequênciadescolhida a partir de:

(a) toda ou parte das sequências SEQ ID N° 1, 5 ou 7, ou da sua cadeia complementar,

(b) qualquer sequência que hibrida com uma sequência (a) e codificante de um polipéptido de acordo com a invenção, e

(c) as sequências derivadas a partir das sequências (a) e (b) como resultado da degenerescência do código genético.

As diferentes sequências nucleotídicas da invenção podem ser de origem artificial ou não. Podem ser sequências genómicas, de ADNc, de ARN, sequências híbridas ou sequências sintéticas ou semi-sintéticas. Estas sequências podem ser obtidas, por exemplo, através de rastreio de bibliotecas de ADN (biblioteca de ADNc, biblioteca de ADN genómico) através de sondas produzidas com base nas sequências SEQ ID N° 1, 5 ou 7. Essas bibliotecas podem ser preparadas a partir de células de diferentes origens através de técnicas convencionais de biologia molecular conhecidas do especialista na técnica. As sequências nucleotídicas da invenção podem ser igualmente preparadas através de síntese química, particularmente de acordo com o método da fosforamidite ou alternativamente através de métodos mistos incluindo a modificação química ou enzimática das sequências obtidas através de rastreio de bibliotecas.

Estas sequências nucleotídicas de acordo com a invenção podem ser utilizadas no domínio farmacêutico, seja para a produção *in vitro* de péptidos da invenção, seja para a realização de sequências anti-sentido ou para a produção de péptidos da invenção no contexto de uma terapia génica, seja alternativamente para a detecção e diagnóstico, para experiências de hibridação, de expressão ou de uma sobre-expressão de um factor de permuta do GDP amplificado, mutado ou reorganizado em amostras biológicas, ou para

isolamento de sequências homólogas a partir de outras fontes celulares.

Para a produção de péptidos da invenção, as sequências de ácidos nucleicos definidas acima são geralmente colocadas sob o controlo de sinais que permitem a sua expressão num hospedeiro celular. A escolha destes sinais (promotores, terminadores, sequência líder de secreção, etc.) pode variar em função do hospedeiro celular utilizado. De um modo preferido, estas sequências nucleotídicas da invenção fazem parte de um vector que pode ter replicação autónoma ou integrativa. Mais particularmente, os vectores de replicação autónoma podem ser preparados utilizando sequências de replicação autónoma no hospedeiro escolhido. Tratando-se de vectores integrativos, estes podem ser preparados, por exemplo, utilizando sequências homólogas a determinadas regiões do genoma do hospedeiro, permitindo a integração do vector através de recombinação homóloga.

Os hospedeiros celulares utilizáveis para a produção de péptidos da invenção são igualmente bons hospedeiros eucarióticos ou procarióticos. Entre os hospedeiros eucarióticos que são adequados, podem ser mencionadas as células animais, as leveduras ou os fungos. Em particular, tratando-se de leveduras, podem ser mencionadas as leveduras do género *Saccharomyces*, *Kluyveromyces*, *Pichia*, *Schwanniomyces* ou *Hansenula*. Tratando-se de células animais, podem ser mencionadas as células COS, CHO, C127, etc. Tratando-se de fungos, podem ser mencionados mais particularmente *Aspergillus* ssp. ou *Trichoderma* ssp. Como hospedeiros procarióticos é preferido utilizar-se as seguintes bactérias *E. coli*, *Bacillus* ou *Streptomyces*.

As sequências de ácidos nucleicos de acordo com a invenção podem ser igualmente utilizadas para a produção de

oligonucleótidos anti-sentido ou de sequências anti-sentido genéticas utilizáveis como agentes farmacêuticos. A inibição da expressão de determinados oncogenes através de sequências anti-sentido provou ser uma estratégia útil na compreensão do papel destes oncogenes e uma via particularmente promissora na realização de um tratamento anticancerígeno. As sequências anti-sentido são oligonucleótidos de tamanho pequeno, complementares à cadeia codificante de um determinado gene, e como resultado, são capazes de hibridar especificamente com o transcrito de ARNm, inibindo a sua tradução em proteína. A invenção tem assim como objectivo sequências anti-sentido capazes de inibir, pelo menos parcialmente, a produção de péptidos que estimulam a permuta do GDP nos complexos p21-GDP. Essas sequências podem ser constituídas por toda ou parte das sequências de ácidos nucleicos definidas acima. Tratam-se geralmente de sequências ou fragmentos de sequências complementares de sequências codificantes de péptidos que estimulam a permuta do GDP. Esses oligonucleótidos podem ser obtidos a partir das sequências SEQ ID N° 1, 5 ou 7, por fragmentação, etc., ou por síntese química. Essas sequências podem ser utilizadas no contexto de terapias génicas para a transferência e a expressão *in vivo* de sequências anti-sentido ou de péptidos capazes de modular os níveis de permutas do GDP nas proteínas ras. A este respeito, as sequências podem ser incorporadas nos vectores, em particular de origem viral.

A invenção refere-se igualmente, como sequências nucleotídicas, a sondas nucleotídicas, sintéticas ou não, capazes de hibridar com as sequências nucleotídicas definidas acima que codificam um péptido da invenção ou com o ARNm correspondente. Essas sondas podem ser utilizadas *in vitro* como ferramenta de diagnóstico para a detecção da expressão do factor de permuta do GDP, ou alternativamente para pôr em evidência anomalias genéticas (*splicing* incorrecto, polimorfismo, mutações

pontuais, etc.). Essas sondas devem ser previamente marcadas, e para tal são conhecidas do especialista na técnica diferentes métodos. As condições de hibridação sob as quais estas sondas podem ser utilizadas são as condições normais de restringência (ver particularmente as técnicas gerais de clonagem abaixo, assim como os exemplos). Estas sondas podem ser igualmente utilizadas para pôr em evidência e para o isolamento de sequências de ácido nucleicos homólogas codificantes de um péptido da invenção, a partir de outras fontes celulares, como ilustrado nos exemplos.

A invenção tem também como objecto qualquer composição farmacêutica compreendendo como princípio activo, pelo menos, um péptido como acima definido.

Tem também como objecto qualquer composição farmacêutica compreendendo como princípio activo, pelo menos, um anticorpo e/ou fragmento de anticorpo como acima definido, assim como qualquer composição farmacêutica compreendendo como princípio activo, pelo menos, uma sequência nucleotídica como acima definido.

Além disso, tem também como objecto composições farmacêuticas em que os péptidos, anticorpos e sequências nucleotídicas acima definidos são associados entre si ou com outros princípios activos.

As composições farmacêuticas de acordo com a invenção podem ser utilizadas para modular a activação das proteínas p21 e, como resultado, modular a proliferação de determinados tipos celulares. Mais particularmente, estas composições farmacêuticas são destinadas ao tratamento de cancros. De facto, numerosos cancros foram associados à presença de proteínas ras oncogénicas. Entre os cancros contendo mais frequentemente genes

ras mutados, podem ser mencionados particularmente os adenocarcinomas do pâncreas, 90% dos quais têm um oncogene Ki-ras mutado no décimo segundo codão [Almoguera *et al.*, *Cell* 53 (1988) 549], os adenocarcinomas do cólon e os cancros da tiróide (50%), ou os carcinomas do pulmão e as leucemias mielóides [30%, Bos, J. L. *Cancer Res.* 49 (1989) 4682].

A invenção tem também como objecto a utilização de moléculas acima descritas para modular a actividade das proteínas p21. Em particular, a invenção refere-se à utilização destas moléculas para inibir, pelo menos parcialmente, a activação das proteínas p21.

A invenção proporciona igualmente um processo de detecção da expressão e/ou de uma sobre-expressão de um gene ras numa amostra biológica. Esse processo compreende, por exemplo, a colocação em contacto dessa amostra com um anticorpo ou fragmento de anticorpo de acordo com a invenção, a revelação dos complexos antigénio-anticorpo e a comparação dos resultados obtidos com uma amostra padrão. Num tal processo, o anticorpo pode estar em suspensão ou previamente imobilizado num suporte. Este processo pode compreender igualmente a colocação em contacto da amostra com uma sonda nucleotídica de acordo com a invenção, pôr em evidência os híbridos obtidos e a comparação com aqueles obtidos no caso de uma amostra padrão.

A presente invenção pode ser utilizada no domínio terapêutico: os péptidos, anticorpos e sequências nucleotídicas da invenção são capazes de modular a actividade de genes ras permitindo, de facto, intervir no processo de desenvolvimento dos cancros. Como ilustrado nos exemplos, as sequências nucleotídicas da invenção permitem particularmente a expressão de péptidos capazes de complementar a sensibilidade à temperatura de leveduras que contêm uma mutação *cdc25*. Estes

permitem igualmente expressar os péptidos capazes de suprimir uma mutação dominante RAS2ts, demonstrando uma competição com o produto normal da expressão do gene CDC25 para a interacção com as proteínas p21. A invenção pode ser igualmente utilizada no domínio do diagnóstico e da tipagem de cancros: os anticorpos e sondas nucleotídicas da invenção permitem, de facto, a identificação de cancros nos quais está implicado um gene ras, assim como o diagnóstico de cancros associados à sobre-expressão de um gene ras normal ou oncogénico.

Outras vantagens da presente invenção tornar-se-ão evidentes com a leitura dos exemplos seguintes que devem ser considerados como ilustrativos e não limitativos.

#### Legenda das figuras

Figura 1: Teste de transactivação: Esta figura mostra em ordenada os níveis de actividade CAT (%relativa ao ruído de fundo) de estirpes CHO recombinantes, em função da sequência de ADNc utilizada para a transformação.

Figura 2: Teste de actividade de permuta de GTP: Esta figura mostra em ordenada a proporção das formas p21-GDP/p21-GTP de estirpes CHO recombinantes, em função da sequência de ADNc utilizada para a transformação.

Figura 3: Teste de actividade de permuta de GDP *in vitro*: Esta figura mostra, em função do tempo e para 2 concentrações de um péptido da invenção, a diminuição da proporção de GDP que permaneceu ligada à proteína p21.

## Técnicas gerais de clonagem

Os métodos classicamente utilizados na biologia molecular, tais como as extracções preparativas de ADN plasmídico, a centrifugação de ADN plasmídico em gradiente de cloreto de céσιο, electroforese em gel de acrilamida ou agarose, a purificação de fragmentos de ADN por eletroeluição, as extracções de proteína em fenol ou fenol-clorofórmio, a precipitação de ADN em meio salino pelo etanol ou isopropanol, a transformação em *Escherichia coli*, etc., são bem conhecidos do especialista na técnica e estão amplamente descritas na literatura [Maniatis, T. *et al.*, "Molecular Cloning, a Laboratory Manual", Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982; Ausubel F. M. *et al.* (eds), "Current Protocols in Molecular Biology", John Wiley and Sons, Nova Iorque, 1987].

As enzimas de restrição foram fornecidas pela New England Biolabs (Biolabs), Bethesda Research Laboratories (BRL) ou Amersham e são utilizadas de acordo com as recomendações dos fornecedores.

Os plasmídeos de tipo pBR322, pUC,  $\lambda$ gt11 e pGEX 2T e os fagos da série M13 são de origem comercial.

Para as ligações, os fragmentos de ADN são separados de acordo com o seu tamanho por electroforese em gel de agarose ou de acrilamida, extraídos em fenol ou por uma mistura de fenol/clorofórmio, precipitados em etanol e depois incubados na presença de ADN ligase do fago T4 (Biolabs) de acordo com as recomendações do fornecedor.

O preenchimento das extremidades 5' proeminentes é realizado pelo fragmento Klenow da ADN polimerase I de *E. coli*

(Biolabs) de acordo com as especificações do fornecedor. A destruição das extremidades 3' proeminentes é realizada na presença de ADN polimerase do fago T4 (Biolabs) utilizada de acordo com as recomendações do fabricante. A destruição das extremidades 5' proeminentes é realizada por um tratamento conduzido pela nuclease S1.

A mutagénese dirigida *in vitro* por oligodesoxinucleótidos sintéticos é efectuada de acordo com o método desenvolvido por Taylor *et al.* [Nucleic Acids Res. 13 (1985) 8749-8764] utilizando o kit distribuído pela Amersham.

A amplificação enzimática de fragmentos de ADN pela técnica denominada de PCR [Reacção em Cadeia da Polimerase, Saiki R. K. *et al.*, Science 230 (1985) 1350-1354; Mullis K. B. e Faloona F. A., Meth. Enzym. 155 (1987) 335-350] é efectuada utilizando um "DNA thermal cycler" (Perkin Elmer Cetus) de acordo com as especificações do fabricante.

A verificação das sequências nucleotídicas é realizada pelo método desenvolvido por Sanger *et al.* [Proc. Natl. Acad. Sci. USA, 74 (1977) 5463-5467] utilizando o kit distribuído pela Amersham.

Para as experiências de hibridação, as condições de restringência normais são geralmente as seguintes: hibridação: 3 x SCC na presença de 5 x Denhart a 65 °C; lavagem: 0,5 x SSC a 65 °C.

1. Isolamento do gene do factor humano de permuta do GDP (hGRF)

Foram rastreados  $5 \times 10^5$  fagos de uma biblioteca de cérebro humano construída no vector  $\lambda$ gt11 [Skolnik et al., Cell 65 (1991) 83] de acordo com as técnicas descritas por Sambrook, Fritsch e Maniatis (Molecular cloning; Cold Spring Harbor Laboratory Press; 1989).

A sonda utilizada para o rastreio desta biblioteca é um fragmento de ADNc humano de 137 pares de bases marcado com  $^{32}\text{P}$ . Esta sonda foi preparada por PCR com ADN total da referida biblioteca, utilizando como iniciadores os oligonucleótidos degenerados seguintes:

- ATC CGT CAG GTA CAT CCC CAG GTA TGG CAC ACA  
T T T T T T T T  
A A A A G A  
G G G

para o oligonucleótido da extremidade 3' do fragmento (SEQ ID N° 11), e

- GCA ATT TTT CGG CTT AAG AAG ACT TGG  
T C C A C A A C  
C A T G A  
G C A G

para o oligonucleótido da extremidade 5' do fragmento (SEQ ID N° 12).

A sonda foi marcada com  $^{32}\text{P}$  por "iniciação aleatória" de acordo com a técnica de Feinberg e Vogelstein [Anal. Biochem.

137 (1984) 266], e a reacção de PCR foi realizada a cerca de 40 °C sob as condições descritas nas técnicas gerais de clonagem.

De entre os diferentes clones positivos obtidos por hibridação com esta sonda, um compreende a totalidade de uma grelha de leitura aberta que contém a actividade de permuta relativamente a Ha-Ras.

Este clone de  $\lambda$ gt11 contém um ADNc de 3 kb que foi introduzido na forma de um fragmento EcoRI no sítio correspondente de um vector M13mp18. Este ADNc foi depois sequenciado com auxílio de iniciadores comerciais "reverso" e "-20" assim como com o auxílio de oligonucleótidos específicos de acordo com a técnica de Sanger (ver técnicas gerais de clonagem).

A sequência nucleotídica do gene do factor humano de permuta do GDP contida no fragmento assim obtido é apresentada na sequência SEQ ID N° 1, assim como nas sequências peptídicas deduzidas (SEQ ID N° 2, 3 e 4).

## 2. Preparação de subfragmentos

Diferentes derivados ou fragmentos do gene assim obtido podem ser preparados e utilizados, particularmente para a expressão de péptidos da invenção. Em particular, os fragmentos seguintes foram preparados através de clivagem enzimática e separados através de eletroeluição:

- um fragmento PstI-EcoRI compreendendo uma parte da região codificante do factor de permuta (do nucleótido 936 até ao

codão de terminação) assim como a região 3' não codificante do fragmento,

- um fragmento EcoNI-EcoRI compreendendo uma parte da região codificante do factor de permuta (do nucleótido 910 até ao codão de terminação) assim como a região 3' não codificante do fragmento,
- um fragmento EagI-EcoRI compreendendo uma parte da região codificante do factor de permuta (do nucleótido 638 até ao codão de terminação) assim como a região 3' não codificante do fragmento,
- um fragmento BalI-EcoRI compreendendo uma parte da região codificante do factor de permuta (do nucleótido 88 até ao codão de terminação) assim como a região 3' não codificante do fragmento, e
- um fragmento NaeI-EcoRI compreendendo uma parte da região codificante do factor de permuta (do nucleótido 196 até ao codão de terminação) assim como a região 3' não codificante do fragmento,

É compreendido que a região 3' não codificante contida nestes fragmentos é acessória e que pode ser eliminada, quer por digestão através de uma nuclease, quer através de clivagem com uma enzima que tem um sítio perto do codão de terminação, tal como particularmente SmaI, cujo o sítio está localizado a cerca de 30 pb após o codão de terminação.

É igualmente compreendido que outros fragmentos podem ser preparados, tal como particularmente os fragmentos que não contêm a região codificante para a parte C-terminal inteira, assim como derivados destes fragmentos, obtidos através de

mutação, substituição, adição ou modificação de natureza química e/ou genética.

### 3. Caracterização biológica

As funcionalidades dos péptidos de acordo com a invenção foram testadas:

- em células de mamíferos,
- na levedura *Saccharomyces cerevisiae* ou então
- *in vitro* na proteína Ha-Ras recombinante.

3.1. Para avaliação funcional em células de mamíferos, as sequências de ADN codificantes de péptidos da invenção, tais como por exemplo, aquelas descritas no Exemplo 2, podem ser colocadas sob controlo do promotor precoce de SV40 no vector pCym1 descrito por Camonis *et al.* [Gene 86 (1990) 263].

Neste exemplo, os fragmentos PstI-EcoRI e EcoNI-EcoRI descritos no Exemplo 2 foram introduzidos neste vector.

Os vectores assim obtidos foram testados para transfecção transitória em células CHO de acordo com o protocolo descrito por Rey *et al.* [Oncogene 6 (1991) 347].

Dois critérios de funcionalidade foram assim estudados:

- a capacidade dos vectores de transactivar um promotor que governa a expressão de um gene repórter, que é aqui o gene bacteriano codificante da cloranfenicol acetiltransferase (gene CAT),

- a sua capacidade para promover a carga em GTP das proteínas Ras de células CHO transfectadas.

a) Para os testes de transactivação, as células CHO a 50% de confluência foram transfectadas (ver, por exemplo, o protocolo descrito por Schweighoffer *et al.*, Science, no prelo), uma parte com 0,5 µg de um vector contendo o gene CAT sob controlo de um promotor sintético composto dum promotor murino do gene da timidina cinase e de 4 elementos PEA1 repetidos derivados do intensificador de polioma [Wasylyk *et al.*, EMBO J. 7 (1988) 2475], e outra parte com 4,5 µg de um vector de expressão contendo sob controlo do promotor precoce SV40, nenhum ADNc codificante (pista 1), o ADNc do Ha-Ras normal (pista 2), o ADNc do Ha-Ras activado em 12 (Val 12) (pista 3), a extremidade 3' do ADNc de SDC25, descrito por Rey *et al.* citado acima (pista 4), e o ADNc codificante de um péptido de acordo com a invenção, acima descrito (pista 5).

Os resultados são apresentados na figura 1. A pista 1 corresponde à activação basal, a pista 3 (obtida para o fragmento PstI-EcoRI: CDC hum.) mostra que a expressão de um péptido da invenção permite a transactivação do promotor sintético utilizado.

O mesmo resultado qualitativo foi obtido para os outros fragmentos estudados.

b) De modo a verificar que esta transactivação envolve de facto uma carga nucleotídica de proteínas ras de células CHO, as mesmas transfecções transitórias foram realizadas, o meio de cultura foi adicionado com ortofosfato marcado com <sup>32</sup>P.

Este protocolo de marcação, assim como aquele de imunoprecipitação de proteínas ras celulares é descrito por Rey *et al.* citado acima. Os resultados obtidos são apresentados na figura 2. Estes mostram que os péptidos da invenção são capazes de modular os níveis de permuta do GDP nas proteínas ras, desde que algumas delas (péptido CDC Hum. expresso pelo fragmento PstI-EcoRI acima descrito) sejam capazes de promover a carga em GTP das proteínas ras de células CHO imunoprecipitadas por anticorpos de Y 13-259.

3.2. Os péptidos da invenção foram igualmente testados funcionalmente na levedura *S. cerevisiae* cdc25. Na verdade, os vectores acima descritos (3.1.) que são vectores vaivém foram introduzidos na estirpe OL97-1-11b de levedura [Camonis e Jacquet, *Mol. Cell. Biol.* 8 (1988) 2980]. Os resultados obtidos mostram que os fragmentos de ADNc de acordo a invenção codificam os péptidos que são capazes de complementar a falta de crescimento desta estirpe a 36 °C. Estes resultados mostram assim que estes fragmentos codificam péptidos funcionais *in vivo* em *S. cerevisiae*.

3.3. A capacidade das sequências de ADN da invenção para codificarem péptidos capazes de promover a permuta do GDP nas proteínas Ha-Ras purificadas foi igualmente demonstrada *in vitro* de acordo com o protocolo descrito por Rey *et al.* *Mol. Cell. Biol.* 9 (1989) 3904].

Para tal, as sequências da invenção são expressas na estirpe TG1 de *E. coli* na forma de proteínas da fusão com a glutathione S-transferase (GST) de acordo com a técnica descrita por Smith e Johnson [*Gene* 67 (1988) 31]. Para isso, os diferentes fragmentos de ADN descritos em 3.1. acima foram

clonados, na forma de fragmentos SmaI-EcoRI, no vector pGEX 2T (Pharmacia), em 3' e em fase com um ADNc codificante da GST. Os fragmentos SmaI-EcoRI são obtidos por adição de um adaptador ao meio de uma ligase. Os vectores assim obtidos são depois utilizados para transformar a estirpe TG1 de *E. coli*. As células assim transformadas são pré-cultivadas de um dia para o outro a 37 °C, diluídas a 1/10 em meio LB, adiciona-se IPTG para induzir a expressão (2 horas, 25 °C) e cultivadas depois durante aproximadamente 21 horas a 25 °C. As células são depois lisadas e as proteínas de fusão produzidas são purificadas por afinidade em coluna de agarose-GSH. Para tal, o lisado bacteriano é incubado na presença do gel (preparado e equilibrado com o tampão de lise) durante 15 minutos a 4 °C. Após 3 lavagens com um tampão Tris-HCl, pH 7,4, as proteínas são eluídas na presença de um tampão Tris-HCl, pH 7,7, contendo um excesso de GSH. O sobrenadante é recolhido e centrifugado.

A actividade da permuta do GDP dos péptidos da invenção nas proteínas Ha-Ras purificadas foi depois demonstrada *in vitro* de acordo com protocolo descrito por Rey *et al.* (Mol. Cell. Biol. citado acima). Os resultados obtidos são apresentados na figura 3. Estes mostram particularmente que o péptido da invenção corresponde à sequência CDC hum expressa pelo fragmento PstI-EcoRI descrito acima estimula a permuta do GDP.

#### 4. Demonstração de sequências homólogas

As sequências de ácidos nucleicos homólogas àquelas apresentadas na figura 1 foram demonstradas por duas estratégias diferentes:

- por PCR, nas condições descritas nas técnicas gerais de clonagem, nos ADN neo-sintetizados a partir de ARNm de

placenta, e utilizando como iniciadores oligonucleótidos degenerados escolhidos para cobrir as sequências conservadas entre a sequência SEQ ID N° 1 e a sequência da proteína SOS [Bonfini *et al.*, Science 255 (1992) 603].

- por rastreio de uma biblioteca de ADNc de placenta utilizando uma sonda constituída pela totalidade da sequência SEQ ID N° 1 marcada com <sup>32</sup>P. O rastreio foi realizado em condições de baixa restringência: hibridação a 50 °C em meio 5 x SSC, 5 x Denhart; mais lavagem a 50 °C em meio 2 x SSC.

Estas duas estratégias permitem revelar as sequências homólogas à sequência SEQ ID N° 1. Estas sequências podem ser prontamente isoladas e caracterizadas. Estas constituem sequências de ácidos nucleicos no sentido da presente invenção quando codificam (ou seus fragmentos ou derivados) péptidos capazes de modular os níveis de permuta do GDP nos complexos p21-GDP.

Em particular, esta estratégia permitiu a identificação a partir de ARNm de placenta e utilizando os oligonucleótidos oligo 2449 (SEQ ID N° 9) e oligo 2451 (SEQ ID N° 10), os 2 ADNC codificantes dos factores designados hSOS1 e hSOS2, cujas sequências parciais estão representadas nas SEQ ID N° 5 e SEQ ID N° 7, respectivamente. Os ARNm que correspondem a estes factores estão presentes em todos os tecidos em que foram procurados, contrariamente ao factor descrito no Exemplo 1 que parece localizado apenas no cérebro. A localização cromossómica destes genes foi realizada e produziu os seguintes resultados:

- h-GRF: 15q2.4.
- h-SOS1: 4q2.1.
- h-SOS2: 14q2.2.

5. Pesquisa de factores de permuta de tipo h-GRF noutros tecidos

Foi preparado um anticorpo anti-h-GRF em coelho, através de imunização com um antigénio correspondente ao fragmento de 280 aminoácidos localizados entre os resíduos 211 e 489 do factor h-GRF apresentado em SEQ ID N° 4.

Estes anticorpos permitiram a detecção, por ELISA, no córtex humano e em células precursoras do cérebro, de proteínas de pesos moleculares aparentes de 30, 55, 75, 95 e 140 kDa. A diversidade de pesos moleculares sugere a presença de múltiplos ADN. A pré-incubação de anticorpos anti-h-GRF com o h-GRF suprime a detecção de proteínas identificadas, demonstrando desse modo a especificidade do sinal.

LISTAGEM DE SEQUÊNCIAS

(1) INFORMAÇÃO GERAL:

(i) REQUERENTE:

(A) NOME: RHONE-POULENC RORER S.A.

(B) RUA: 20, avenue Raymond ARON

(C) CIDADE: ANTONY

(E) PAIS: FRANÇA

(F) CÓDIGO POSTAL: 92165

(ii) TÍTULO DA INVENÇÃO: PÉPTIDOS QUE INIBEM A ACTIVIDADE DE  
PROTEÍNAS RAS, PREPARAÇÃO E UTILIZAÇÃO

(iii) NÚMERO DE SEQUÊNCIAS: 12

(iv) FORMA DE LEITURA POR COMPUTADOR:

(A) TIPO DE MEIO: disquete

(B) COMPUTADOR: IBM PC compatível

(C) SISTEMA OPERATIVO: PC-DOS/MS-DOS

(D) PROGRAMA INFORMÁTICO: Patentin Release #1.0, Versão  
#1.25 (OEB)

(2) INFORMAÇÃO PARA A SEQ ID N°:1:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 2652 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: dupla

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(iii) HIPOTÉTICA: NÃO

(iii) ANTI-SENTIDO: NÃO

(ix) CARACTERÍSTICA ADICIONAL:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 1..2445

(ix) CARACTERÍSTICA ADICIONAL:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 445..2445 (SEQ ID N° 3)

(ix) CARACTERÍSTICA ADICIONAL:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 976..2445 (SEQ ID N° 4)

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 1:

GGC	GAT	GGC	TGT	AAG	ATC	CTC	CTG	GAC	ACC	AGC	CAG	ACC	TTT	GTG	AGA	48
Gly	Asp	Gly	Cys	Lys	Ile	Leu	Leu	Asp	Thr	Ser	Gln	Thr	Phe	Val	Arg	
1				5					10					15		
CAA	GGT	TCC	CTC	ATT	CAG	GTG	CCC	ATG	TCT	GAA	AAG	GGC	AAG	ATC	ACC	96
Gln	Gly	Ser	Leu	Ile	Gln	Val	Pro	Met	Ser	Glu	Lys	Gly	Lys	Ile	Thr	
			20					25					30			
AGG	GGG	CGC	CTG	GGG	TCT	CTC	TCC	CTA	AAG	AAA	GAG	GGC	GAG	CGA	CAG	144
Arg	Gly	Arg	Leu	Gly	Ser	Leu	Ser	Leu	Lys	Lys	Glu	Gly	Glu	Arg	Gln	
		35					40					45				
TGC	TTC	CTG	TTT	TCT	AAG	CAT	CTG	ATT	ATC	TGT	ACC	AGA	GGC	TCT	GGA	192
Cys	Phe	Leu	Phe	Ser	Lys	His	Leu	Ile	Ile	Cys	Thr	Arg	Gly	Ser	Gly	
	50					55					60					
GGG	AAG	CTT	CAC	TTG	ACC	AAG	AAT	GGA	GTC	ATA	TCC	CTC	ATT	GAC	TGC	240
Gly	Lys	Leu	His	Leu	Thr	Lys	Asn	Gly	Val	Ile	Ser	Leu	Ile	Asp	Cys	
65					70					75					80	

ACT TTA TTG GAG GAG CCA GAA AGC ACG GAG GAG GAA GCC AAA GGA TCC	288
Thr Leu Leu Glu Glu Pro Glu Ser Thr Glu Glu Glu Ala Lys Gly Ser	
85 90 95	
GGC CAA GAC ATA GAT CAC TTG GAT TTT AAA ATC GGG GTG GAG CCA AAG	336
Gly Gln Asp Ile Asp His Leu Asp Phe Lys Ile Gly Val Glu Pro Lys	
100 105 110	
GAT TCC CCG CCC TTT ACA GTC ATC CTA GTG GCC TCG TCC AGA CAG GAG	384
Asp Ser Pro Pro Phe Thr Val Ile Leu Val Ala Ser Ser Arg Gln Glu	
115 120 125	
AAG GCA GCG TGG ACC AGT GAC ATC AGC CAG TGT GTT GGT AAC ATC CGA	432
Lys Ala Ala Trp Thr Ser Asp Ile Ser Gln Cys Val Gly Asn Ile Arg	
130 135 140	
TGC AAT GGG CTC ATG ATG AAG CCA TTT GAA GAA AAT TCC AAG GTC ACT	480
Cys Asn Gly Leu Met Met Lys Pro Phe Glu Glu Asn Ser Lys Val Thr	
145 150 155 160	
GTG CCG CAG ATG ATC AAG TCC GAC GCC TCC TTA TAT TGT GAT GAT GTT	528
Val Pro Gln Met Ile Lys Ser Asp Ala Ser Leu Tyr Cys Asp Asp Val	
165 170 175	
GAC ATT CGC TTC AGC AAA ACC ATG AAC TCC TGC AAA GTG CTG CAG ATC	576
Asp Ile Arg Phe Ser Lys Thr Met Asn Ser Cys Lys Val Leu Gln Ile	
180 185 190	
GCC TAC GCC AGT GTG GAG CGG CTG CTG GAG AGG CTG ACG GAC CTG CGC	624
Ala Tyr Ala Ser Val Glu Arg Leu Leu Glu Arg Leu Thr Asp Leu Arg	
195 200 205	
TTC CTG AGC ATC GAC TTC CTC AAC ACC TTC CTG CAC TCC TAC CGC GTC	672
Phe Leu Ser Ile Asp Phe Leu Asn Thr Phe Leu His Ser Tyr Arg Val	
210 215 220	
TTC ACC ACC GCC ATC GTG GTC CTG GAC AAG CTC ATT ACC ATC TAC AAG	720
Phe Thr Thr Ala Ile Val Val Leu Asp Lys Leu Ile Thr Ile Tyr Lys	
225 230 235 240	
AAG CCT ATC AGT GCC ATT CCT GCC AGG TCG CTG GAG CTC CTG TTT GCC	768
Lys Pro Ile Ser Ala Ile Pro Ala Arg Ser Leu Glu Leu Leu Phe Ala	
245 250 255	
AGT GGC CAG AAC AAT AAG CTC CTG TAC GGT GAA CCC CCC AAG TCC CCG	816
Ser Gly Gln Asn Asn Lys Leu Leu Tyr Gly Glu Pro Pro Lys Ser Pro	
260 265 270	
CGC GCC ACC CGC AAG TTC TCC TCG CCG CCA CCT CTG TCC ATC ACC AAG	864
Arg Ala Thr Arg Lys Phe Ser Ser Pro Pro Pro Leu Ser Ile Thr Lys	
275 280 285	
ACA TCG TCA CCG AGC CGC CGG CGG AAG CTC TCC CTG AAC ATC CCC ATC	912
Thr Ser Ser Pro Ser Arg Arg Arg Lys Leu Ser Leu Asn Ile Pro Ile	
290 295 300	
ATC ACT GGC GGC AAG GCC CTG GAC CTG GCC GCC CTC AGC TGC AAC TCC	960
Ile Thr Gly Gly Lys Ala Leu Asp Leu Ala Ala Leu Ser Cys Asn Ser	
305 310 315 320	
AAT GGC TAC ACC AGC ATG TAC TCG GCC ATG TCA CCC TTC AGC AAG GCC	1008
Asn Gly Tyr Thr Ser Met Tyr Ser Ala Met Ser Pro Phe Ser Lys Ala	
325 330 335	

ACG	CTG	GAC	ACC	AGC	AAG	CTC	TAT	GTG	TCC	AGC	AGC	TTC	ACC	AAC	AAG	1056
Thr	Leu	Asp	Thr	Ser	Lys	Leu	Tyr	Val	Ser	Ser	Ser	Phe	Thr	Asn	Lys	
			340					345					350			
ATT	CCA	GAT	GAG	GGC	GAT	ACG	ACC	CCT	GAG	AAG	CCC	GAA	GAC	CCT	TCA	1104
Ile	Pro	Asp	Glu	Gly	Asp	Thr	Thr	Pro	Glu	Lys	Pro	Glu	Asp	Pro	Ser	
		355					360					365				
GCG	CTC	AGC	AAG	CAG	AGC	TCA	GAA	GTC	TCC	ATG	AGA	GAG	GAG	TCA	GAT	1152
Ala	Leu	Ser	Lys	Gln	Ser	Ser	Glu	Val	Ser	Met	Arg	Glu	Glu	Ser	Asp	
	370					375					380					
ATT	GAT	CAA	AAC	CAG	AGT	GAT	GAT	GGT	GAT	ACT	GAA	ACA	TCA	CCA	ACT	1200
Ile	Asp	Gln	Asn	Gln	Ser	Asp	Asp	Gly	Asp	Thr	Glu	Thr	Ser	Pro	Thr	
	385				390				395						400	
AAA	TCT	CCA	ACA	ACA	CCC	AAA	TCA	GTC	AAA	AAC	AAA	AAT	TCT	TCA	GAG	1248
Lys	Ser	Pro	Thr	Thr	Pro	Lys	Ser	Val	Lys	Asn	Lys	Asn	Ser	Ser	Glu	
			405					410						415		
TTC	CCA	CTC	TTT	TCC	TAT	AAC	AAT	GGA	GTC	GTC	ATG	ACC	TCC	TGT	CGT	1296
Phe	Pro	Leu	Phe	Ser	Tyr	Asn	Asn	Gly	Val	Val	Met	Thr	Ser	Cys	Arg	
			420					425					430			
GAA	CTG	GAC	AAT	AAC	CGC	AGT	GCC	TTG	TCG	GCC	GCC	TCT	GCC	TTT	GCC	1344
Glu	Leu	Asp	Asn	Asn	Arg	Ser	Ala	Leu	Ser	Ala	Ala	Ser	Ala	Phe	Ala	
		435				440						445				
ATA	GCA	ACC	GCC	GGG	GCC	AAC	GAG	GGC	ACC	CCA	AAC	AAG	GAG	AAG	TAC	1392
Ile	Ala	Thr	Ala	Gly	Ala	Asn	Glu	Gly	Thr	Pro	Asn	Lys	Glu	Lys	Tyr	
	450					455					460					
CGG	AGG	ATG	TCC	TTA	GCC	AGT	GCA	GGG	TTT	CCC	CCA	GAC	CAG	AGG	AAT	1440
Arg	Arg	Met	Ser	Leu	Ala	Ser	Ala	Gly	Phe	Pro	Pro	Asp	Gln	Arg	Asn	
	465				470					475					480	
GGA	GAC	AAG	GAG	TTT	GTG	ATC	CGC	AGA	GCA	GCC	ACC	AAT	CGT	GTC	TTG	1488
Gly	Asp	Lys	Glu	Phe	Val	Ile	Arg	Arg	Ala	Ala	Thr	Asn	Arg	Val	Leu	
			485				490							495		
AAC	GTG	CTC	CGC	CAC	TGG	GTG	TCC	AAG	CAC	TCT	CAG	GAC	TTT	GAG	ACC	1536
Asn	Val	Leu	Arg	His	Trp	Val	Ser	Lys	His	Ser	Gln	Asp	Phe	Glu	Thr	
		500					505						510			
AAC	GAT	GAG	CTC	AAA	TGC	AAG	GTG	ATC	GGC	TTC	CTG	GAA	GAA	GTC	ATG	1584
Asn	Asp	Glu	Leu	Lys	Cys	Lys	Val	Ile	Gly	Phe	Leu	Glu	Glu	Val	Met	
		515				520						525				
CAC	GAC	CCG	GAG	CTC	CTG	ACC	CAG	GAG	CGG	AAG	GCT	GCA	GCC	AAC	ATC	1632
His	Asp	Pro	Glu	Leu	Leu	Thr	Gln	Glu	Arg	Lys	Ala	Ala	Ala	Asn	Ile	
	530					535					540					
ATC	AGG	ACT	CTG	ACC	CAG	GAG	GAC	CCA	GGT	GAC	AAC	CAG	ATC	ACG	CTG	1680
Ile	Arg	Thr	Leu	Thr	Gln	Glu	Asp	Pro	Gly	Asp	Asn	Gln	Ile	Thr	Leu	
	545				550					555					560	
GAG	GAG	ATC	ACG	CAG	ATG	GCT	GAA	GGC	GTG	AAG	GCT	GAG	CCC	TTT	GAA	1728
Glu	Glu	Ile	Thr	Gln	Met	Ala	Glu	Gly	Val	Lys	Ala	Glu	Pro	Phe	Glu	
			565					570						575		

AAC CAC TCA GCC CTG GAG ATC GCG GAG CAG CTG ACC CTG CTA GAT CAC	1776
Asn His Ser Ala Leu Glu Ile Ala Glu Gln Leu Thr Leu Leu Asp His	
580 585 590	
CTC GTC TTC AAG AAG ATT CCT TAT GAG GAG TTC TTC GGA CAA GGA TGG	1824
Leu Val Phe Lys Lys Ile Pro Tyr Glu Glu Phe Phe Gly Gln Gly Trp	
595 600 605	
ATG AAA CTG GAA AAG AAT GAA AGG ACC CCT TAT ATC ATG AAA ACC ACT	1872
Met Lys Leu Glu Lys Asn Glu Arg Thr Pro Tyr Ile Met Lys Thr Thr	
610 615 620	
AAG CAC TTC AAT GAC ATC AGT AAC TTG ATT GCT TCA GAA ATC ATC CGC	1920
Lys His Phe Asn Asp Ile Ser Asn Leu Ile Ala Ser Glu Ile Ile Arg	
625 630 635 640	
AAT GAG GAC ATC AAC GCC AGG GTG AGC GCC ATC GAG AAG TGG GTG GCC	1968
Asn Glu Asp Ile Asn Ala Arg Val Ser Ala Ile Glu Lys Trp Val Ala	
645 650 655	
GTA GCT GAC ATA TGC CGC TGC CTC CAC AAC TAC AAT GCC GTA CTG GAG	2016
Val Ala Asp Ile Cys Arg Cys Leu His Asn Tyr Asn Ala Val Leu Glu	
660 665 670	
ATC ACC TCG TCC ATG AAC CGC AGT GCA ATC TTC CGG CTC AAA AAG ACG	2064
Ile Thr Ser Ser Met Asn Arg Ser Ala Ile Phe Arg Leu Lys Lys Thr	
675 680 685	
TGG CTC AAA GTC TCT AAG CAG ACT AAA GCT TFG ATT GAT AAG CTC CAA	2112
Trp Leu Lys Val Ser Lys Gln Thr Lys Ala Leu Ile Asp Lys Leu Gln	
690 695 700	
AAG CTT GTG TCA TCT GAG GGC AGA TTT AAG AAT CTC AGA GAA GCT TTG	2160
Lys Leu Val Ser Ser Glu Gly Arg Phe Lys Asn Leu Arg Glu Ala Leu	
705 710 715 720	
AAA AAT TGT GAC CCA CCC TGT GTC CCT TAC CTG GGG ATG TAC CTC ACC	2208
Lys Asn Cys Asp Pro Pro Cys Val Pro Tyr Leu Gly Met Tyr Leu Thr	
725 730 735	
GAC CTG GCC TTC ATC GAG GAG GGG ACG CCC AAT TAC ACG GAA GAC GGC	2256
Asp Leu Ala Phe Ile Glu Glu Gly Thr Pro Asn Tyr Thr Glu Asp Gly	
740 745 750	
CTG GTC AAC TTC TCC AAG ATG AGG ATG ATA TCC CAT ATT ATC CGA GAG	2304
Leu Val Asn Phe Ser Lys Met Arg Met Ile Ser His Ile Ile Arg Glu	
755 760 765	
ATT CGC CAG TTT CAA CAA ACT GCC TAC AAA ATA GAG CAC CAA GCA AAG	2352
Ile Arg Gln Phe Gln Gln Thr Ala Tyr Lys Ile Glu His Gln Ala Lys	
770 775 780	
GTA ACG CAA TAT TTA CTG GAC CAA TCT TTT GTA ATG GAT GAA GAA AGC	2400
Val Thr Gln Tyr Leu Leu Asp Gln Ser Phe Val Met Asp Glu Glu Ser	
785 790 795 800	
CTC TAC GAG TCT TCT CTC CGA ATA GAA CCA AAA CTC CCC ACC TGAAGCTGTG	2452
Leu Tyr Glu Ser Ser Leu Arg Ile Glu Pro Lys Leu Pro Thr	
805 810 815	
CCCAGCCCAG ACCCAGCTGC TCCCAGGGAC ATGTGCTAGA TGATACTGTA CATATTCGTT	2512
TGGTTTCACT GGATTTTCTT CTTCACTATG TGCTTCTCCA AGAATACAAA TCGTCCTTGT	2572
TCTTAGATTC CTGTAGAACC GGAATATGAA TTTCTGCACC GTTTCAGACT TCGCCCACCC	2632
ATCCCTCCCC TCGCCCGAAT	2652

(2) INFORMAÇÃO PARA A SEQ ID N°: 2:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 814 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 2:

Gly Asp Gly Cys Lys Ile Leu Leu Asp Thr Ser Gln Thr Phe Val Arg  
1 5 10 15  
Gln Gly Ser Leu Ile Gln Val Pro Met Ser Glu Lys Gly Lys Ile Thr  
20 25 30  
Arg Gly Arg Leu Gly Ser Leu Ser Leu Lys Lys Glu Gly Glu Arg Gln  
35 40 45  
Cys Phe Leu Phe Ser Lys His Leu Ile Ile Cys Thr Arg Gly Ser Gly  
50 55 60  
Gly Lys Leu His Leu Thr Lys Asn Gly Val Ile Ser Leu Ile Asp Cys  
65 70 75 80  
Thr Leu Leu Glu Glu Pro Glu Ser Thr Glu Glu Glu Ala Lys Gly Ser  
85 90 95  
Gly Gln Asp Ile Asp His Leu Asp Phe Lys Ile Gly Val Glu Pro Lys  
100 105 110  
Asp Ser Pro Pro Phe Thr Val Ile Leu Val Ala Ser Ser Arg Gln Glu  
115 120 125  
Lys Ala Ala Trp Thr Ser Asp Ile Ser Gln Cys Val Gly Asn Ile Arg  
130 135 140  
Cys Asn Gly Leu Met Met Lys Pro Phe Glu Glu Asn Ser Lys Val Thr  
145 150 155 160  
Val Pro Gln Met Ile Lys Ser Asp Ala Ser Leu Tyr Cys Asp Asp Val  
165 170 175  
Asp Ile Arg Phe Ser Lys Thr Met Asn Ser Cys Lys Val Leu Gln Ile  
180 185 190  
Ala Tyr Ala Ser Val Glu Arg Leu Leu Glu Arg Leu Thr Asp Leu Arg  
195 200 205

Phe Leu Ser Ile Asp Phe Leu Asn Thr Phe Leu His Ser Tyr Arg Val  
210 215 220  
Phe Thr Thr Ala Ile Val Val Leu Asp Lys Leu Ile Thr Ile Tyr Lys  
225 230 235 240  
Lys Pro Ile Ser Ala Ile Pro Ala Arg Ser Leu Glu Leu Leu Phe Ala  
245 250 255  
Ser Gly Gln Asn Asn Lys Leu Leu Tyr Gly Glu Pro Pro Lys Ser Pro  
260 265 270  
Arg Ala Thr Arg Lys Phe Ser Ser Pro Pro Pro Leu Ser Ile Thr Lys  
275 280 285  
Thr Ser Ser Pro Ser Arg Arg Arg Lys Leu Ser Leu Asn Ile Pro Ile  
290 295 300  
Ile Thr Gly Gly Lys Ala Leu Asp Leu Ala Ala Leu Ser Cys Asn Ser  
305 310 315 320  
Asn Gly Tyr Thr Ser Met Tyr Ser Ala Met Ser Pro Phe Ser Lys Ala  
325 330 335  
Thr Leu Asp Thr Ser Lys Leu Tyr Val Ser Ser Ser Phe Thr Asn Lys  
340 345 350  
Ile Pro Asp Glu Gly Asp Thr Thr Pro Glu Lys Pro Glu Asp Pro Ser  
355 360 365  
Ala Leu Ser Lys Gln Ser Ser Glu Val Ser Met Arg Glu Glu Ser Asp  
370 375 380  
Ile Asp Gln Asn Gln Ser Asp Asp Gly Asp Thr Glu Thr Ser Pro Thr  
385 390 395 400  
Lys Ser Pro Thr Thr Pro Lys Ser Val Lys Asn Lys Asn Ser Ser Glu  
405 410 415  
Phe Pro Leu Phe Ser Tyr Asn Asn Gly Val Val Met Thr Ser Cys Arg  
420 425 430  
Glu Leu Asp Asn Asn Arg Ser Ala Leu Ser Ala Ala Ser Ala Phe Ala  
435 440 445  
Ile Ala Thr Ala Gly Ala Asn Glu Gly Thr Pro Asn Lys Glu Lys Tyr  
450 455 460  
Arg Arg Met Ser Leu Ala Ser Ala Gly Phe Pro Pro Asp Gln Arg Asn  
465 470 475 480  
Gly Asp Lys Glu Phe Val Ile Arg Arg Ala Ala Thr Asn Arg Val Leu  
485 490 495  
Asn Val Leu Arg His Trp Val Ser Lys His Ser Gln Asp Phe Glu Thr  
500 505 510  
Asn Asp Glu Leu Lys Cys Lys Val Ile Gly Phe Leu Glu Glu Val Met  
515 520 525  
His Asp Pro Glu Leu Leu Thr Gln Glu Arg Lys Ala Ala Ala Asn Ile  
530 535 540

Ile Arg Thr Leu Thr Gln Glu Asp Pro Gly Asp Asn Gln Ile Thr Leu  
 545 550 555 560  
 Glu Glu Ile Thr Gln Met Ala Glu Gly Val Lys Ala Glu Pro Phe Glu  
 565 570 575  
 Asn His Ser Ala Leu Glu Ile Ala Glu Gln Leu Thr Leu Leu Asp His  
 580 585 590  
 Leu Val Phe Lys Lys Ile Pro Tyr Glu Glu Phe Phe Gly Gln Gly Trp  
 595 600 605  
 Met Lys Leu Glu Lys Asn Glu Arg Thr Pro Tyr Ile Met Lys Thr Thr  
 610 615 620  
 Lys His Phe Asn Asp Ile Ser Asn Leu Ile Ala Ser Glu Ile Ile Arg  
 625 630 635 640  
 Asn Glu Asp Ile Asn Ala Arg Val Ser Ala Ile Glu Lys Trp Val Ala  
 645 650 655  
 Val Ala Asp Ile Cys Arg Cys Leu His Asn Tyr Asn Ala Val Leu Glu  
 660 665 670  
 Ile Thr Ser Ser Met Asn Arg Ser Ala Ile Phe Arg Leu Lys Lys Thr  
 675 680 685  
 Trp Leu Lys Val Ser Lys Gln Thr Lys Ala Leu Ile Asp Lys Leu Gln  
 690 695 700  
 Lys Leu Val Ser Ser Glu Gly Arg Phe Lys Asn Leu Arg Glu Ala Leu  
 705 710 715 720  
 Lys Asn Cys Asp Pro Pro Cys Val Pro Tyr Leu Gly Met Tyr Leu Thr  
 725 730 735  
 Asp Leu Ala Phe Ile Glu Glu Gly Thr Pro Asn Tyr Thr Glu Asp Gly  
 740 745 750  
 Leu Val Asn Phe Ser Lys Met Arg Met Ile Ser His Ile Ile Arg Glu  
 755 760 765  
 Ile Arg Gln Phe Gln Gln Thr Ala Tyr Lys Ile Glu His Gln Ala Lys  
 770 775 780  
 Val Thr Gln Tyr Leu Leu Asp Gln Ser Phe Val Met Asp Glu Glu Ser  
 785 790 795 800  
 Leu Tyr Glu Ser Ser Leu Arg Ile Glu Pro Lys Leu Pro Thr  
 805 810

(2) INFORMAÇÃO PARA A SEQ ID N°: 3:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 666 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 3:

Met	Met	Lys	Pro	Phe	Glu	Glu	Asn	Ser	Lys	Val	Thr	Val	Pro	Gln	Met	
1				5					10					15		
Ile	Lys	Ser	Asp	Ala	Ser	Leu	Tyr	Cys	Asp	Asp	Val	Asp	Ile	Arg	Phe	
			20					25					30			
Ser	Lys	Thr	Met	Asn	Ser	Cys	Lys	Val	Leu	Gln	Ile	Ala	Tyr	Ala	Ser	
		35					40					45				
Val	Glu	Arg	Leu	Leu	Glu	Arg	Leu	Thr	Asp	Leu	Arg	Phe	Leu	Ser	Ile	
	50					55					60					
Asp	Phe	Leu	Asn	Thr	Phe	Leu	His	Ser	Tyr	Arg	Val	Phe	Thr	Thr	Ala	
65					70					75					80	
Ile	Val	Val	Leu	Asp	Lys	Leu	Ile	Thr	Ile	Tyr	Lys	Lys	Pro	Ile	Ser	
				85					90					95		
Ala	Ile	Pro	Ala	Arg	Ser	Leu	Glu	Leu	Leu	Phe	Ala	Ser	Gly	Gln	Asn	
			100					105					110			
Asn	Lys	Leu	Leu	Tyr	Gly	Glu	Pro	Pro	Lys	Ser	Pro	Arg	Ala	Thr	Arg	
		115					120					125				
Lys	Phe	Ser	Ser	Pro	Pro	Pro	Leu	Ser	Ile	Thr	Lys	Thr	Ser	Ser	Pro	
	130					135					140					
Ser	Arg	Arg	Arg	Lys	Leu	Ser	Leu	Asn	Ile	Pro	Ile	Ile	Thr	Gly	Gly	
145					150					155					160	
Lys	Ala	Leu	Asp	Leu	Ala	Ala	Leu	Ser	Cys	Asn	Ser	Asn	Gly	Tyr	Thr	
				165					170					175		
Ser	Met	Tyr	Ser	Ala	Met	Ser	Pro	Phe	Ser	Lys	Ala	Thr	Leu	Asp	Thr	
			180					185					190			

Ser Lys Leu Tyr Val Ser Ser Ser Phe Thr Asn Lys Ile Pro Asp Glu  
 195 200 205

Gly Asp Thr Thr Pro Glu Lys Pro Glu Asp Pro Ser Ala Leu Ser Lys  
 210 215 220

Gln Ser Ser Glu Val Ser Met Arg Glu Glu Ser Asp Ile Asp Gln Asn  
 225 230 235 240

Gln Ser Asp Asp Gly Asp Thr Glu Thr Ser Pro Thr Lys Ser Pro Thr  
 245 250 255

Thr Pro Lys Ser Val Lys Asn Lys Asn Ser Ser Glu Phe Pro Leu Phe  
 260 265 270

Ser Tyr Asn Asn Gly Val Val Met Thr Ser Cys Arg Glu Leu Asp Asn  
 275 280 285

Asn Arg Ser Ala Leu Ser Ala Ala Ser Ala Phe Ala Ile Ala Thr Ala  
 290 295 300

Gly Ala Asn Glu Gly Thr Pro Asn Lys Glu Lys Tyr Arg Arg Met Ser  
 305 310 315 320

Leu Ala Ser Ala Gly Phe Pro Pro Asp Gln Arg Asn Gly Asp Lys Glu  
 325 330 335

Phe Val Ile Arg Arg Ala Ala Thr Asn Arg Val Leu Asn Val Leu Arg  
 340 345 350

His Trp Val Ser Lys His Ser Gln Asp Phe Glu Thr Asn Asp Glu Leu  
 355 360 365

Lys Cys Lys Val Ile Gly Phe Leu Glu Glu Val Met His Asp Pro Glu  
 370 375 380

Leu Leu Thr Gln Glu Arg Lys Ala Ala Ala Asn Ile Ile Arg Thr Leu  
 385 390 395 400

Thr Gln Glu Asp Pro Gly Asp Asn Gln Ile Thr Leu Glu Glu Ile Thr  
 405 410 415

Gln Met Ala Glu Gly Val Lys Ala Glu Pro Phe Glu Asn His Ser Ala  
 420 425 430

Leu Glu Ile Ala Glu Gln Leu Thr Leu Leu Asp His Leu Val Phe Lys  
 435 440 445

Lys Ile Pro Tyr Glu Glu Phe Phe Gly Gln Gly Trp Met Lys Leu Glu  
 450 455 460

Lys Asn Glu Arg Thr Pro Tyr Ile Met Lys Thr Thr Lys His Phe Asn  
 465 470 475 480

Asp Ile Ser Asn Leu Ile Ala Ser Glu Ile Ile Arg Asn Glu Asp Ile  
 485 490 495

Asn Ala Arg Val Ser Ala Ile Glu Lys Trp Val Ala Val Ala Asp Ile  
 500 505 510  
 Cys Arg Cys Leu His Asn Tyr Asn Ala Val Leu Glu Ile Thr Ser Ser  
 515 520 525  
 Met Asn Arg Ser Ala Ile Phe Arg Leu Lys Lys Thr Trp Leu Lys Val  
 530 535 540  
 Ser Lys Gln Thr Lys Ala Leu Ile Asp Lys Leu Gln Lys Leu Val Ser  
 545 550 555 560  
 Ser Glu Gly Arg Phe Lys Asn Leu Arg Glu Ala Leu Lys Asn Cys Asp  
 565 570 575  
 Pro Pro Cys Val Pro Tyr Leu Gly Met Tyr Leu Thr Asp Leu Ala Phe  
 580 585 590  
 Ile Glu Glu Gly Thr Pro Asn Tyr Thr Glu Asp Gly Leu Val Asn Phe  
 595 600 605  
 Ser Lys Met Arg Met Ile Ser His Ile Ile Arg Glu Ile Arg Gln Phe  
 610 615 620  
 Gln Gln Thr Ala Tyr Lys Ile Glu His Gln Ala Lys Val Thr Gln Tyr  
 625 630 635 640  
 Leu Leu Asp Gln Ser Phe Val Met Asp Glu Glu Ser Leu Tyr Glu Ser  
 645 650 655  
 Ser Leu Arg Ile Glu Pro Lys Leu Pro Thr  
 660 665

(2) INFORMAÇÃO PARA A SEQ ID N°: 4:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 489 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 4:

Met Tyr Ser Ala Met Ser Pro Phe Ser Lys Ala Thr Leu Asp Thr Ser  
1 5 10 15  
Lys Leu Tyr Val Ser Ser Ser Phe Thr Asn Lys Ile Pro Asp Glu Gly  
20 25 30  
Asp Thr Thr Pro Glu Lys Pro Glu Asp Pro Ser Ala Leu Ser Lys Gln  
35 40 45  
Ser Ser Glu Val Ser Met Arg Glu Glu Ser Asp Ile Asp Gln Asn Gln  
50 55 60  
Ser Asp Asp Gly Asp Thr Glu Thr Ser Pro Thr Lys Ser Pro Thr Thr  
65 70 75 80  
Pro Lys Ser Val Lys Asn Lys Asn Ser Ser Glu Phe Pro Leu Phe Ser  
85 90 95  
Tyr Asn Asn Gly Val Val Met Thr Ser Cys Arg Glu Leu Asp Asn Asn  
100 105 110  
Arg Ser Ala Leu Ser Ala Ala Ser Ala Phe Ala Ile Ala Thr Ala Gly  
115 120 125  
Ala Asn Glu Gly Thr Pro Asn Lys Glu Lys Tyr Arg Arg Met Ser Leu  
130 135 140  
Ala Ser Ala Gly Phe Pro Pro Asp Gln Arg Asn Gly Asp Lys Glu Phe  
145 150 155 160  
Val Ile Arg Arg Ala Ala Thr Asn Arg Val Leu Asn Val Leu Arg His  
165 170 175  
Trp Val Ser Lys His Ser Gln Asp Phe Glu Thr Asn Asp Glu Leu Lys  
180 185 190  
Cys Lys Val Ile Gly Phe Leu Glu Glu Val Met His Asp Pro Glu Leu  
195 200 205  
Leu Thr Gln Glu Arg Lys Ala Ala Ala Asn Ile Ile Arg Thr Leu Thr  
210 215 220  
Gln Glu Asp Pro Gly Asp Asn Gln Ile Thr Leu Glu Glu Ile Thr Gln  
225 230 235 240  
Met Ala Glu Gly Val Lys Ala Glu Pro Phe Glu Asn His Ser Ala Leu  
245 250 255  
Glu Ile Ala Glu Gln Leu Thr Leu Leu Asp His Leu Val Phe Lys Lys  
260 265 270

Ile Pro Tyr Glu Glu Phe Phe Gly Gln Gly Trp Met Lys Leu Glu Lys  
 275 280 285  
 Asn Glu Arg Thr Pro Tyr Ile Met Lys Thr Thr Lys His Phe Asn Asp  
 290 295 300  
 Ile Ser Asn Leu Ile Ala Ser Glu Ile Ile Arg Asn Glu Asp Ile Asn  
 305 310 315 320  
 Ala Arg Val Ser Ala Ile Glu Lys Trp Val Ala Val Ala Asp Ile Cys  
 325 330 335  
 Arg Cys Leu His Asn Tyr Asn Ala Val Leu Glu Ile Thr Ser Ser Met  
 340 345 350  
 Asn Arg Ser Ala Ile Phe Arg Leu Lys Lys Thr Trp Leu Lys Val Ser  
 355 360 365  
 Lys Gln Thr Lys Ala Leu Ile Asp Lys Leu Gln Lys Leu Val Ser Ser  
 370 375 380  
 Glu Gly Arg Phe Lys Asn Leu Arg Glu Ala Leu Lys Asn Cys Asp Pro  
 385 390 395 400  
 Pro Cys Val Pro Tyr Leu Gly Met Tyr Leu Thr Asp Leu Ala Phe Ile  
 405 410 415  
 Glu Glu Gly Thr Pro Asn Tyr Thr Glu Asp Gly Leu Val Asn Phe Ser  
 420 425 430  
 Lys Met Arg Met Ile Ser His Ile Ile Arg Glu Ile Arg Gln Phe Gln  
 435 440 445  
 Gln Thr Ala Tyr Lys Ile Glu His Gln Ala Lys Val Thr Gln Tyr Leu  
 450 455 460  
 Leu Asp Gln Ser Phe Val Met Asp Glu Glu Ser Leu Tyr Glu Ser Ser  
 465 470 475 480  
 Leu Arg Ile Glu Pro Lys Leu Pro Thr  
 485

(2) INFORMAÇÃO PARA A SEQ ID N°: 5:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 1092 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: dupla

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(iii) HIPOTÉTICA: NÃO

(iii) ANTI-SENTIDO: NÃO

(ix) CARACTERÍSTICA ADICIONAL:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 1..1092

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 5:

ATT	ACT	AAA	ATA	ATC	CAA	AGG	AAA	AAA	ATT	GCA	AGA	GAC	AAT	GGA	CCA	48
Ile	Thr	Lys	Ile	Ile	Gln	Arg	Lys	Lys	Ile	Ala	Arg	Asp	Asn	Gly	Pro	
1				5					10					15		
GGT	CAT	AAT	ATT	ACA	TTT	CAG	AGT	TCA	CCT	CCC	ACA	GTT	GAG	TGG	CAT	96
Gly	His	Asn	Ile	Thr	Phe	Gln	Ser	Ser	Pro	Pro	Thr	Val	Glu	Trp	His	
			20					25					30			
ATA	AGC	AGA	CCT	GGG	CAC	ATA	GAG	ACT	TTT	GAC	CTG	CTC	ACC	TTA	CAC	144
Ile	Ser	Arg	Pro	Gly	His	Ile	Glu	Thr	Phe	Asp	Leu	Leu	Thr	Leu	His	
		35					40					45				
CCA	ATA	GAA	ATT	GCT	CGA	CAA	CTC	ACT	TTA	CTT	GAT	TCA	GAT	CTA	TAC	192
Pro	Ile	Glu	Ile	Ala	Arg	Gln	Leu	Thr	Leu	Leu	Asp	Ser	Asp	Leu	Tyr	
	50					55					60					
CGA	GCT	GTA	CAG	CCA	TCA	GAT	TTA	GTT	GGA	AGT	GTG	TGG	ACA	AAA	GAA	240
Arg	Ala	Val	Gln	Pro	Ser	Asp	Leu	Val	Gly	Ser	Val	Trp	Thr	Lys	Glu	
	65				70					75					80	
GAC	AAA	GAA	ATT	AAC	TCT	CCT	AAT	CTT	CTG	AAA	ATG	ATT	CGA	CAT	ACC	288
Asp	Lys	Glu	Ile	Asn	Ser	Pro	Asn	Leu	Leu	Lys	Met	Ile	Arg	His	Thr	
				85					90					95		
ACC	AAC	CTC	ACT	CTG	TGG	TTT	GAG	AAA	TGT	ATT	GTA	GAA	ACT	GAA	AAT	336
Thr	Asn	Leu	Thr	Leu	Trp	Phe	Glu	Lys	Cys	Ile	Val	Glu	Thr	Glu	Asn	
			100					105					110			
TTA	GAA	GAA	AGA	GTA	GCT	GTG	GTG	AGT	CGA	ATT	ATT	GAG	ATT	CTA	CAA	384
Leu	Glu	Glu	Arg	Val	Ala	Val	Val	Ser	Arg	Ile	Ile	Glu	Ile	Leu	Gln	
		115					120					125				

GTC	TTT	CAA	GAG	TTG	AAC	AAC	TTT	AAT	GGG	GTC	CTT	GAG	GTT	GTC	AGT	432
Val	Phe	Gln	Glu	Leu	Asn	Asn	Phe	Asn	Gly	Val	Leu	Glu	Val	Val	Ser	
	130					135					140					
GCT	ATG	AAT	TCC	TCA	CCT	GTT	TAC	AGA	CTA	GAC	CAC	ACA	TTT	GAG	CAA	480
Ala	Met	Asn	Ser	Ser	Pro	Val	Tyr	Arg	Leu	Asp	His	Thr	Phe	Glu	Gln	
	145				150					155					160	
ATA	CCA	AGT	CGC	CAG	AAG	AAA	ATT	TTA	GAA	GAA	GCT	CAT	GAA	TTG	AGT	528
Ile	Pro	Ser	Arg	Gln	Lys	Lys	Ile	Leu	Glu	Glu	Ala	His	Glu	Leu	Ser	
				165					170					175		
GAA	GAT	CAC	TAT	AAG	AAA	TAT	TTG	GCA	AAA	CTC	AGG	TCT	ATT	AAT	CCA	576
Glu	Asp	His	Tyr	Lys	Lys	Tyr	Leu	Ala	Lys	Leu	Arg	Ser	Ile	Asn	Pro	
		180						185					190			
CCA	TGT	GTG	CCT	TTC	TTT	GGA	ATT	TAT	CTA	CAT	AAT	ATC	TTG	AAA	ACA	624
Pro	Cys	Val	Pro	Phe	Phe	Gly	Ile	Tyr	Leu	His	Asn	Ile	Leu	Lys	Thr	
		195				200						205				
GAA	GAA	GGC	AAC	CCT	GAG	GTC	CTA	AAA	AGA	CAT	GGA	AAA	GAG	CTT	ATA	672
Glu	Glu	Gly	Asn	Pro	Glu	Val	Leu	Lys	Arg	His	Gly	Lys	Glu	Leu	Ile	
	210					215					220					
AAC	TTT	AGC	AAA	AGG	AGG	AAA	GTA	GCA	GAA	ATA	ACA	GGA	GAG	ATC	CAG	720
Asn	Phe	Ser	Lys	Arg	Arg	Lys	Val	Ala	Glu	Ile	Thr	Gly	Glu	Ile	Gln	
	225				230					235					240	
CAG	TAC	CAA	AAT	CAG	CCT	TAC	TGT	TTA	CGA	GTA	GAA	TCA	GAT	ATC	AAA	768
Gln	Tyr	Gln	Asn	Gln	Pro	Tyr	Cys	Leu	Arg	Val	Glu	Ser	Asp	Ile	Lys	
				245					250					255		
AGG	TTC	TTT	GAA	AAC	TTG	AAT	CCG	ATG	GGA	AAT	AGC	ATG	GAG	AGG	GAA	816
Arg	Phe	Phe	Glu	Asn	Leu	Asn	Pro	Met	Gly	Asn	Ser	Met	Glu	Arg	Glu	
			260					265					270			
TTT	ACA	GAT	TAT	CTT	TTC	AAC	AAA	TCC	CTA	GAA	ATA	GAA	CCA	CGA	AAC	864
Phe	Thr	Asp	Tyr	Leu	Phe	Asn	Lys	Ser	Leu	Glu	Ile	Glu	Pro	Arg	Asn	
		275					280					285				
CCT	AAG	CCT	CTC	CCA	AGA	TTT	CCA	AAA	AAA	TAT	ACG	TAT	CCC	CTA	AAA	912
Pro	Lys	Pro	Leu	Pro	Arg	Phe	Pro	Lys	Lys	Tyr	Thr	Tyr	Pro	Leu	Lys	
	290					295					300					
TCT	CCT	GGT	GTC	CGG	CCA	TCA	AAC	CCA	AGA	CCG	GGT	ACC	ATG	AGG	ATC	960
Ser	Pro	Gly	Val	Arg	Pro	Ser	Asn	Pro	Arg	Pro	Gly	Thr	Met	Arg	Ile	
	305				310					315					320	
CCC	ACC	CCT	CTA	CAG	CAG	GAA	CCA	CGA	AAA	ATA	AGT	TAT	AGT	AGA	ATA	1008
Pro	Thr	Pro	Leu	Gln	Gln	Glu	Pro	Arg	Lys	Ile	Ser	Tyr	Ser	Arg	Ile	
				325					330					335		
CCA	GAG	TCA	GAG	ACA	GAG	AGT	ACT	GCT	AGT	GCA	CCT	AAT	TCA	CCA	AGG	1056
Pro	Glu	Ser	Glu	Thr	Glu	Ser	Thr	Ala	Ser	Ala	Pro	Asn	Ser	Pro	Arg	
			340					345					350			
ACA	CCT	CTA	ACA	CCT	CCA	CCT	GCA	TCA	GGA	ACA	TCA					1092
Thr	Pro	Leu	Thr	Pro	Pro	Pro	Ala	Ser	Gly	Thr	Ser					
		355					360									

(2) INFORMAÇÃO PARA A SEQ ID N°: 6:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 364 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 6:

Ile Thr Lys Ile Ile Gln Arg Lys Lys Ile Ala Arg Asp Asn Gly Pro  
1 5 10 15  
Gly His Asn Ile Thr Phe Gln Ser Ser Pro Pro Thr Val Glu Trp His  
20 25 30  
Ile Ser Arg Pro Gly His Ile Glu Thr Phe Asp Leu Leu Thr Leu His  
35 40 45  
Pro Ile Glu Ile Ala Arg Gln Leu Thr Leu Leu Asp Ser Asp Leu Tyr  
50 55 60  
Arg Ala Val Gln Pro Ser Asp Leu Val Gly Ser Val Trp Thr Lys Glu  
65 70 75 80  
Asp Lys Glu Ile Asn Ser Pro Asn Leu Leu Lys Met Ile Arg His Thr  
85 90 95  
Thr Asn Leu Thr Leu Trp Phe Glu Lys Cys Ile Val Glu Thr Glu Asn  
100 105 110  
Leu Glu Glu Arg Val Ala Val Val Ser Arg Ile Ile Glu Ile Leu Gln  
115 120 125  
Val Phe Gln Glu Leu Asn Asn Phe Asn Gly Val Leu Glu Val Val Ser  
130 135 140  
Ala Met Asn Ser Ser Pro Val Tyr Arg Leu Asp His Thr Phe Glu Gln  
145 150 155 160  
Ile Pro Ser Arg Gln Lys Lys Ile Leu Glu Glu Ala His Glu Leu Ser  
165 170 175  
Glu Asp His Tyr Lys Lys Tyr Leu Ala Lys Leu Arg Ser Ile Asn Pro  
180 185 190

Pro Cys Val Pro Phe Phe Gly Ile Tyr Leu His Asn Ile Leu Lys Thr  
 195 200 205  
 Glu Glu Gly Asn Pro Glu Val Leu Lys Arg His Gly Lys Glu Leu Ile  
 210 215 220  
 Asn Phe Ser Lys Arg Arg Lys Val Ala Glu Ile Thr Gly Glu Ile Gln  
 225 230 235 240  
 Gln Tyr Gln Asn Gln Pro Tyr Cys Leu Arg Val Glu Ser Asp Ile Lys  
 245 250 255  
 Arg Phe Phe Glu Asn Leu Asn Pro Met Gly Asn Ser Met Glu Arg Glu  
 260 265 270  
 Phe Thr Asp Tyr Leu Phe Asn Lys Ser Leu Glu Ile Glu Pro Arg Asn  
 275 280 285  
 Pro Lys Pro Leu Pro Arg Phe Pro Lys Lys Tyr Thr Tyr Pro Leu Lys  
 290 295 300  
 Ser Pro Gly Val Arg Pro Ser Asn Pro Arg Pro Gly Thr Met Arg Ile  
 305 310 315 320  
 Pro Thr Pro Leu Gln Gln Glu Pro Arg Lys Ile Ser Tyr Ser Arg Ile  
 325 330 335  
 Pro Glu Ser Glu Thr Glu Ser Thr Ala Ser Ala Pro Asn Ser Pro Arg  
 340 345 350  
 Thr Pro Leu Thr Pro Pro Pro Ala Ser Gly Thr Ser  
 355 360

(2) INFORMAÇÃO PARA A SEQ ID N°: 7:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 1956 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: dupla
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(iii) HIPOTÉTICA: NÃO

(iii) ANTI-SENTIDO: NÃO

(ix) CARACTERÍSTICA ADICIONAL:

(A) NOME/CHAVE: CDS

(B) LOCALIZAÇÃO: 1.1956

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 7:

AGG	TTT	GAA	ATC	CCA	GAG	CCA	GAA	CCT	ACA	GAA	GCA	GAT	AAA	CTA	GCA	48
Arg	Phe	Glu	Ile	Pro	Glu	Pro	Glu	Pro	Thr	Glu	Ala	Asp	Lys	Leu	Ala	
1				5					10					15		
CTT	GAG	AAA	GGA	GAA	CAA	CCA	ATC	TCT	GCA	GAT	CTA	AAG	AGG	TTC	AGA	96
Leu	Glu	Lys	Gly	Glu	Gln	Pro	Ile	Ser	Ala	Asp	Leu	Lys	Arg	Phe	Arg	
			20					25					30			
AAG	GAA	TAT	ATC	CAA	CCA	GTA	CAG	CTA	CGG	GTG	TTG	AAC	GTG	CAG	CGG	144
Lys	Glu	Tyr	Ile	Gln	Pro	Val	Gln	Leu	Arg	Val	Leu	Asn	Val	Gln	Arg	
		35					40					45				
CAC	TGG	GTT	GAA	CAT	CAC	CCC	CAT	GAC	TTT	GAA	AGA	GAC	TTG	GAA	CTG	192
His	Trp	Val	Glu	His	His	Pro	His	Asp	Phe	Glu	Arg	Asp	Leu	Glu	Leu	
	50					55					60					
CTC	GAA	AGA	CTA	GAA	TCC	TTC	ACC	TCA	AGC	GCT	CAC	AGA	GCG	AAA	GCA	240
Leu	Glu	Arg	Leu	Glu	Ser	Phe	Thr	Ser	Ser	Ala	His	Arg	Ala	Lys	Ala	
65					70					75					80	
ATG	AAG	AAG	TGG	GTA	GAG	AGC	ATC	GCT	AAG	ACC	ATC	AGG	AGG	AAG	AAG	288
Met	Lys	Lys	Trp	Val	Glu	Ser	Ile	Ala	Lys	Thr	Ile	Arg	Arg	Lys	Lys	
				85					90					95		
CAA	GCT	CAG	GCA	AAT	GGA	GTA	AGC	CAT	AAT	ATT	ACC	TTT	GAA	AGT	CCA	336
Gln	Ala	Gln	Ala	Asn	Gly	Val	Ser	His	Asn	Ile	Thr	Phe	Glu	Ser	Pro	
			100					105					110			
CCT	CCA	CCA	ATT	GAA	TGG	CAT	ATC	AGC	AAA	CCA	GGA	CAG	TTT	GAA	ACA	384
Pro	Pro	Pro	Ile	Glu	Trp	His	Ile	Ser	Lys	Pro	Gly	Gln	Phe	Glu	Thr	
		115					120					125				
TTT	GAT	CTC	ATG	ACA	CTT	CAT	CCA	ATA	GAA	ATT	GCA	CGT	CAG	CTG	ACA	432
Phe	Asp	Leu	Met	Thr	Leu	His	Pro	Ile	Glu	Ile	Ala	Arg	Gln	Leu	Thr	
	130					135					140					
CTT	TTG	GAG	TCT	GAT	CTT	TAC	AGG	AAA	GTT	CAA	CCG	TCT	GAA	CTT	GTA	480
Leu	Leu	Glu	Ser	Asp	Leu	Tyr	Arg	Lys	Val	Gln	Pro	Ser	Glu	Leu	Val	
145					150					155					160	
GGG	AGT	GTG	TGG	ACC	AAA	GAA	GAT	AAA	GAA	ATA	AAT	TCT	CCA	AAT	TTA	528
Gly	Ser	Val	Trp	Thr	Lys	Glu	Asp	Lys	Glu	Ile	Asn	Ser	Pro	Asn	Leu	
				165					170					175		
TTA	AAA	ATG	ATT	CGC	CAT	ACC	ACA	AAT	CTC	ACC	CTC	TGG	TTT	GAA	AAA	576
Leu	Lys	Met	Ile	Arg	His	Thr	Thr	Asn	Leu	Thr	Leu	Trp	Phe	Glu	Lys	
			180					185					190			
TGC	ATT	GTG	GAA	GCA	GAA	AAT	TTT	GAA	GAA	CGG	GTG	GCA	GTA	CTA	AGT	624
Cys	Ile	Val	Glu	Ala	Glu	Asn	Phe	Glu	Glu	Arg	Val	Ala	Val	Leu	Ser	
		195					200					205				
AGA	ATT	ATA	GAA	ATT	CTG	CAA	GTT	TTT	CGA	GAT	TTG	AAT	AAT	TTC	AAT	672

Arg	Ile	Ile	Glu	Ile	Leu	Gln	Val	Phe	Arg	Asp	Leu	Asn	Asn	Phe	Asn		
	210					215					220						
GGC	GTA	TTG	GAG	ATA	GTC	AGT	GCA	GTA	AAT	TCA	GTG	TCA	GTA	TAC	AGA	720	
Gly	Val	Leu	Glu	Ile	Val	Ser	Ala	Val	Asn	Ser	Val	Ser	Val	Tyr	Arg		240
	225				230					235							
CTA	GAC	CAT	ACC	TTT	GAG	GCA	CTG	CAG	GAA	AGG	AAA	AGG	AAA	ATT	TTG	768	
Leu	Asp	His	Thr	Phe	Glu	Ala	Leu	Gln	Glu	Arg	Lys	Arg	Lys	Ile	Leu		255
				245					250								
GAC	GAA	GCT	GTG	GAA	TTA	AGT	CAA	GAT	CAC	TTT	AAA	AAA	TAC	CTA	GTA	816	
Asp	Glu	Ala	Val	Glu	Leu	Ser	Gln	Asp	His	Phe	Lys	Lys	Tyr	Leu	Val		270
			260					265									
AAA	CTT	AAG	TCA	ATC	AAT	CCA	CCT	TGT	GTG	CCT	TTT	TTT	GGA	ATA	TAT	864	
Lys	Leu	Lys	Ser	Ile	Asn	Pro	Pro	Cys	Val	Pro	Phe	Phe	Gly	Ile	Tyr		285
		275					280										
TTA	ACA	AAT	ATT	CTG	AAG	ACC	GAA	GAA	GGG	AAT	AAT	GAT	TTT	TTA	AAA	912	
Leu	Thr	Asn	Ile	Leu	Lys	Thr	Glu	Glu	Gly	Asn	Asn	Asp	Phe	Leu	Lys		300
	290					295											
AAG	AAA	GGG	AAA	GAT	TTA	ATC	AAT	TTC	AGT	AAG	AGG	AGG	AAA	GTA	GCT	960	
Lys	Lys	Gly	Lys	Asp	Leu	Ile	Asn	Phe	Ser	Lys	Arg	Arg	Lys	Val	Ala		320
	305				310					315							
GAA	ATT	ACT	GGA	GAA	ATT	CAG	CAG	TAT	CAG	AAT	CAG	CCG	TAC	TGC	CTA	1008	
Glu	Ile	Thr	Gly	Glu	Ile	Gln	Gln	Tyr	Gln	Asn	Gln	Pro	Tyr	Cys	Leu		335
				325					330								
CGG	ATA	GAA	CCA	GAT	ATG	AGG	AGA	TTC	TTT	GAA	AAC	CTT	AAC	CCC	ATG	1056	
Arg	Ile	Glu	Pro	Asp	Met	Arg	Arg	Phe	Phe	Glu	Asn	Leu	Asn	Pro	Met		350
			340					345									
GGA	AGT	GCA	TGT	GAA	AAA	GAG	TTT	ACA	GAT	TAT	TTG	TTC	AAC	AAA	AGT	1104	
Gly	Ser	Ala	Cys	Glu	Lys	Glu	Phe	Thr	Asp	Tyr	Leu	Phe	Asn	Lys	Ser		365
		355					360										
TTA	GAA	ATA	GAA	CCA	CGA	AAT	TGT	AAA	CAG	CCA	CCA	CGA	TTT	CCA	CGA	1152	
Leu	Glu	Ile	Glu	Pro	Arg	Asn	Cys	Lys	Gln	Pro	Pro	Arg	Phe	Pro	Arg		380
	370					375											
AAA	AGT	ACA	TTT	GAA	CTA	AAA	GAA	CCA	GGA	ATA	CGA	CCA	AAT	GCA	GGA	1200	
Lys	Ser	Thr	Phe	Glu	Leu	Lys	Glu	Pro	Gly	Ile	Arg	Pro	Asn	Ala	Gly		400
	385				390					395							
CGA	CAT	GGA	GAA	ACA	AGT	GGA	ACA	AGA	GGA	CAT	CCA	ACA	CCT	CTA	GAA	1248	
Arg	His	Gly	Glu	Thr	Ser	Gly	Thr	Arg	Gly	His	Pro	Thr	Pro	Leu	Glu		415
				405					410								
AGA	GAA	CCA	TAT	AAA	ATA	GAA	TTT	GAA	AGA	ATA	GCT	GAA	ACA	GAA	CTA	1296	
Arg	Glu	Pro	Tyr	Lys	Ile	Glu	Phe	Glu	Arg	Ile	Ala	Glu	Thr	Glu	Leu		430
			420				425										
GAA	AGT	ACA	GTA	AGT	GCA	CCA	ACA	AGT	CCA	AAT	ACT	CCC	TCA	ACA	CCA	1344	
Glu	Ser	Thr	Val	Ser	Ala	Pro	Thr	Ser	Pro	Asn	Thr	Pro	Ser	Thr	Pro		445
		435					440										
CCA	GTT	TCA	GCA	TCA	TCA	GAT	CAC	TCA	GTA	TTT	TTA	GAT	GTA	GAT	CTA	1392	
Pro	Val	Ser	Ala	Ser	Ser	Asp	His	Ser	Val	Phe	Leu	Asp	Val	Asp	Leu		460
	450					455					460						

AAC Asn 465	AGT Ser	AGT Ser	CAC His	GGA Gly 470	TCA Ser 470	AAC Asn	ACA Thr	ATC Ile	TTT Phe	GCA Ala 475	CCT Pro	GTG Val	CTA Leu	CTA Leu	CCA Pro 480	1440
AAG Lys	TCC Ser	AAG Lys	TCT Ser	TTC Phe 485	TTT Phe	AGT Ser	TCA Ser	TGT Cys	GGA Gly 490	AGT Ser	TTA Leu	CAT His	AAA Lys	CTA Leu	AGT Ser 495	1488
GAA Glu	GAG Glu	CCC Pro	CTG Leu 500	ATT Ile	CCT Pro	CCT Pro	CCT Pro	CTT Leu 505	CCT Pro	CCT Pro	CGA Arg	AAA Lys	AAG Lys 510	TTT Phe	GAT Asp	1536
CAT His	GAT Asp	GGC Gly 515	TCA Ser	AAT Asn	TCC Ser	AAG Lys	GGA Gly 520	AAT Asn	ATG Met	AAA Lys	TCT Ser	GAT Asp 525	GAT Asp	GAT Asp	CCT Pro	1584
CCT Pro	GCT Ala 530	ATT Ile	CCA Pro	CCG Pro	AGA Arg	CAG Gln 535	CCT Pro	CCT Pro	CCT Pro	CCA Pro	AAG Lys 540	GTA Val	AAA Lys	CCC Pro	AGA Arg	1632
GTT Val 545	CCT Pro	GTT Val	CCT Pro	ACT Thr 550	GGT Gly 550	GCA Ala	TTT Phe	GAT Asp	GGG Gly 555	CCT Pro 555	CTG Leu	CAT His	AGT Ser	CCA Pro	CCT Pro 560	1680
CCG Pro	CCA Pro	CCA Pro	CCA Pro	AGA Arg 565	GAT Asp	CCT Pro	CTT Leu	CCT Pro	GAT Asp 570	ACC Thr	CCT Pro	CCA Pro	CCA Pro	GTT Val 575	CCC Pro	1728
CTT Leu	CGG Arg	CCT Pro	CCA Pro 580	GAA Glu	CAC His	TTT Phe	ATA Ile	AAC Asn 585	TGT Cys	CCA Pro	TTT Phe	AAT Asn 590	CTT Leu	CAG Gln	CCA Pro	1776
CCT Pro	CCA Pro	CTG Leu 595	GGG Gly	CAT His	CTT Leu	CAC His	AGA Arg 600	GAT Asp	TCA Ser	GAC Asp	TGG Trp	CTC Leu	AGA Arg 605	GAC Asp	ATT Ile	1824
AGT Ser	ACG Thr 610	TGT Cys	CCA Pro	AAT Asn	TCG Ser	CCA Pro 615	AGC Ser	ACT Thr	CCT Pro	CCT Pro	AGC Ser 620	ACA Thr	CCC Pro	TCT Ser	CCA Pro	1872
AGG Arg 625	GTA Val	CCG Pro	CGT Arg	CGA Arg	TGC Cys 630	TAT Tyr	GTG Val	CTC Leu	AGT Ser	TCT Ser 635	AGT Ser	CAG Gln	AAT Asn	AAT Asn	CTT Leu 640	1920
GCT Ala	CAT His	CCT Pro	CCA Pro	GCT Ala 645	CCC Pro	CCT Pro	GTT Val	CCA Pro	CCA Pro 650	AGG Arg	GAG Glu					1956

(2) INFORMAÇÃO PARA A SEQ ID N°: 8:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 652 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 8:

Arg Phe Glu Ile Pro Glu Pro Glu Pro Thr Glu Ala Asp Lys Leu Ala  
1 5 10 15  
Leu Glu Lys Gly Glu Gln Pro Ile Ser Ala Asp Leu Lys Arg Phe Arg  
20 25 30  
Lys Glu Tyr Ile Gln Pro Val Gln Leu Arg Val Leu Asn Val Gln Arg  
35 40 45  
His Trp Val Glu His His Pro His Asp Phe Glu Arg Asp Leu Glu Leu  
50 55 60  
Leu Glu Arg Leu Glu Ser Phe Thr Ser Ser Ala His Arg Ala Lys Ala  
65 70 75 80  
Met Lys Lys Trp Val Glu Ser Ile Ala Lys Thr Ile Arg Arg Lys Lys  
85 90 95  
Gln Ala Gln Ala Asn Gly Val Ser His Asn Ile Thr Phe Glu Ser Pro  
100 105 110  
Pro Pro Pro Ile Glu Trp His Ile Ser Lys Pro Gly Gln Phe Glu Thr  
115 120 125  
Phe Asp Leu Met Thr Leu His Pro Ile Glu Ile Ala Arg Gln Leu Thr  
130 135 140  
Leu Leu Glu Ser Asp Leu Tyr Arg Lys Val Gln Pro Ser Glu Leu Val  
145 150 155 160  
Gly Ser Val Trp Thr Lys Glu Asp Lys Glu Ile Asn Ser Pro Asn Leu  
165 170 175  
Leu Lys Met Ile Arg His Thr Thr Asn Leu Thr Leu Trp Phe Glu Lys  
180 185 190  
Cys Ile Val Glu Ala Glu Asn Phe Glu Glu Arg Val Ala Val Leu Ser  
195 200 205  
Arg Ile Ile Glu Ile Leu Gln Val Phe Arg Asp Leu Asn Asn Phe Asn  
210 215 220  
Gly Val Leu Glu Ile Val Ser Ala Val Asn Ser Val Ser Val Tyr Arg  
225 230 235 240  
Leu Asp His Thr Phe Glu Ala Leu Gln Glu Arg Lys Arg Lys Ile Leu  
245 250 255  
Asp Glu Ala Val Glu Leu Ser Gln Asp His Phe Lys Lys Tyr Leu Val  
260 265 270

Lys Leu Lys Ser Ile Asn Pro Pro Cys Val Pro Phe Phe Gly Ile Tyr  
 275 280 285  
 Leu Thr Asn Ile Leu Lys Thr Glu Glu Gly Asn Asn Asp Phe Leu Lys  
 290 295 300  
 Lys Lys Gly Lys Asp Leu Ile Asn Phe Ser Lys Arg Arg Lys Val Ala  
 305 310 315 320  
 Glu Ile Thr Gly Glu Ile Gln Gln Tyr Gln Asn Gln Pro Tyr Cys Leu  
 325 330 335  
 Arg Ile Glu Pro Asp Met Arg Arg Phe Phe Glu Asn Leu Asn Pro Met  
 340 345 350  
 Gly Ser Ala Cys Glu Lys Glu Phe Thr Asp Tyr Leu Phe Asn Lys Ser  
 355 360 365  
 Leu Glu Ile Glu Pro Arg Asn Cys Lys Gln Pro Pro Arg Phe Pro Arg  
 370 375 380  
 Lys Ser Thr Phe Glu Leu Lys Glu Pro Gly Ile Arg Pro Asn Ala Gly  
 385 390 395 400  
 Arg His Gly Glu Thr Ser Gly Thr Arg Gly His Pro Thr Pro Leu Glu  
 405 410 415  
 Arg Glu Pro Tyr Lys Ile Glu Phe Glu Arg Ile Ala Glu Thr Glu Leu  
 420 425 430  
 Glu Ser Thr Val Ser Ala Pro Thr Ser Pro Asn Thr Pro Ser Thr Pro  
 435 440 445  
 Pro Val Ser Ala Ser Ser Asp His Ser Val Phe Leu Asp Val Asp Leu  
 450 455 460  
 Asn Ser Ser His Gly Ser Asn Thr Ile Phe Ala Pro Val Leu Leu Pro  
 465 470 475 480  
 Lys Ser Lys Ser Phe Phe Ser Ser Cys Gly Ser Leu His Lys Leu Ser  
 485 490 495  
 Glu Glu Pro Leu Ile Pro Pro Pro Leu Pro Pro Arg Lys Lys Phe Asp  
 500 505 510  
 His Asp Gly Ser Asn Ser Lys Gly Asn Met Lys Ser Asp Asp Asp Pro  
 515 520 525  
 Pro Ala Ile Pro Pro Arg Gln Pro Pro Pro Pro Lys Val Lys Pro Arg  
 530 535 540  
 Val Pro Val Pro Thr Gly Ala Phe Asp Gly Pro Leu His Ser Pro Pro  
 545 550 555 560  
 Pro Pro Pro Pro Arg Asp Pro Leu Pro Asp Thr Pro Pro Pro Val Pro  
 565 570 575

Leu Arg Pro Pro Glu His Phe Ile Asn Cys Pro Phe Asn Leu Gln Pro  
 580 585 590  
 Pro Pro Leu Gly His Leu His Arg Asp Ser Asp Trp Leu Arg Asp Ile  
 595 600 605  
 Ser Thr Cys Pro Asn Ser Pro Ser Thr Pro Pro Ser Thr Pro Ser Pro  
 610 615 620  
 Arg Val Pro Arg Arg Cys Tyr Val Leu Ser Ser Ser Gln Asn Asn Leu  
 625 630 635 640  
 Ala His Pro Pro Ala Pro Pro Val Pro Pro Arg Glu  
 645 650

(2) INFORMAÇÃO PARA A SEQ ID N°:9:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 41 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADEIA: simples
- (D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(iii) HIPOTÉTICA: NÃO

(iii) ANTI-SENTIDO: NÃO

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°:9:

GATATCGAAT TCCGIGTIYT IAAYGTIYTI MGICAYTGGG T 41

(2) INFORMAÇÃO PARA A SEQ ID N°:10:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

- (A) COMPRIMENTO: 34 pares de bases
- (B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(iii) HIPOTÉTICA: NÃO

(iii) ANTI-SENTIDO: NÃO

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°:10:

AAGCTTGAAT TCCKIMKYTT ISWRAARTTI AKIA

34

(2) INFORMAÇÃO PARA A SEQ ID N°: 11:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 33 bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNc

(iii) HIPOTÉTICA: NÃO

(iii) ANTI-SENTIDO: NÃO

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 11:

ATCHGTNAGG TACATNCCHA GGTAKGGNAC ACA

33

(2) INFORMAÇÃO PARA A SEQ ID N°: 12:

(i) CARACTERÍSTICAS DA SEQUÊNCIA:

(A) COMPRIMENTO: 27 bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADEIA: simples

(D) TOPOLOGIA: linear

(ii) TIPO DE MOLÉCULA: ADNC

(xi) DESCRIÇÃO DA SEQUÊNCIA: SEQ ID N°: 12:

GCNATHTTYC GNCTNAARAA RACNTGG 27

Lisboa, 27 de Agosto de 2007

## REIVINDICAÇÕES

1. Péptido capaz de modular os níveis de permuta do GDP nos complexos p21-GDP, compreendendo as sequências SEQ ID N° 2, 3, 4, 6 ou 8.
2. Péptido de acordo com a reivindicação 1, caracterizado por retardar ou inibir a permuta do GDP no complexo p21-GDP.
3. Péptido de sequência SEQ ID N° 2, 3, 4, 6 ou 8.
4. Anticorpos ou fragmento de anticorpos dirigidos contra um péptido de acordo com qualquer uma das reivindicações 1 a 3.
5. Anticorpos ou fragmento de anticorpos de acordo com a reivindicação 4, caracterizados por serem dirigidos contra a sequência peptídica SEQ ID N° 1.
6. Anticorpos ou fragmento de anticorpos de acordo com a reivindicação 5, caracterizados por possuírem capacidade para inibir, pelo menos parcialmente, a permuta do GDP no complexo p21-GDP.
7. Sequência nucleotídica codificante de um péptido como definido em qualquer uma das reivindicações 1 a 3.
8. Sequência nucleotídica de acordo com a reivindicação 7, caracterizada por ser escolhida a partir de:
  - (a) sequências SEQ ID N° 1, 5 ou 7 e codificam um péptido capaz de modular os níveis de permuta do GDP nos complexos p21-GDP,

(b) cadeia complementar das sequências (a),

(c) sequências derivadas a partir das sequências (a) e (b) como resultado da degenerescência do código genético.

9. Sequência anti-sentido capaz de inibir a produção de péptidos de acordo com a reivindicação 3 e derivados das sequências SEQ ID N° 1, SEQ ID N° 5 ou SEQ ID N° 7 ou a sua cadeia complementar.
10. Utilização de uma sequência de acordo com a reivindicação 8(b) para a detecção *in vitro* da expressão do factor de permuta do GDP ou para a demonstração de anomalias genéticas (*splicing* incorrecto, polimorfismo, mutações pontuais).
11. Composição farmacêutica compreendendo como princípio activo pelo menos um péptido de acordo com uma das reivindicações 1 a 3 ou um anticorpo ou fragmento de anticorpo de acordo com uma das reivindicações 4 a 6 ou uma sequência nucleotídica de acordo com a reivindicação 8.
12. Composição farmacêutica de acordo com a reivindicação 11, destinada a modular a activação das proteínas p21.
13. Composição farmacêutica de acordo com a reivindicação 12, destinada a inibir, pelo menos, parcialmente a activação das proteínas p21.
14. Composição farmacêutica de acordo com a reivindicação 11, destinada ao tratamento de cancros.

15. Utilização de um anticorpo ou fragmento de anticorpo de acordo com uma das reivindicações 4 a 6 e/ou de uma sequência nucleotídica de acordo com a reivindicação 8(b) para a detecção da expressão e/ou de uma sobre-expressão de um factor de permuta do GDP amplificado, mutado ou reorganizado numa amostra biológica.
16. Utilização de um anticorpo ou fragmento de anticorpo de acordo com uma das reivindicações 4 a 6 e/ou de uma sequência nucleotídica de acordo com a reivindicação 8(b) para a tipagem de cancros *in vitro*.
17. Processo de preparação de um péptido de acordo com qualquer uma das reivindicações 1 a 3, caracterizado por se cultivar uma célula contendo uma sequência nucleotídica de acordo com a reivindicação 7 nas condições de expressão da referida sequência e recuperar o péptido produzido.

Lisboa, 27 de Agosto de 2007

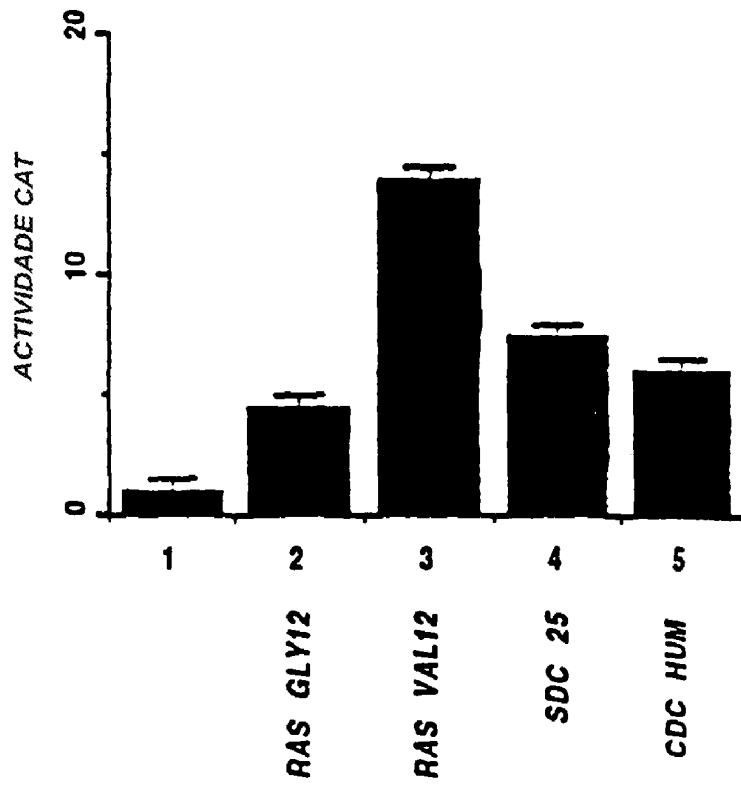


FIGURA 1

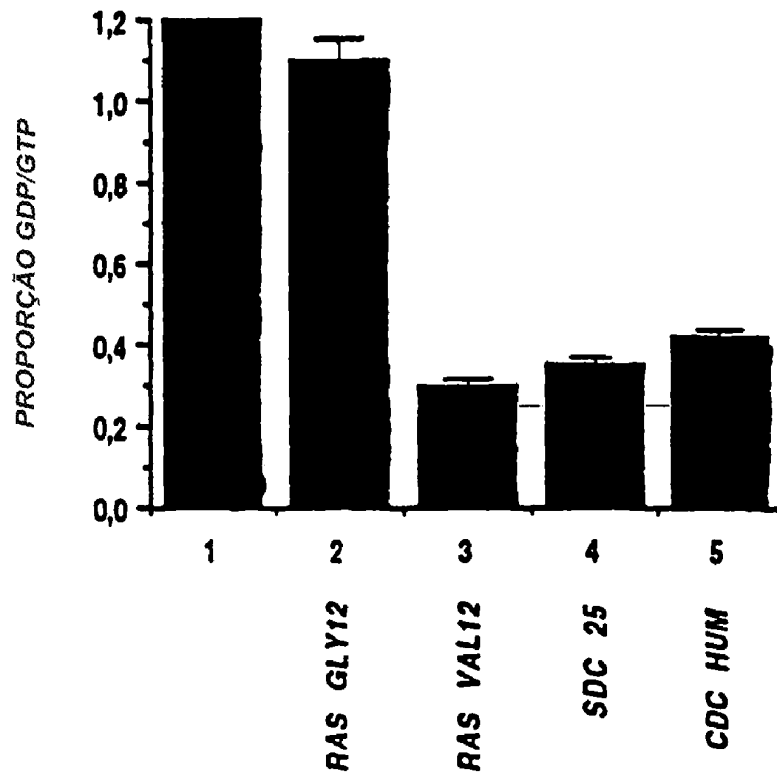


FIGURA 2

## Actividade de permuta da proteína de fusão GST-CDCHum

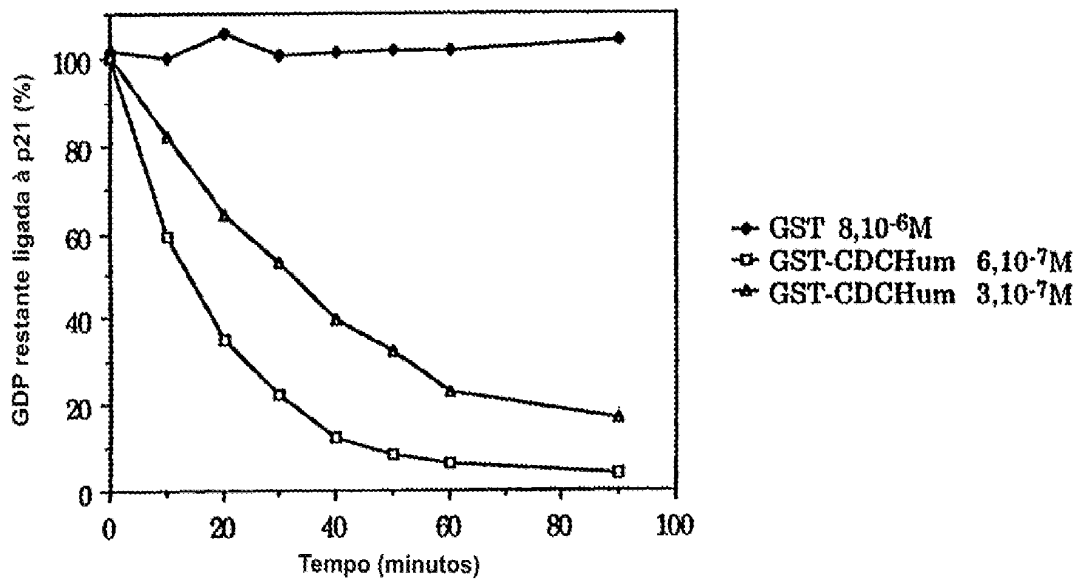


FIGURA 3